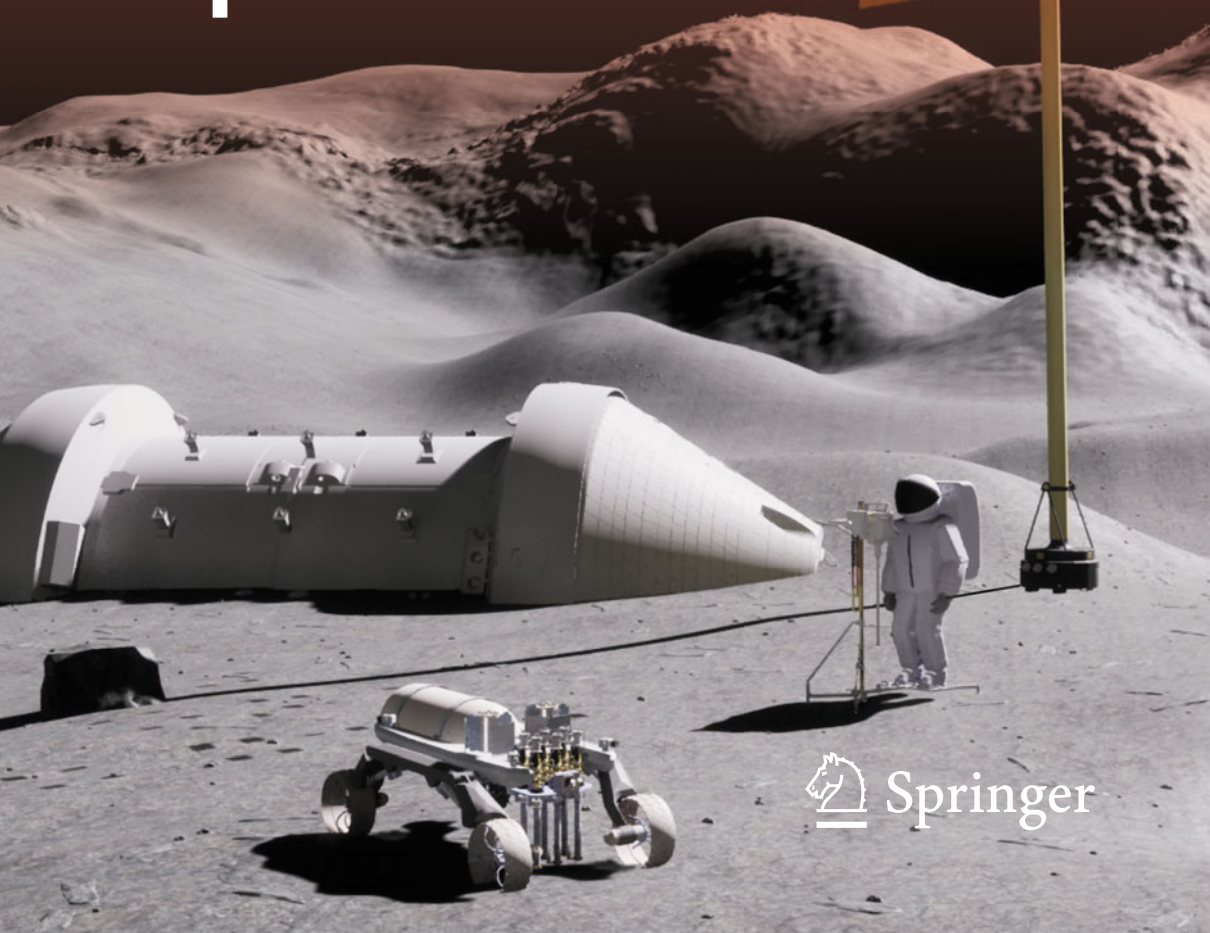


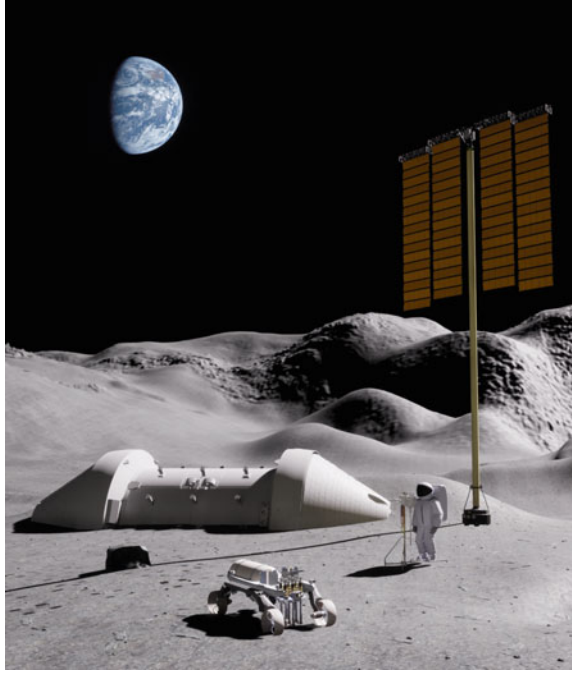
Viorel Badescu
Kris Zacny
Yoseph Bar-Cohen
Editors

Handbook of Space Resources



 Springer

Handbook of Space Resources



Viorel Badescu · Kris Zacny · Yoseph Bar-Cohen
Editors

Handbook of Space Resources

 Springer

Editors

Viorel Badescu
Candida Oancea Institute
Polytechnic University of Bucharest
Bucharest, Romania

Kris Zacny
Honeybee Robotics
Altadena, CA, USA

Yoseph Bar-Cohen
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, CA, USA

ISBN 978-3-030-97912-6 ISBN 978-3-030-97913-3 (eBook)
<https://doi.org/10.1007/978-3-030-97913-3>

© Springer Nature Switzerland AG 2023, corrected publication 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to all space miners

Preface

This book contains the latest perspectives on the space energy and material resources for human and robotic exploration and exploitation of the Solar System. It covers the latest advances as well as advantages and limitations of various space related systems and their potential applications to other fields. The book reviews various concepts and innovative options. It is a good resource for readers who are seeking background on various aspects of space-related activities.

The book is structured along logical lines of progressive thought and is divided into seven sections.

The first section deals with Technologies for Planetary Exploration and contains seven chapters. Chapter 1 is dealing with displaced non-Keplerian orbits for Sun and inner planet observation. Chapter 2 focuses on Dynamics and Control of Electrostatic Flight. Chapter 3 focuses on Tracking and thrust vectoring of E-sail-based spacecraft for solar activity monitoring. Chapter 4 deals with space elevators for space resource mining. Chapter 5 describes the orbital hub providing a LEO-infrastructure for multi-disciplinary science and commercial use cases. Chapter 6 covers instrumentation for planetary exploration. Finally, Chap. 7 covers space debris recycling by electromagnetic melting.

The second section of the book deals with Mercury and Venus and contains two chapters. Chapter 8 covers details about planetary exploration of Mercury with Bepi-Colombo and prospects of studying Venus during its cruise phase. Chapter 9 reports on the analysis of Smart Dust-based frozen orbits around Mercury.

The third section of the book, deals with the Moon, as a Steppingstone for In Situ Resource Utilization (ISRU) and it consists of six chapters. Chapter 10 deals with simulants in ISRU Technology Development. Chapter 11 focuses on regolith processing. Chapter 12 covers details on sintering as a method for construction of off-Earth infrastructure from off-Earth Materials. In Chap. 13 one can find information about the effects of mineral variations on the basalt sintering process and implications for ISRU. Chapter 14 proposes rocket mining for Lunar and Mars ISRU and Chap. 15 covers results about penetration investigations in Lunar regolith & simulants.

The fourth section of the book covers Mars, and it contains six chapters. Ice resource mapping of Mars is presented in Chap. 16 while Chap. 17 presents the

design and modeling of an electrochemical device producing methane/oxygen and polyethylene from in-situ resources on Mars. Chapter 18 covers details about mobile Mars habitation and Chap. 19 proposes local resource creation on Mars. The planetary exploration of Mars is covered in Chaps. 20 and 21 and deal with robotic deployment and installation of payloads on planetary surface.

The fifth section of the book refers to Asteroids and Comets and it consists of three chapters. In Chap. 22 the reader can find information about asteroid habitats and how one can live inside a hollow celestial body. Chapters 23 and 24 deals with resources from asteroids and comets and the asteroids as small bodies with big potential, respectively, while Chap. 25 focuses on the exploration of asteroids and comets with innovative propulsion systems.

The sixth section of the book deals with Ocean Worlds and it contains four chapters. Chapter 26 presents the Ocean Worlds and their interior processes and physical environments. Robotic mobility and sampling systems for Ocean World bodies are described in Chap. 27 while Chap. 28 focuses on the communication and obstacles detection using piezoelectric transducers in melting penetrator of deep ice at ocean worlds. Ice Melting Probes are covered in Chap. 29.

Finally, the seventh section of the book deals with economics and policies, and it contains five chapters. The Lunar Ore Reserves Standards 101 (LORS-101) are presented in Chap. 30. Chapter 31 presents the economics of space resources with details about future markets and value chains. In Chap. 32 the reader can find details about the lifetime embodied energy and a theory about the value of new space economy. Policy and legal processes and precedent for space mining are covered in Chap. 33 while Chap. 34 presents legal considerations for space resources.

The book allows the reader to acquire a clear understanding of the scientific, legal and policy fundamentals behind specific technologies to be used for the exploration and exploitation of space resources. The principal audience may consist of researchers and engineers who are involved or are interested in space exploration in general and in specific bodies exploration in specific. Also, the book may be useful for industry developers interested in taking advantage of national or international space programs towards implementing space related technologies. Finally, it may be used for undergraduates, graduates and postgraduates as well as doctoral studies and teaching.

Bucharest, Romania
Altadena, CA, USA
Pasadena, CA, USA

Viorel Badescu
Kris Zacny
Yoseph Bar-Cohen

Acknowledgments

A critical part of writing any book is the review process, and the authors and editors are very much obliged to the following researchers who patiently helped them read through subsequent chapters and who made valuable suggestions: Prof. Angel Abbud-Madrid (Colorado School of Mines, USA), Dr. Basu Abhijit (University of Indiana, USA), Dr. Antaeres Dawn Antoniuk-Pablant, Robert Ash, Dr. Nadya Balabanska (Jet Propulsion Laboratory, USA), Dr. Victor Baker, Dr. Akos Bazso, Dr. Mark Berggren, Dr. Sarah Deitrick (NASA Johnson Space Center, USA), Dr. Pierluigi Di Lizia (Politecnico di Milano, Italy), Dr. Antonio Diaz-Calderon, (Johns Hopkins University/Applied Physics Laboratory, USA), Dr. Christopher Dreyer (Colorado School of Mines, USA), Dr. Adriano Drummond C. Trindade (Pinheiro Neto Company), Dr. Rudolf Dvorak, Dr. Kyla Edison (Pacific International Space Center for Exploration Systems, USA), Dr. Bradley Edwards, Dr. Farid Gangami, Dr. Gonzalo Garcia-Atance Fatjo (University of Central Lancashire), Dr. Henry B. Garrett (Jet Propulsion Laboratory, USA), Dr. Kevin Grossman (NASA), Prof. Giorgio Guglieri (Politecnico di Torino, Italy), Dr. Axel Hagermann, Dr. Daoru (Frank) Han (Missouri University of Science and Technology, USA), Prof. Andrea Harrington (Colorado School of Mines, USA), Dr. Amy Hofmann (Jet Propulsion Laboratory, USA), Dr. Ken Hurst (Jet Propulsion Laboratory, USA), Dr. Erica Jawin (Smithsonian Institute, USA), Prof. Andrew Keane Woods (University of Arizona, USA), Dr. Beverley Kemmerer (NASA Kennedy Space Center, USA), Dr. Goro Komatsu (Università degli Studi “G. d’Annunzio”, Pescara, Italy), Dr. Julien Lamamy, Mr. Tim Lexen, Mr. Tom Lunne, Dr. Joe MacGregor (NASA Goddard Space Flight Center), Dr. Mohit Melwani-Daswani (Jet Propulsion Laboratory, USA), Dr. Daniel Miller (Lawrence Berkeley National Laboratory), Jeffrey Montes (Blue Origin), Dr. Yaniv Mordecai (Massachusetts Institute of Technology), Dr. Tony Muscatello, Dr. Heather Ann Oravec, Dr. Mark Panning (Jet Propulsion Laboratory, USA), Dr. Andrew Petruska (Colorado School of Mines, USA), Dr. Marco Quadrelli (Jet Propulsion Laboratory, USA), Stephen Ransom (SR Consultancy, Bremen, Germany), Prof. Laura Ray (Dartmouth College, USA), Dr. Florian Ruhhammer (ESA), Dr. Sarag Saikia (Spacefaring Technologies Pvt. Ltd., India), Dr. Jennifer Scully (Jet Propulsion Laboratory, USA), Dr. Vlada Stamenkovic (Blue

Origin), Prof. Paul van Susante (Michigan Technological University), Prof. Serkan Saydam (UNSW Sydney, Australia), Dr. Peter Swan (ISEC), Prof. Pavel Talalay, Dr. Fabio Tronchetti (University of Mississippi Law School, USA), Dr. Dave Vaniman, Dr. Hunter Williams (Honeybee Robotics, USA), Dr. Dale Winebrenner, Dr. Olivier Witasse (ESA) and Dr. Siggy Zerweckh.

The editors, furthermore, owe a debt of gratitude to all authors. Collaborating with these stimulating colleagues has been a privilege and a very satisfying experience.

Viorel Badescu
Kris Zacny
Yoseph Bar-Cohen

About This Book

Earth has limited material and energy resources while these resources in space are virtually unlimited. Moreover, further development of humanity will require going beyond our planet and this requires utilization of extra terrestrial resources.

This book covers present-day perspectives on the space energy and material resources for potential human use. It reviews the latest advances as well as advantages and limitations of various space related systems and their potential applications to other fields. The book reviews proposed concepts and innovative options as well as solutions. It is a good resource for readers who are seeking background on various aspects of space-related activities.

Written for researchers, engineers, students and businessmen who are interested in space resources exploration and exploitation.

Contents

Part I Technologies for Planetary Exploration

1 Displaced Non-Keplerian Orbits for Sun and Inner Planet Observation	3
Lorenzo Niccolai, Alessandro A. Quarta, and Giovanni Mengali	
2 Dynamics and Control of Electrostatic Flight	41
Marco B. Quadrelli, Michele Bechini, Joseph Wang, and Shota Kikuchi	
3 Tracking and Thrust Vectoring of E-Sail-Based Spacecraft for Solar Activity Monitoring	167
Marco Bassetto	
4 Space Elevator for Space-Resource Mining	229
Yoji Ishikawa	
5 <i>Orbital Hub</i>: Providing an LEO Infrastructure for Multi-disciplinary Science and Commercial Use Cases	249
Volker Maiwald, Dominik Quantius, Claudia Philpot, and Vincent Vrakking	
6 Instrumentation for Planetary Exploration	277
Emily F. Klonicki-Ference, Michael J. Malaska, Mark P. Panning, Sarah E. Waller, and Patrick J. Gasda	
7 Space Debris Recycling by Electromagnetic Melting	309
Jan Walter Schroeder, Gary Douglas Calnan, Abdoul-Aziz Bogno, Toby Joseph Daniel Mould, Romain Pecher Pecher, Joseph W. Pawelski, and Kai Staats	

Part II Mercury and Venus

8 Planetary Exploration of Mercury With BepiColombo and Prospects of Studying Venus During Its Cruise Phase 337
 Johannes Benkhoff and Joe Zender

9 Analysis of Smart Dust-Based Frozen Orbits Around Mercury 357
 Generoso Aliasi, Lorenzo Niccolai, Alessandro A. Quarta, and Giovanni Mengali

Part III The Moon, a Steppingstone to Planetary ISRU

10 Simulants in In-Situ Resource Utilization Technology Development 381
 Hunter Williams

11 Regolith Processing 399
 Kevin M. Cannon and Robert P. Mueller

12 Sintering: A Method for Construction of Off-Earth Infrastructure from Off-Earth Materials 429
 Liz Scott and Thao Nguyen

13 The Effects of Mineral Variations on the Basalt Sintering Process and Implications for In-Situ Resource Utilization (ISRU) 463
 Kyla P. Edison, G. Jeffery Taylor, Christian B. Andersen, and Rodrigo F. V. Romo

14 Rocket Mining for Lunar and Mars ISRU 491
 Matthew Kuhns, Roger J. Kuhns, Forrest Meyan, Philip Metzger, Hunter Williams, and Kris Zacny

15 Penetration Investigations in Lunar Regolith and Simulants 521
 Jared Atkinson

Part IV Mars

16 Ice Resource Mapping on Mars 583
 Nathaniel E. Putzig, Gareth A. Morgan, Hanna G. Sizemore, David M. Hollibaugh Baker, Eric I. Petersen, Asmin V. Pathare, Colin M. Dundas, Ali M. Bramson, Samuel W. Courville, Matthew R. Perry, Stefano Nerozzi, Zachary M. Bain, Rachael H. Hoover, Bruce A. Campbell, Marco Mastrogioseppe, Michael T. Mellon, Roberto Seu, and Isaac B. Smith

17	Design and Modeling of an Electrochemical Device Producing Methane/Oxygen and Polyethylene from In-Situ Resources on Mars	617
	Jeffery B. Greenblatt	
18	Mobile Mars Habitation	645
	Kürşad Özdemir and Süheyla Müge Halıcı	
19	Local Resource Creation on Mars	669
	Robert Zubrin	
20	Planetary Exploration of Mars	689
	Robert C. Anderson, James M. Dohm, Debra Buczkowski, Danielle Y. Wyrick, and Timothy J. Parker	
21	Robotic Deployment and Installation of Payloads on Planetary Surface	721
	Ashitey Trebi-Ollennu, Khaled Ali, Cristina Sorice, Won Kim, Steven Myint, Omair Khan, Philip Bailey, Hallie Abarca, Robert G. Deen, Jeng Yen, Justin N. Maki, Grace Lim, Nythi Udomkesmalee, and Jeffrey Umland	
 Part V Asteroids and Comets		
22	Asteroid Habitats—Living Inside a Hollow Celestial Body	763
	Werner Grandl and Clemens Böck	
23	Resources from Asteroids and Comets	787
	Daniel Britt and Kevin Cannon	
24	Asteroids: Small Bodies, Big Potential	803
	Akbar D. Whizin	
25	Exploration of Asteroids and Comets with Innovative Propulsion Systems	841
	Lorenzo Niccolai, Alessandro A. Quarta, and Giovanni Mengali	
 Part VI Ocean Worlds		
26	Ocean Worlds: Interior Processes and Physical Environments	873
	Samuel M. Howell and Erin J. Leonard	
27	Robotic Mobility and Sampling Systems for Ocean-World Bodies	907
	Hari Nayar, Junggon Kim, Gareth Meirion-Griffith, Brendan Chamberlain-Simon, Kalind Carpenter, Anna Boettcher, Michael Hans, Brian Wilcox, Jason Carlton, and Justin Jenkins	

28 Communication and Obstacles Detection Using Piezoelectric Transducers in a Penetrator Melting Deep Ice on Ocean Worlds 935
 Yoseph Bar-Cohen, Xiaoqi Bao, Hyeong Jae Lee, Benjamin Hockman, Mircea Badescu, Stewart Sherrit, and Shyh-Shiuh Lih

29 Ice Melting Probes 955
 Bernd Dachwald, Stephan Ulamec, Julia Kowalski, Marc S. Boxberg, Fabian Baader, Jens Biele, and Norbert Kömle

Part VII Economics and Policies

30 Lunar Ore Reserves Standards 101 (LORS-101) 999
 Carlos Daniel Espejel, Sophia Casanova, Serkan Saydam, and Julien-Alexandre Lamamy

31 The Economics of Space Resources: Future Markets and Value Chains 1023
 Mathias Link, Gary Martin, and Joseph Mousel

32 Lifetime Embodied Energy: A Theory of Value for the New Space Economy 1053
 George C. Lordos, Jeffrey A. Hoffman, and Olivier L. de Weck

33 Policy, Legal Processes and Precedents for Space Mining 1109
 Scot W. Anderson, Korey J. Christensen, Julia La Manna, Katherine Wood, Alex Gilbert, and Morgan Bazilian

34 Legal Considerations for Space Resources 1163
 Austin C. Murnane

Correction to: Instrumentation for Planetary Exploration C1
 Emily F. Klonicki-Ference, Michael J. Malaska, Mark P. Panning, Sarah E. Waller, and Patrick J. Gasda

Index 1201

About the Editors

Viorel Badescu is Professor of Engineering Thermodynamics and affiliated with Candida Oancea Institute at Polytechnic University of Bucharest. His mainstream scientific contributions consist of more than 300 papers and forty books related to various fields in science and engineering. Most of his research areas refer to terrestrial and space solar energy applications, including research on photo-thermal energy conversion and solar power plants, the physics of radiation, photovoltaic conversion of solar energy, solar radiation properties and solar radiation distribution and forecasting. Other fields of interest are statistical physics and thermodynamics, the physics of semiconductors, optimal control of thermal engineering processes, terraforming and macro-engineering. He is Associate Editor and Member of the editorial board of several international journals including *Space Power, Energy, Renewable Energy and Journal of Energy Engineering* and Reviewer for more than fifty international journals. He is also Member of several scientific societies including the International Solar Energy Society and the European Astronomical Society and Member of the Romanian Academy.

Dr. Kris Zacny is Vice President and Senior Research Scientist at Honeybee Robotics. His focus is on developing space mining robots for In Situ Resource Utilization (ISRU) and space exploration. Prior to joining Honeybee Robotics, Dr. Zacny worked in South African underground gold, diamond and coal mines. Dr. Zacny received his Ph.D. from UC Berkeley in Extraterrestrial Drilling and Mining. He participated in several Antarctic, Arctic, Atacama and Greenland expeditions. Dr. Zacny has over 200 publications, including several co-edited books, and multiple NASA Group Achievement Awards. Dr. Zacny is Principal Investigator of TRIDENT ice mining drill scheduled to launch to the Moon in 2023 and 2024, as part of PRIME1 lander and VIPER rover, respectively. He is also Principal Investigator of PlanetVac pneumatic mining systems scheduled to launch to the Moon in 2024.

Dr. Yoseph Bar-Cohen is Senior Research Scientist and Group Supervisor at Jet Propulsion Laboratory (JPL). He received his Ph.D. in Physics from the Hebrew University, Jerusalem, Israel, in 1979. His research is focused on electroactive mechanisms, biomimetics and NDE. He has edited and co-authored 12 books (including this one), co-authored over 460 publications and co-chaired 56 international conferences. He documented his co-conceived inventions in 134 New Technology Reports (NTR) and 42 registered patents. His notable initiatives include the SPIE conference on electroactive polymers (EAP) and the EAP-in-Action Session that he chaired for 21 years as well as his challenging engineers and scientists worldwide to develop a robotic arm driven by artificial muscles to wrestle with human and win. For his contributions to the field of artificial muscles, *Business Week* named him in April 2003 one of five technology gurus who are 'Pushing Tech's Boundaries'. His scientific and engineering accomplishments earned him two NASA Honor Award Medals, two SPIE's Lifetime Achievement Awards, Fellow of two technical societies: ASNT and SPIE, as well as many other honors and awards.

Part I
Technologies for Planetary Exploration

Chapter 1

Displaced Non-Keplerian Orbits for Sun and Inner Planet Observation



Lorenzo Niccolai, Alessandro A. Quarta, and Giovanni Mengali

Abstract A displaced non-Keplerian orbit is a trajectory whose orbital plane does not contain the center of mass of the primary body, so that its orbital maintenance requires the application of a suitable continuous thrust. Although the latter could be provided, in principle, by a low-thrust electric propulsion system, innovative propellantless propulsive technologies are well suited to such a mission scenario, due to their ability to generate thrust without requiring any propellant, thus significantly extending mission lifetime. This chapter focuses on the possibility of maintaining a displaced non-Keplerian orbit by means of both solar sails and electric solar wind sails (or E-sails). In fact, these advanced propulsion systems are both capable of generating a propulsive acceleration without consuming any propellant, by exploiting the solar radiation pressure (in case of solar sails) or the solar wind dynamic pressure (E-sails). This analysis uses recent models to provide a mathematical description of the propulsive acceleration generated by both propulsion systems, and different scenarios involving non-Keplerian orbits are analyzed. Particular focus is given to Type II displaced orbits, non-Keplerian orbits lying on the ecliptic plane, and heliostationary positions. Performance and attitude requirements are provided for each scenario. A linear stability analysis is also performed, in order to identify the combination of orbital parameters that characterize stable non-Keplerian orbits. The results suggest the feasibility of the mission scenarios discussed, but for most of them performance requirements are very demanding. A possible exception is non-Keplerian orbits lying on the ecliptic, which represent a very promising near-term scenario.

L. Niccolai (✉) · A. A. Quarta · G. Mengali
Department of Civil and Industrial Engineering, University of Pisa, Via G. Caruso 8, 56122 Pisa,
Italy
e-mail: lorenzo.niccolai@ing.unipi.it

A. A. Quarta
e-mail: a.quarta@ing.unipi.it

G. Mengali
e-mail: g.mengali@ing.unipi.it

List of Acronyms

CR3BP	Circular restricted three-body problem
DNKO	Displaced non-Keplerian orbit
E-sail	Electric solar wind sail
L_1	First collinear Lagrangian point
PDFO	Planet following displaced orbit

1.1 Introduction

A displaced non-Keplerian orbit (DNKO) is a closed spacecraft trajectory characterized by the fact that the primary body does not lie on the orbital plane, so that its orbital maintenance must be guaranteed by the application of a suitable continuous thrust. The first mission concept involving DNKOs was proposed by Forward (1984), who suggested the utilization of a classical (photonic) solar sail to generate a geosynchronous orbit whose orbital plane was either above or below the Earth's equatorial plane. The DNKO concept has been extensively investigated in the literature, and is surveyed in detail by McKay et al. (2011).

Although the generation of a DNKO is, in principle, obtainable with a generic low-thrust electric propulsion systems (Macdonald et al. 2011; Pan et al. 2019), the scientific investigation focused on DNKOs has recently received a significant impulse, mainly due to the renewed interest in propellantless propulsive systems, with special attention to solar sails and electric solar wind sails (or E-sails). These advanced propulsion systems are both capable of generating a propulsive acceleration without consuming any propellant, by exploiting the solar radiation pressure (in the case of solar sails) or the solar wind dynamic pressure (E-sails). A detailed description of such propellantless propulsive systems is given in Sect. 1.1.2, while a short review of possible applications of DNKOs is discussed in Sect. 1.1.1.

1.1.1 Mission Applications

An extensive literature exists on DNKO maintained by different propulsion means (McInnes 1998; McKay et al. 2011), including solar sails (Bookless et al. 2006; Gong et al. 2014a, b; Song et al. 2016), E-sails (Mengali et al. 2009; Niccolai et al. 2017a, 2018; Pan et al. 2020), and hybrid propulsion systems (Ceriotti et al. 2014; Mengali et al. 2007b, c). In particular, a number of possible DNKO-based mission scenarios have been proposed. Among them, some of the most important and promising options are described below. Ceriotti et al. (2011, 2014) discuss the possibility of observing the planetary polar regions (providing a continuous coverage) by means of a hybrid

propulsive system composed of a solar sail and an electric thruster. In their analysis, the Sun's and planet's gravitational attractions and the propulsive accelerations generated by the sail and by the electric thruster are all taken into account. Heiligers et al. (2011) analyze the maintenance of an out-of-equatorial plane geostationary orbit, a special case of circular DNKO which is synchronous with the Earth's rotation. Heiligers et al. (2014) propose the observation of high-latitude solar regions through DNKOs oscillating above and below the ecliptic plane, thus relaxing the constraint that requires the spacecraft to be constantly positioned at high heliocentric latitudes. Macdonald et al. (2011) discuss the creation of a communication relay for (near) future Mars exploration. They propose the utilization of highly non-Keplerian orbits to guarantee continuous communications between the Earth and Mars during solar conjunctions. Such orbits could be generated with an electric thruster or (more efficiently) with a hybrid system composed of an electric thruster and a solar sail.

A special application of the DNKO concept is the generation of a non-Keplerian orbit without displacement, that is, a closed trajectory whose orbital plane contains the primary body. These in-plane non-Keplerian orbits have the same shape as a classical conic section, but an orbital period different from that given by Kepler's third law. In a heliocentric scenario, such a particular closed orbit could have interesting scientific outcomes if designed to be circular with a radius not far from 1 au. In fact, in that case the planetary gravitational attraction should also be accounted for, so that the mission scenario would correspond to the maintenance of an artificial equilibrium point in the Sun-[Earth + Moon] circular restricted three-body problem (Aliasi et al. 2011, 2013a). An interesting potential application of such a scenario would be a solar warning mission placed at an L_1 -type artificial equilibrium point, closer to the Sun than the natural L_1 point, in order to increase the feasible warning time in case of dangerous solar flares (Aliasi et al. 2015; Vulpetti et al. 2017).

1.1.2 Propellantless System Options

The non-Keplerian nature of DNKOs implies that their generation requires a continuous thrust. Even though a quasi-DNKO could in principle be obtained with a succession of impulsive maneuvers (Caruso et al. 2019; McInnes 2011; Simo 2017), this analysis focuses on actual DNKOs. The requirement of a constantly acting thrust makes such a mission scenario well suited for propellantless propulsion systems, while DNKOs would be difficult to maintain with more conventional (chemical or electrical) thrusters. Accordingly, our analysis will be confined to solar sails and E-sails.

1.1.2.1 Solar Sail

A solar sail (see Fig. 1.1) is a thin reflective membrane that exchanges momentum with the impinging photons, that is, it exploits the solar radiation pressure as its

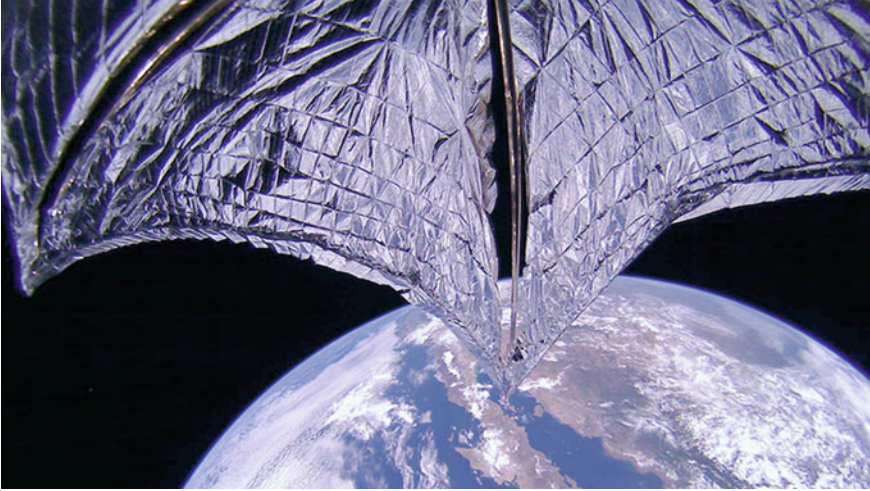


Fig. 1.1 In-space picture of the deployed solar sail of LightSail-2 mission. *Credits* The Planetary Society

propulsive source. A comprehensive review of solar sailing may be found in many references, including Fu et al. (2016), McInnes (1999), Vulpetti et al. (2015), and Wright (1992).

The solar sail working principle is based on solar radiation pressure, which has been well known for more than 50 years, as is confirmed by its use for attitude control purposes in the Mariner 10 mission. However, the possibility of using a solar sail as a primary propulsion system has long been called into question.

Indeed, before the last passage of Halley's comet in 1986, NASA was planning to perform a cometary rendezvous by means of a spacecraft propelled by a solar sail, but the project was eventually discarded due to the high associated risks. The solar sail concept received a renewed impulse at the end of the last century, mainly thanks to the progress in material sciences, which led to the first flight of a solar sail-based spacecraft, JAXA's Interplanetary Kitecraft Accelerated by Radiation Of the Sun (IKAROS) (Mori et al. 2010; Tsuda et al. 2011). IKAROS, which was launched in May 2010, successfully performed a Venus flyby, demonstrating the solar sail deployment capability and the effectiveness of an attitude control system based on reflectivity control devices (Funase et al. 2011). More recently, in January 2011, the NanoSail-D2 mission by NASA tested the deployment of a small square solar sail (with a side-length of 3.2 m) in a LEO (Johnson et al. 2011). The Planetary Society, a private company, launched the first private solar sail satellite, the LightSail-1 (Nye et al. 2016), equipped with a 32 m² square sail, which performed a fast deorbiting from a LEO thanks to the augmented atmospheric drag. Recently, the LightSail-2 mission (Betts et al. 2019) was the first to be capable of effectively modifying the spacecraft orbital parameters by means of a solar sail. The recent success of solar sails is also demonstrated by future space missions that will be equipped with such a propulsion

system, including JAXA's OKEANOS mission (currently still not financed) towards the Trojan asteroids (Funase et al. 2012; Mori et al. 2019), which should make use of a solar sail combined with an ion thruster powered by the solar cells placed on the sail surface, and NASA's Near Earth Asteroid Scout (NEA Scout), launched in November 2022, which should perform a flyby with a near-earth asteroid (Russell Lockett et al. 2019) by using a solar sail as primary propulsion system during the cruise phase (McNutt et al. 2014).

Modeling the propulsive acceleration generated by a solar sail is, in general, a complex task. A simple and effective tool is the so-called ideal thrust model, which assumes the sail shape to be a perfectly flat plane (referred to as nominal plane), and all of the photons impinging on the sail membrane to be specularly reflected. These hypotheses imply that the propulsive acceleration magnitude is overestimated, and its direction is considered parallel to the normal to the sail nominal plane in the direction opposite to the Sun. These results are neither realistic nor conservative, so that Dachwald (2004) proposed the use of a non-perfect specular reflection model, which accounts for the reduction of the propulsive acceleration magnitude by introducing a reflection efficiency, but leaves the direction unaffected and still assumes a flat shape. A further improvement in solar sail thrust modeling is constituted by the optical force model, in which the optical characteristics of the sail reflective surface are considered in the calculation of the propulsive acceleration. The optical force model will be used in our analysis, since it represents a good compromise between simplicity and accuracy, as implied by its use in the preliminary mission design phase of NASA's NEA-Scout mission. Other more complex models should also account for the variations of the optical properties with temperature (Ancona et al. 2017; Kezerashvili, 2008, 2014; Mengali et al. 2007a), or the influences of the light polarization and the features of the sail surface on the thrust generated. In this regard, the interested reader may refer to work by Vulpetti (2013) and Zola et al. (2018), where Fresnel reflection laws are taken into account to model the solar sail-generated thrust. Finally, the (small) thrust fluctuations associated with the variations of solar radiation pressure are also neglected. Further information may be found in the papers by Caruso et al. (2020) and Niccolai et al. (2019), where the use of reflectivity control devices is suggested to compensate for these environmental variations.

When the solar sail propulsive acceleration is described by an optical force model, the contributions of the absorbed, specularly reflected, scattered, and emitted photons are all taken into account, and the following expression for the propulsive acceleration vector \mathbf{a} (Heaton et al. 2015; McInnes 1999) is obtained

$$\mathbf{a} = \beta \left(\frac{\mu_{\odot}}{r^2} \right) \frac{\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}}{b_1 + b_2 + b_3} \{ b_1 \hat{\mathbf{r}} + [b_2 (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) + b_3] \hat{\mathbf{n}} \} \quad (1.1)$$

where β is the (dimensionless) lightness number of the solar sail-based spacecraft, that is, the ratio of the maximum propulsive acceleration magnitude that the sail can generate at a given Sun-spacecraft distance to the local Sun's gravitational acceleration magnitude, r is the Sun-spacecraft distance, μ_{\odot} is the Sun's gravitational

Table 1.1 Optical and force coefficients for a solar sail with an optical force model (Heaton et al. 2015, 2017)

Parameter	c_r	c_f	B_f	B_b	ε_f	ε_b	b_1	b_2	b_3
Value	0.91	0.89	0.79	0.67	0.025	0.27	0.095	0.8099	0.0151

parameter, $\hat{\mathbf{r}}$ is the Sun—spacecraft unit vector, and $\hat{\mathbf{n}}$ is the unit vector normal to the sail plane in the direction opposite to the Sun. Finally, the terms b_1 , b_2 and b_3 , referred to as force coefficients, may be obtained from the sail film optical properties as (Mengali et al. 2005)

$$b_1 = \frac{1 - c_r c_s}{2} \quad (1.2)$$

$$b_2 = c_r c_s \quad (1.3)$$

$$b_3 = \frac{B_f c_r (1 - c_s)}{2} + \frac{(1 - c_r)(\varepsilon_f B_f - \varepsilon_b B_b)}{2(\varepsilon_f + \varepsilon_b)} \quad (1.4)$$

where c_r is the reflection coefficient, c_s is the fraction of reflected photons that are specularly reflected, B_f (or B_b) is the front (or back) sail surface non-Lambertian coefficient, and ε_f (or ε_b) is the front (or back) sail surface emissivity coefficient. A recent estimation of the sail optical parameters in Eqs. (1.2)–(1.4) has been obtained during the preliminary design of the NEA-Scout mission. The experimental campaign has updated previous measurements (Heaton et al. 2015) and has relaxed the assumption of flat sail by accounting for the presence of millimeter-scale wrinkles that reduce the specular reflection fraction. The nominal values of the optical properties estimated by Heaton et al. (2017) are reported in Table 1.1.

An equivalent version of Eq. (1.1) is

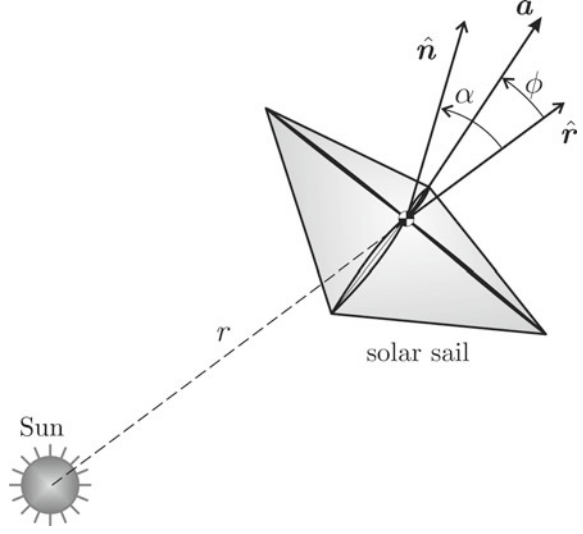
$$\mathbf{a} = a_c \left(\frac{r_\oplus}{r} \right)^2 \frac{\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}}{b_1 + b_2 + b_3} \{ b_1 \hat{\mathbf{r}} + [b_2 (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) + b_3] \hat{\mathbf{n}} \} \quad (1.5)$$

where the characteristic acceleration a_c is used as a performance parameter, that is, the maximum propulsive acceleration magnitude that the sail can generate at a Sun–Earth distance $r = r_\oplus \triangleq 1$ au. The characteristic acceleration of a solar sail-based spacecraft is

$$a_c = \beta \left(\frac{\mu_\odot}{r_\oplus^2} \right) \simeq \beta \times 5.93 \text{ mm/s}^2 \quad (1.6)$$

According to the thrust model of Eq. (1.5), the solar sail propulsive acceleration \mathbf{a} lies in the plane spanned by the normal unit vector $\hat{\mathbf{n}}$ and the radial direction defined by $\hat{\mathbf{r}}$, and its orientation can be controlled by suitably adjusting the attitude of the sail nominal plane. Let $\alpha \in [-\pi/2, \pi/2]$ rad be the sail pitch angle, that is, the angle

Fig. 1.2 Solar sail thrust vector characteristics



between the direction of \hat{r} and that of \hat{n} ; see Fig. 1.2. Positive (or negative) values of the pitch angle correspond to positive (or negative) values of the projection of \hat{n} along the specific angular momentum vector $\mathbf{h} \triangleq \mathbf{r} \times \mathbf{v}$, where \mathbf{v} is the spacecraft velocity vector. The sail pitch angle α is therefore given by

$$\alpha \triangleq \text{sign}(\mathbf{h} \cdot \hat{n}) \arccos(\hat{r} \cdot \hat{n}) \quad (1.7)$$

so that the propulsive acceleration vector can be rewritten by introducing the radial (a_r) and transverse (a_θ) components, defined as

$$a_r \triangleq \mathbf{a} \cdot \hat{r} = a_c \left(\frac{r_\oplus}{r} \right)^2 \frac{b_1 \cos \alpha + b_2 \cos^3 \alpha + b_3 \cos^2 \alpha}{b_1 + b_2 + b_3} \quad (1.8)$$

$$a_\theta \triangleq \|\mathbf{a} - a_r \hat{r}\| = a_c \left(\frac{r_\oplus}{r} \right)^2 \frac{b_2 \cos^2 \alpha \sin \alpha + b_3 \cos \alpha \sin \alpha}{b_1 + b_2 + b_3} \quad (1.9)$$

while the propulsive acceleration magnitude $a \triangleq \|\mathbf{a}\|$ is

$$a = \sqrt{a_r^2 + a_\theta^2} \quad (1.10)$$

The sail attitude modifies the thrust direction, as can be observed from Eq. (1.5). To quantify this effect, let $\phi \in [-\pi/2, \pi/2]$ rad be the sail cone angle, that is, the angle between the propulsive acceleration direction and the radial direction, viz.

$$\phi \triangleq \text{sign}(\alpha) \arccos\left(\frac{\hat{r} \cdot \mathbf{a}}{\|\mathbf{a}\|}\right) \equiv \text{sign}(\alpha) \arccos\left(\frac{a_r}{a}\right) \quad (1.11)$$

where a_r and a are given by Eqs. (1.8) and (1.10), respectively. The variation of ϕ with α is illustrated in Fig. 1.3a, from which it is clear that the same thrust angle can be obtained with two different sail pitch angles (that is, with two different sail attitudes). A further interesting consequence of Eq. (1.11) is that a solar sail can generate a maximum thrust angle less than 55° . Finally, note that an attitude variation (i.e., an orientation change of $\hat{\mathbf{n}}$) also modifies the magnitude of \mathbf{a} ; see Eq. (1.5). To account for this effect, a sort of “efficiency” parameter $\gamma \in [0, 1]$ is now introduced, defined as the ratio of the effective magnitude of the propulsive acceleration vector $\|\mathbf{a}\|$ to the maximum value of $\|\mathbf{a}\|$ (obtained when $\alpha = 0$, that is, in a Sun-facing condition), viz.

$$\gamma \triangleq \frac{\|\mathbf{a}\|}{\|\mathbf{a}\|_{\alpha=0}} \quad (1.12)$$

Clearly, γ gives the effective dimensionless magnitude of the propulsive acceleration. Equation (1.12) can be specialized to the solar sail case as

$$\gamma = \frac{a}{a_c \left(\frac{r_\oplus}{r}\right)^2} \quad (1.13)$$

The variation of γ as a function of the pitch angle α with an optical force model is shown in Fig. 1.3b, which highlights that smaller values of α correspond to larger propulsive acceleration magnitudes.

Therefore, to minimize the required sail performance, the pitch angle to be chosen for a given thrust angle is the minimum between the two possible values. Under such

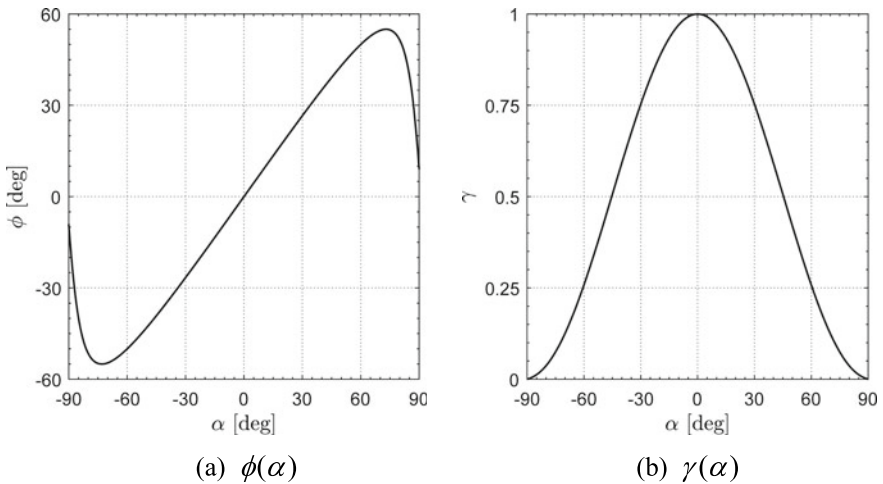
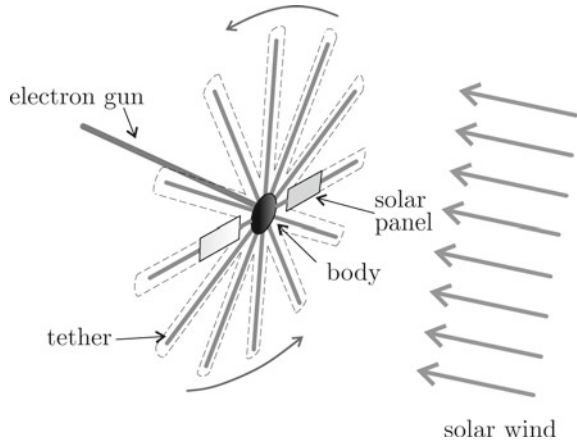


Fig. 1.3 Variation of the cone angle ϕ and the dimensionless acceleration γ as functions of the pitch angle α for a solar sail with an optical force model

Fig. 1.4 Basic sketch of an E-sail typical structure



an assumption, the function $\phi = \phi(\alpha)$ becomes invertible, see Fig. 1.3a, so that the value of α necessary for generating a thrust angle ϕ may be obtained with standard numerical methods.

1.1.2.2 *E-Sail*

The E-sail propulsion concept consists of a spinning grid of tethers, kept at a high (usually positive) potential by means of an electron gun (Janhunen 2004; Mengali et al. 2008). When the E-sail is immersed in a surrounding plasma, such as the solar wind, the electrostatic interaction between the charged grid and the incoming ions generates a momentum exchange and thus a net propulsive acceleration. A sketch of the basic structure of an E-sail is shown in Fig. 1.4, while Fig. 1.5 an artistic rendering.

A first validation test of the E-sail working principle was attempted with the Estonian satellite EstCube-1 (Lätt et al. 2014), whose aim was to test the plasma brake concept (Janhunen 2010), a derivation of the E-sail working principle useful for spacecraft deorbiting from LEO (Bassetto et al. 2018; Niccolai et al. 2017b; Orsini et al. 2018). Unfortunately, the tether unreel mechanism failed, probably due to vibrational loads during the launch phase (Slavinskis et al. 2015). The first experimental in-situ data on the E-sail principle should therefore be provided by the Finnish satellite Aalto-1 (Kestilä et al. 2013), which was launched in June 2018 and is equipped with a 100 m-long plasma brake tether to perform an end-of-life deorbiting phase (Khurshid et al. 2014).

The most recent tool for describing the thrust generated by an E-sail is the model proposed by Huo et al. (2018), according to which the propulsive acceleration vector \mathbf{a} is given by

$$\mathbf{a} = \tau \frac{a_c}{2} \left(\frac{r_{\oplus}}{r} \right) [\hat{\mathbf{r}} + (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}}] \quad (1.14)$$

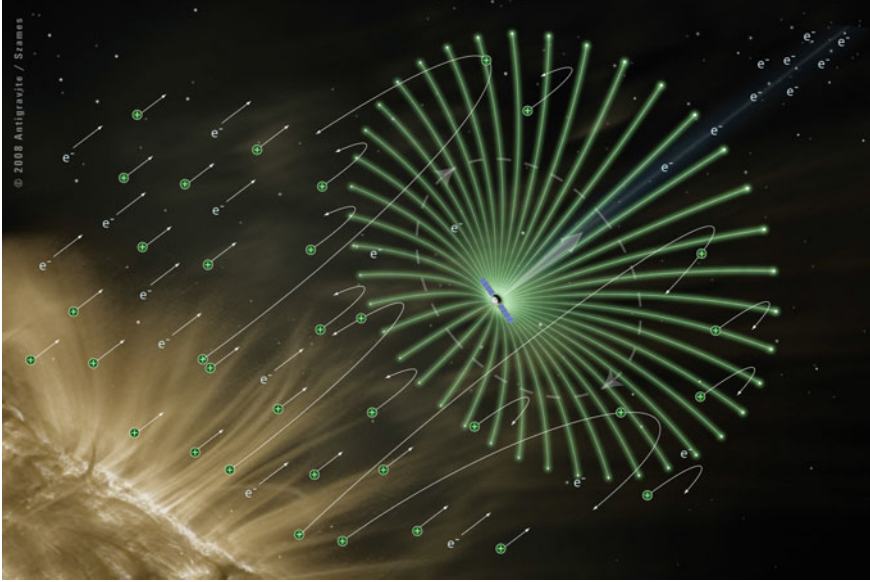


Fig. 1.5 E-sail artistic rendering by Alexandre Szames, Antigravité (Paris)

where the same nomenclature as that of Eq. (1.1) is adopted. In Eq. (1.14), $\tau \in \{0, 1\}$ is a switching dimensionless parameter that accounts for the possibility of switching the electron gun either on ($\tau = 1$) or off ($\tau = 0$), while a_c is the characteristic acceleration, with the same definition as that used for a solar sail.

Similarly to the solar sail case, Eq. (1.14) implies that the E-sail propulsive acceleration \mathbf{a} belongs to the plane defined by the normal unit vector $\hat{\mathbf{n}}$ and the radial unit vector $\hat{\mathbf{r}}$; see Fig. 1.6.

Using the same definitions for the pitch angle α , see Eq. (1.7), and the cone angle ϕ , see Eq. (1.11), the following relation $\phi = \phi(\alpha)$ can be derived from Eq. (1.14)

$$\phi = \arccos\left(\frac{1 + \cos^2 \alpha}{\sqrt{1 + 3 \cos^2 \alpha}}\right) \quad (1.15)$$

which is illustrated in Fig. 1.7a. The latter highlights that the maximum thrust angle is about 20° and, as such, an E-sail has a limited capability of generating a transverse thrust component (Quarta et al. 2016).

Moreover, similarly to the solar sail case, the same value of ϕ can be obtained with two different values of α . Indeed, Eq. (1.15) can be analytically inverted to obtain

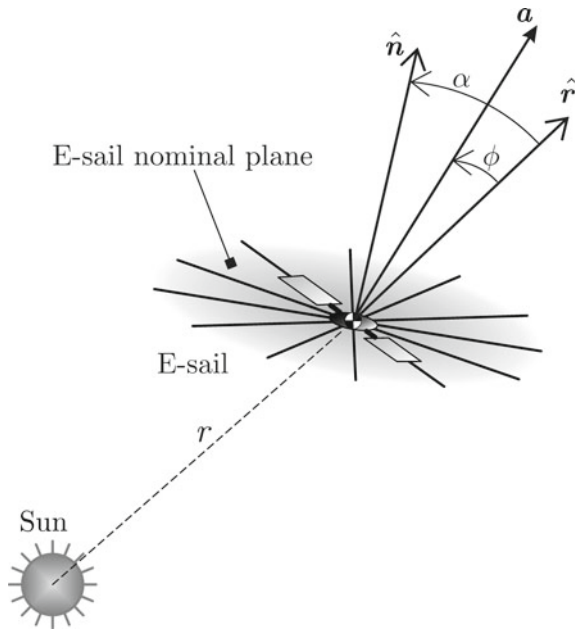


Fig. 1.6 E-sail thrust vector characteristics

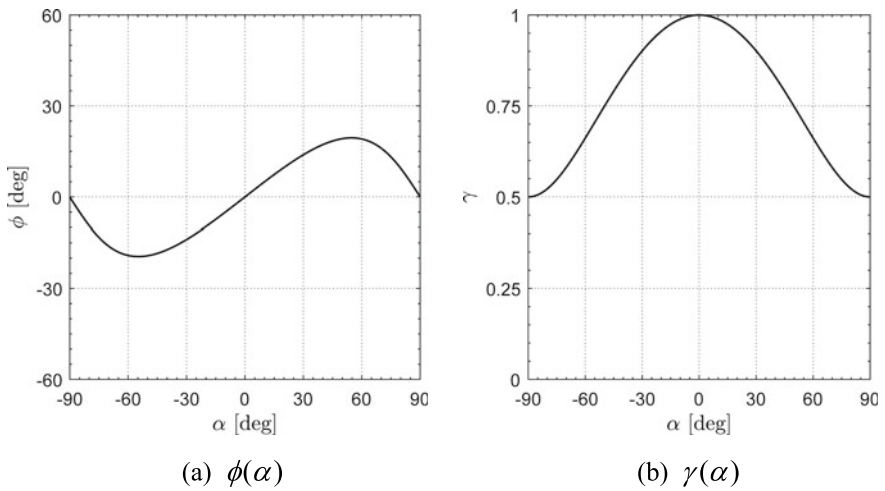


Fig. 1.7 Variation of cone angle ϕ and dimensionless acceleration γ as a function of the pitch angle α for an E-sail

$$\alpha = \begin{cases} \text{sign}(\phi) \arccos \left[\frac{\sqrt{2(3 \cos^2 \phi + \cos \phi \sqrt{9 \cos^2 \phi - 8} - 2)}}{2} \right] \\ \text{sign}(\phi) \arccos \left[\frac{\sqrt{2(3 \cos^2 \phi - \cos \phi \sqrt{9 \cos^2 \phi - 8} - 2)}}{2} \right] \end{cases} \quad (1.16)$$

where the first expression gives the smaller value of α for a given value of ϕ . The magnitude of \mathbf{a} can be expressed from Eq. (1.14) as

$$a = \tau \frac{a_c}{2} \left(\frac{r_\oplus}{r} \right) \sqrt{1 + 3 \cos^2 \alpha} \quad (1.17)$$

and the dimensionless parameter γ defined in Eq. (1.12) can be adapted to the E-sail case as

$$\gamma = \frac{\|\mathbf{a}\|}{\tau a_c \left(\frac{r_\oplus}{r} \right)} = \frac{\sqrt{1 + 3 \cos^2 \alpha}}{2} \quad (1.18)$$

which gives the results shown in Fig. 1.7b. Again, smaller pitch angles correspond to larger values of γ , so that, when a specific thrust angle must be reached, the smaller value of pitch angle is preferable in terms of performance requirements. Accordingly, only the first expression reported in Eq. (1.16) will be used in this analysis. A further consideration that may be derived from Fig. 1.7b is that an E-sail generates a nonzero thrust even when $\alpha = \pm\pi/2$ rad, and the only way to track a Keplerian arc is therefore to switch the electron gun off, which amounts to setting $\tau = 0$ in Eq. (1.14).

1.2 Displaced Non-Keplerian Orbits in a Heliocentric Scenario

This chapter is focused on the analysis of a heliocentric DNKO scenario maintained with a solar sail or an E-sail. Circular and elliptic DNKO cases are considered in Sects. 1.2.1 and 1.2.2, respectively. Examples of possible applications are given in Sect. 1.2.3, while a linear stability analysis of circular DNKOs is provided in Sect. 1.2.4.

Let S be the center of mass of a spacecraft equipped with a propellantless propulsion system, which moves under the gravitational attraction of the Sun only and generates a propulsive acceleration \mathbf{a} . The spacecraft is tracking a DNKO with angular velocity $\boldsymbol{\omega}$ and, without loss of generality, we assume that its orbital plane is parallel to the ecliptic plane. The Sun's center of mass is located at point O , whose projection

on the orbital plane is C . Let $\hat{\rho}$ be the unit vector from C to S . Hence, the scalar quantities ρ and r denote the C - S and O - S distances, respectively. Note that ρ , r and ω are all constant quantities in a circular DNKO, while they are time-varying in an elliptic DNKO.

The generation of a DNKO is possible only by maintaining the equilibrium between gravitational, propulsive, and centrifugal accelerations acting on the spacecraft along the whole orbit. Note that, for symmetry reasons, \hat{a} and $\hat{\rho}$ must belong to the plane containing \hat{r} and ω . Since for both a solar sail and an E-sail a lies in the plane spanned by \hat{r} and \hat{n} , the gravitational, propulsive, and centrifugal forces all belong to the same plane. These considerations about orbital maintenance may immediately be extended to the general case of a DNKO whose orbital plane is not parallel to the ecliptic.

1.2.1 Circular DNKOs

Assume that the spacecraft S is tracking a circular DNKO, as is illustrated in Fig. 1.8.

The constant spacecraft angular velocity is ω , with $\omega \triangleq \|\omega\|$ and its orbital period is $T = 2\pi/\omega$. Since in a circular DNKO both ρ and r are fixed, the elevation angle

$$\psi \triangleq \arccos\left(\frac{\rho}{r}\right) \quad (1.19)$$

is a constant of motion. Each circular DNKO is fully characterized by three independent parameters, as, for example, the elevation angle ψ (or, equivalently, the displacement $z \triangleq r \sin \psi$), the angular velocity ω , and $\omega_K \triangleq \sqrt{\mu_{\odot}/r^3}$ (equivalently, the Sun-spacecraft distance r or the orbital radius ρ). Note that ω_K denotes the angular velocity of a spacecraft moving in a Keplerian orbit with radius r . According to the traditional classification proposed by McInnes (1999), DNKOs can be divided into three categories.

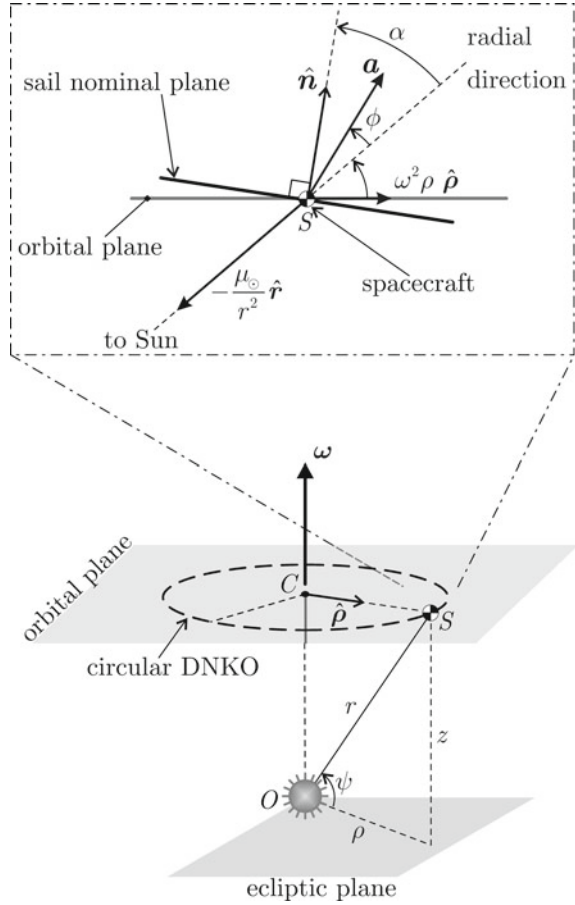
More precisely, a Type I DNKO is characterized by an angular velocity equal to the Earth's mean motion, that is, $\omega = \omega_{\oplus} = 0.9856$ deg/day. Hence, Type I DNKOs are defined by two parameters only (ϕ and ω_K). Instead, a spacecraft placed in a Type II DNKO has an angular velocity $\omega = \omega_K$. Finally, circular DNKOs with unconstrained values of ω , ψ and ρ are referred to as Type III DNKOs.

The condition for orbital maintenance of a generic DNKO can be conveniently expressed by means of two scalar equations, each one imposing the force balance along a direction of the plane defined by \hat{r} and $\hat{\omega}$. The first equation involves the component along \hat{r} , while the second is written along the direction orthogonal to \hat{r} , viz.

$$a \cos \phi = \frac{\mu_{\odot}}{r^2} - \omega^2 \rho \cos \psi \quad (1.20)$$

$$a \sin \phi = \omega^2 \rho \sin \psi \quad (1.21)$$

Fig. 1.8 Sketch of a circular displaced non-Keplerian orbit



Equations (1.20) and (1.21) must be specialized to the specific propulsion system that is considered. In particular, Eqs. (1.20) and (1.21) can be rewritten as

$$a_c \left(\frac{r_\oplus}{r} \right)^\eta \gamma \cos \phi = \frac{\mu_\odot}{r^2} - \omega^2 \rho \cos \psi \quad (1.22)$$

$$a_c \left(\frac{r_\oplus}{r} \right)^\eta \gamma \sin \phi = \omega^2 \rho \sin \psi \quad (1.23)$$

where γ is defined by Eq. (1.13) and $\eta = \{1, 2\}$ identifies the propulsive system, either a solar sail (when $\eta = 2$) or an E-sail (when $\eta = 1$). In addition, in the E-sail case we assume $\tau \equiv 1$ to guarantee a constantly acting thrust. Recalling that $\rho = r \cos \psi$, Eqs. (1.22) and (1.23) give the following requirements in terms of thrust angle ϕ and characteristic acceleration a_c

$$\phi = \arctan \left[\frac{(\omega/\omega_K)^2 \tan \psi}{1 + \tan^2 \psi - (\omega/\omega_K)^2} \right] \quad (1.24)$$

$$a_c = \frac{\mu_\odot}{\gamma r_\oplus^2} \left(\frac{r_\oplus}{r} \right)^{2-\eta} \left[1 - \frac{(\omega/\omega_K)^2}{1 + \tan^2 \psi} \right] \sqrt{1 + \frac{\tan^2 \psi}{[(1 + \tan^2 \psi)/(\omega/\omega_K)^2 - 1]^2}} \quad (1.25)$$

A relevant implication of Eq. (1.24) is that the thrust angle necessary for maintaining a circular DNKO is independent of the specific propulsive system (in fact, it is independent of η). Equations (1.24) and (1.25) refer to a generic Type III circular DNKO and can be specialized to the cases of Type I or Type II orbits. In fact, a Type I DNKO is simply characterized by $\omega = \omega_\oplus$, while a Type II DNKO is obtained when $\omega = \omega_K$.

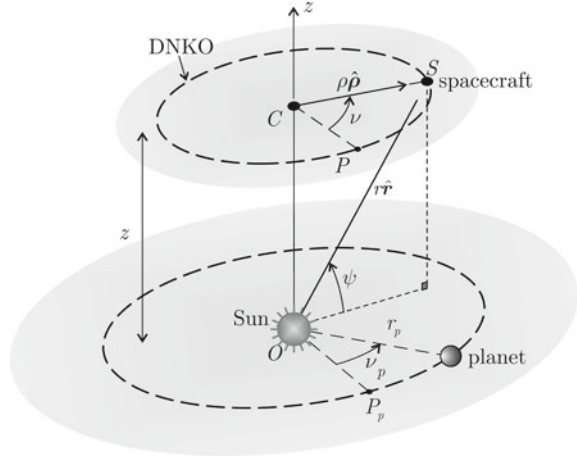
The procedure for determining the conditions to be met for the maintenance of a given circular DNKO can be summarized as follows. The parameters ω , ω_K and ψ are fixed when the DNKO is selected. The required value of the cone angle ϕ is found by means of Eq. (1.24). From ϕ , the possible values of the pitch angle α are obtained through numerical or graphical methods (see Fig. 1.3b) in the solar sail case, or through Eq. (1.16) in the E-sail case. As already stated, the smaller value of the pitch angle α is preferable, because it gives a larger value of γ (see Figs. 1.3b and 1.7b). The corresponding value of the dimensionless propulsive acceleration γ is then found by Eq. (1.13) or by Eq. (1.18). Finally, the characteristic acceleration required for orbital maintenance is obtained from the DNKO parameters and the value of γ through Eq. (1.25).

1.2.2 Elliptic DNKOs

The analysis discussed in the previous section can be extended to the case of elliptic DNKOs. The simplest application of an elliptic DNKO is in the observation and the scientific analysis of the polar regions above a planet. Accordingly, the analysis in this section will concentrate on this scenario. The situation is schematically illustrated in Fig. 1.9, where P_p and P (or ν_p and ν) are the planet and spacecraft perihelion (or true anomaly), respectively, and the other quantities have the same definition as those previously used.

Bearing in mind that in our analysis the gravitational attraction of the planet is neglected, the obtained results will be acceptable as long as the spacecraft remains outside the planetary sphere of influence. The orbital maintenance requires the gravitational, propulsive and centrifugal acceleration to be balanced on the plane defined by the radial unit vector and the angular velocity vector, or equivalently by $\hat{\rho}$ and the normal to the DNKO plane, directed along the z -axis in Fig. 1.9.

Fig. 1.9 Sketch of an elliptic DNKO



Accordingly, the spacecraft motion in an elliptic DNKO may be characterized by writing its equations of motion along the radial direction (that is, along $\hat{\rho}$) on the DNKO plane and along the vertical component (or along the unit vector \hat{k} associated to the z -axis), viz.

$$\ddot{\rho} = -\frac{\mu_{\odot}}{r^3}\rho + a_{\rho} + \frac{h_z^2}{\rho^3} \quad (1.26)$$

$$\ddot{z} = -\frac{\mu_{\odot}}{r^3}z + a_z \quad (1.27)$$

where $h_z \triangleq \rho^2 \dot{\nu}$ is the (constant) component of the angular momentum vector perpendicular to the DNKO orbital plane, a_{ρ} and a_z are the in-plane radial and vertical components of the propulsive acceleration, given by

$$a_{\rho} = \tau a_c \left(\frac{r_{\oplus}}{r} \right)^{\eta} \gamma \cos(\phi + \psi) \quad (1.28)$$

$$a_z = \tau a_c \left(\frac{r_{\oplus}}{r} \right)^{\eta} \gamma \sin(\phi + \psi) \quad (1.29)$$

Unlike the circular DNKO case, a switching parameter $\tau \in [0, 1]$ has been inserted in the equations because the spacecraft thrust must now be modulated for orbital maintenance. In the solar sail case, such a thrust modulation may be achieved by means of electrochromic control devices (Aliasi et al. 2013b; Funase et al. 2011; Lücking et al. 2012; Mengali et al. 2016), which change their optical properties when a voltage is applied (Monk et al. 2007). On the other hand, for an E-sail the thrust modulation can be obtained by adjusting the grid voltage (Toivanen et al. 2013, 2017), which is directly proportional to the propulsive acceleration magnitude (Huo et al. 2018).

Since in our analysis elliptic DNKO are used to observe the polar regions of a planet with a non-negligible (heliocentric) orbital eccentricity, it is assumed that the DNKO has the same shape and the same orbital period as the target planet's orbit, which lies on a plane parallel to the planet's orbital plane. In this case, the Sun, the spacecraft, and the target planet always belong to the same plane, and the orbit is referred to as planet following displaced orbit (PFDO). A PFDO is characterized by a parameter χ defined as

$$\chi \triangleq \frac{\rho}{r_p} \quad (1.30)$$

where r_p is the Sun–planet distance. From its definition, χ is constant in a PFDO, as is the vertical coordinate z . The following constraints must therefore be met for orbital maintenance

$$\dot{\chi} = \ddot{\chi} = 0, \quad \dot{z} = \ddot{z} = 0 \quad (1.31)$$

Substituting Eq. (1.30) into Eqs. (1.26) and (1.27), and enforcing the constraints (1.31), we get

$$a_\rho = \frac{\mu_\odot}{r_p^2} \chi \left(\frac{\cos^3 \psi}{\chi^3} - 1 \right) \quad (1.32)$$

$$a_z = \frac{\mu_\odot}{r_p^2} \frac{\cos^2 \psi \sin \psi}{\chi^2} \quad (1.33)$$

where $r_p = a_p(1 - e_p^2)/(1 + e_p \cos v_p)$, while a_p and e_p are the planet's heliocentric orbit semimajor axis and eccentricity, respectively. The equilibrium conditions become

$$a_\rho = \tau a_c \left(\frac{r_\oplus}{r} \right)^\eta \gamma \cos(\phi + \psi) = \frac{\mu_\odot}{r_p^2} \chi \left(\frac{\cos^3 \psi}{\chi^3} - 1 \right) \quad (1.34)$$

$$a_z = \tau a_c \left(\frac{r_\oplus}{r} \right)^\eta \gamma \sin(\phi + \psi) = \frac{\mu_\odot}{r_p^2} \frac{\cos^2 \psi \sin \psi}{\chi^2} \quad (1.35)$$

and the following expression for the cone angle is eventually obtained

$$\phi = \arctan \left(\frac{\cos^2 \psi \sin \psi}{\cos^3 \psi - \chi^3} \right) - \psi \quad (1.36)$$

The latter equation provides a constraint to the spacecraft attitude, similarly to what happens in the circular DNKO case.

The orbital maintenance requires the thrust magnitude to be suitably adjusted, as implied by Eqs. (1.34) and (1.35). This variation cannot be achieved by changing the spacecraft attitude, which is constrained according to Eq. (1.36). Therefore, it is necessary to implement a control law that varies the generated thrust by modulating the switching parameter τ as

$$\tau = \frac{\mu_{\odot} \chi^{\eta}}{a_c \gamma r_{\oplus}^{\eta} r_p^{(2-\eta)} \cos^{\eta} \psi} \sqrt{\chi^2 + \frac{\cos^4 \psi}{\chi^4} - 2 \frac{\cos^3 \psi}{\chi}} \quad (1.37)$$

where γ is a function of the cone angle α .

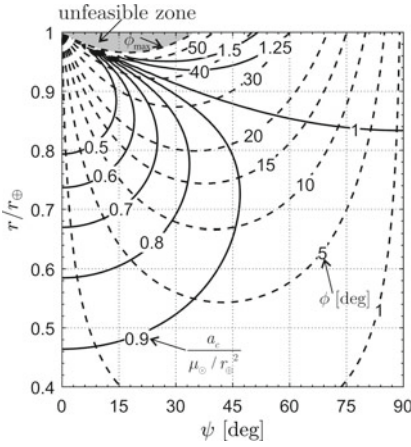
The procedure required to implement the control law for orbital maintenance of a PFDO can be summarized in the following steps. It is assumed that the planet's orbital parameters $\{a_p, e_p\}$ are known, and the PFDO characteristics $\{z, \chi\}$ are selected. For each value of the spacecraft true anomaly ν on the PFDO (with $\nu = \nu_p$), the elevation angle is

$$\psi = \arctan\left(\frac{z}{\chi r_p}\right) \equiv \arctan\left(\frac{z(1 - e_p \cos \nu_p)}{\chi a_p (1 - e_p^2)}\right) \quad (1.38)$$

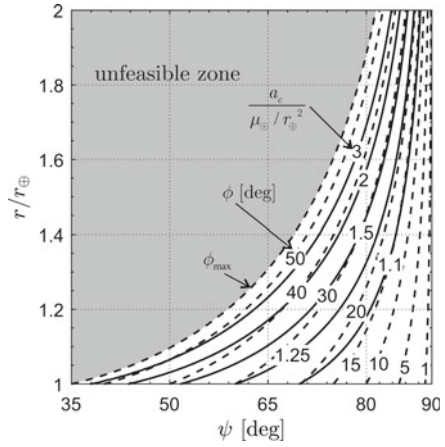
while ϕ is given by Eq. (1.36) as a function of the propulsive system (i.e., η), χ and ψ , the latter being a function of the true anomaly (or, equivalently, of time). The sail pitch angle is obtained by means of numerical or graphical methods (see Fig. 1.3b) in the solar sail case, or by Eq. (1.16) in the E-sail case. The corresponding value of γ is provided by Eq. (1.13) or Eq. (1.18), depending on the thruster type. The propulsive requirement τa_c is given by Eq. (1.37). Note that τa_c is the product of a fixed value (a_c), which characterizes the propulsive system performance, and a varying parameter (τ) that accounts for the thrust modulation during the flight.

1.2.3 Case Study

Circular DNKO could in principle be used to provide a continuous observation of the polar regions of the Sun. In order to estimate the feasibility of such mission scenarios, Figs. 1.10 and 1.11 show the requirements for orbital maintenance of a Type I DNKO (i.e., a DNKO with period equal to 1 year), in terms of thrust angle and characteristic acceleration, both for a solar sail- and an E-sail-based spacecraft. The empty zones in the graphs are characterized by required thrust angles that are unfeasible with the specific propulsive systems. Clearly, the thrust constraint is more demanding for an E-sail-based spacecraft. The obtained results highlight that using a solar sail is more convenient for $r < r_{\oplus}$, whereas an E-sail is better for $r > r_{\oplus}$. Finally, a general consideration of the presented results is that circular DNKO with large elevation angles and small heliocentric distances are very demanding in terms of propulsive requirements.

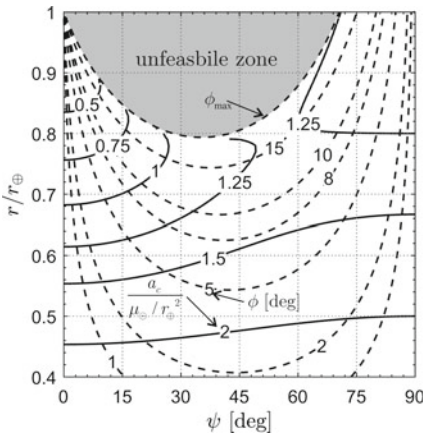


(a) Case of $r < r_{\oplus}$

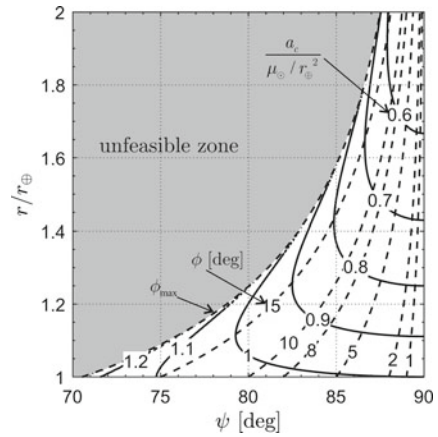


(b) Case of $r > r_{\oplus}$

Fig. 1.10 Thrust angle and propulsive requirements for orbital maintenance of a Type I circular DNKO by means of a solar sail-generated thrust



(a) Case of $r < r_{\oplus}$



(b) Case of $r > r_{\oplus}$

Fig. 1.11 Thrust angle and propulsive requirements for orbital maintenance of a Type I circular DNKO by means of an E-sail-generated thrust

A special mission application of circular DNKO is constituted by a Type II DNKO (McInnes 1999), whose orbital period is equal to that of a Keplerian orbit with radius r . A spacecraft placed on such a trajectory is able to observe the polar regions of a planet with a nearly circular orbit. For a Type II DNKO, the condition $\omega = \omega_K$ holds, and Eq. (1.24) reduces to

$$\phi = \frac{\pi}{2} - \psi \quad (1.39)$$

A relevant implication of Eq. (1.39) is that a minimum value of possible elevation angle exists, and is given by

$$\psi_{\min} = \frac{\pi}{2} - \phi_{\max} \quad (1.40)$$

where $\phi_{\max} \simeq 55^\circ$ in the solar sail case (see Fig. 1.3a) whereas $\phi_{\max} \simeq 19.5^\circ$ in the E-sail case (see Fig. 1.7a). Accordingly, these orbits are physically feasible for propellantless propulsive systems only when large values of elevation angle are considered, that is, about $\psi > 35^\circ$ in the solar sail case and $\psi > 70^\circ$ in the E-sail case.

Figure 1.12 shows the propulsive requirements for Type II DNKO maintenance as a function of the elevation angle and the Sun–spacecraft distance. The first interesting implication of Fig. 1.12a is that, for a solar sail-based spacecraft, the required characteristic acceleration is a function of the elevation angle only and is not affected by the Sun–spacecraft distance. This is not the case when an E-sail-based spacecraft is considered, due to the different nature of the propulsive system that affects the thrust dependence on r . In this case, indeed, the required value of a_c is a decreasing function of ψ . However, it is evident that the characteristic accelerations needed for orbital maintenance are large for both propulsive systems, which makes this fascinating mission scenario beyond the current or near-term technology level. In particular, considering a solar sail-generated Type II DNKO, the required characteristic acceleration is always larger than the reference gravitational acceleration magnitude $\mu_{\odot}/r_{\oplus}^2 \simeq 5.93 \text{ mm/s}^2$. This corresponds to a lightness number β greater than one, while the current state of the art represented by the planned NEA-Scout mission is $\beta \simeq 0.0101$ (Pezent et al. 2019). Similarly, in the E-sail case, a circular DNKO with a radius close to r_{\oplus} is achievable only for an E-sail with $a_c \simeq 6 \text{ mm/s}^2$, which is at least six times larger than the currently hypothesized maximum value of 1 mm/s^2 . Another special case of circular DNKO is obtained when the elevation angle ψ is zero, that is, when the DNKO degenerates in an orbit lying on the ecliptic plane. This could be exploited to generate an artificial Lagrangian point (Aliasi et al. 2011; Baig et al. 2008; Morimoto et al. 2007) closer to the Sun than the natural L_1 point by suitably adjusting the orbital period. In this case the equilibrium heliocentric distance is denoted by r_e . For a circular DNKO lying on the ecliptic, Eq. (1.24) gives $\phi = 0$, which corresponds to a Sun-facing attitude, i.e., $\alpha = 0$ (see Figs. 1.3a and 1.7a) and $\gamma = 1$ (see Figs. 1.3b and 1.7b) for both a solar sail and an E-sail. The required propulsive acceleration (1.25) becomes

$$a_c = \frac{\mu_{\odot}}{r_{\oplus}^2} \left(\frac{r_{\oplus}}{r_e} \right)^{2-\eta} \left[1 - \left(\frac{\omega}{\omega_K} \right)^2 \right] \quad (1.41)$$

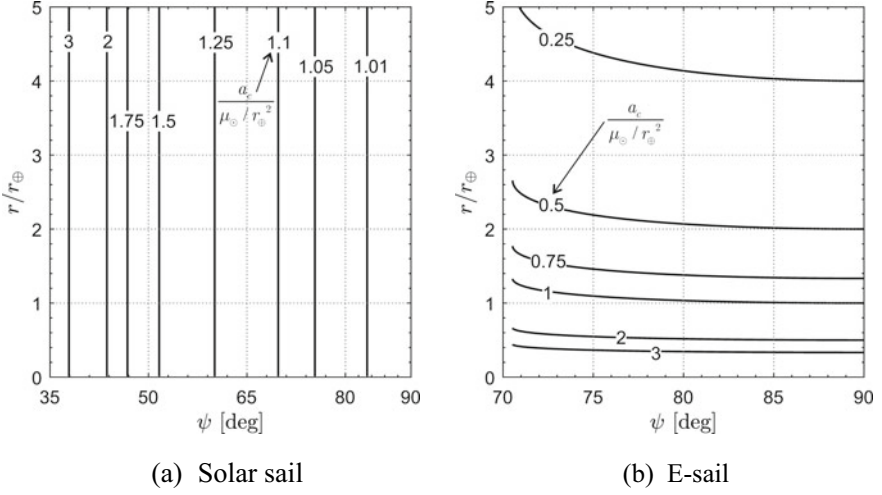


Fig. 1.12 Propulsive requirement for Type II DNKO maintenance as a function of Sun–spacecraft distance and elevation angle

Note that Eq. (1.41) gives $a_c = 0$ for $\omega = \omega_K$, corresponding to a Keplerian orbit. Moreover, Eq. (1.41) implies that, for a solar sail-propelled spacecraft ($\eta = 2$), the propulsive requirement only depends on the ratio ω/ω_K , but this is not true in the E-sail case.

Figure 1.13 shows the achievable heliocentric distance of an ecliptic circular DNKO for different values of characteristic acceleration and orbital period for a solar sail- or an E-sail-based spacecraft, respectively. The generation of an artificial Lagrangian point in the Sun-[Earth + Moon] system ($T = 1$ year) seems feasible for near-term technology level, but the propulsive requirements imply that a very small geocentric distance must be assumed, so that the assumption of two-body dynamics is unrealistic. A thorough analysis of this mission scenario will be given in Sect. 1.3, when the planetary gravity will also be considered.

The last discussed mission application consists in the maintenance of a heliostationary condition, that is, a circular orbit that degenerates in a single point placed above a Sun’s pole at a distance r_H (McInnes, 2003; Mengali et al. 2007a; Quarta et al. 2020). In this case, the spacecraft is fixed with respect to a generic heliocentric inertial reference frame τ_I , as illustrated in Fig. 1.14, and is therefore able to perform a constant observation of one of the Sun’s polar regions. The elevation angle required for maintaining a heliostationary position is $\psi = 90^\circ$, which corresponds to a Sun-facing attitude, i.e., $\alpha = 0$, $\phi = 0$, and $\gamma = 1$. The propulsive requirements are found by balancing the propulsive acceleration, viz.

$$a_c = \frac{\mu_\odot}{r_H^2} \left(\frac{r_H}{r_\oplus} \right)^\eta \quad (1.42)$$

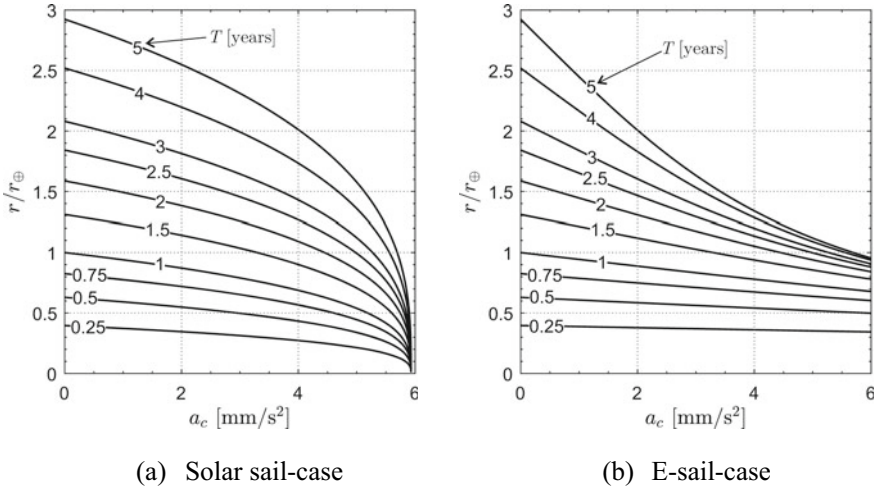
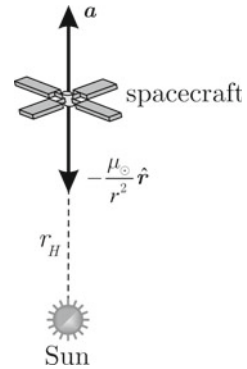


Fig. 1.13 Heliocentric radius r of ecliptic non-Keplerian orbits as a function of a_c and T

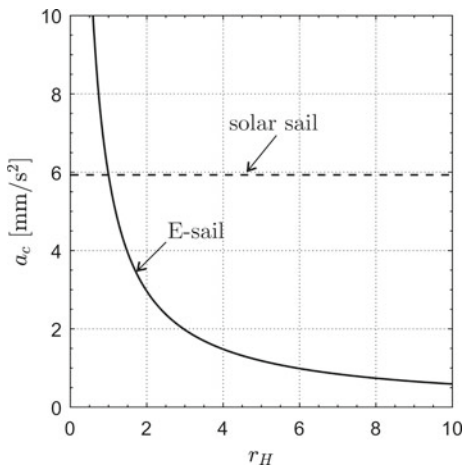
Fig. 1.14 Sketch of the heliostationary position maintenance mission scenario



Equation (1.42) implies that the required characteristic acceleration for a heliostationary position is $a_c \simeq 5.93 \text{ mm/s}^2$ (that is, $\beta = 1$), which is a very demanding technology requirement, as already stated. On the other hand, for an E-sail-propelled spacecraft, the propulsive requirement is a function of the levitating distance r_H , as is shown in Fig. 1.15. Not surprisingly, the solar sail option poses fewer demanding requirements only for $r < r_\oplus$. However, for both propulsive systems, it is evident that the heliostationary condition maintenance poses very demanding propulsive requirements.

As far as elliptic DNKO are concerned, the most interesting application is the possibility of observing the polar regions of a planet with non-negligible orbital eccentricity. In this regard, a PFDO could constitute a promising mission scenario. The simplest PFDO is obtained by assuming $\chi = 1$. This implies that the PFDO

Fig. 1.15 Heliostationary distance r_H as a function of a_c : solar sail case (dashed line) versus E-sail case (solid line)



is obtained with a simple vertical translation of the planet’s heliocentric orbit. A spacecraft tracking such a PFDO would be always placed above the celestial body’s vertical; see Fig. 1.16.

The cone angle constraint can be calculated from Eq. (1.36) by substituting $\chi = 1$. However, the values of ϕ derived with this substitution are always greater than 90 deg, far beyond the maximum allowable value ϕ_{\max} for both a solar sail and an E-sail, making PFDOs with $\chi = 1$ physically unfeasible with these two propulsive systems.

In order to generate feasible solutions, the constraint of PFDO could be relaxed, allowing χ to be smaller than one. This corresponds to the generation of a displaced orbit with the same eccentricity as the planetary orbit, but with a smaller semimajor axis, see Fig. 1.17. Such a mission scenario could guarantee the feasibility of the DNKO, requiring a lightness number variation of a few percentage points with respect

Fig. 1.16 Scheme of a PDFO with $\chi = 1$

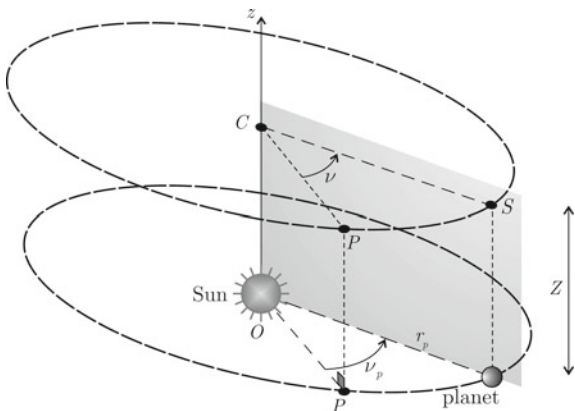
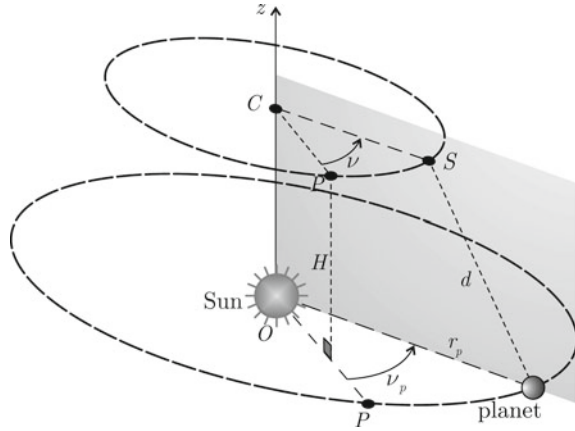


Fig. 1.17 Scheme of a PDFO with $\chi < 1$: the “shrunk” displaced orbit



to the nominal value (Gong et al. 2014a; Niccolai et al. 2017a), but it significantly increases the spacecraft–planet distance. Therefore, the mission performance of such a polar observation mission would not be satisfying.

1.2.4 Linear Stability Analysis

In order to investigate the feasibility of the DNKO concept, a linear stability analysis is now discussed. For the sake of simplicity, only circular DNKOs will be analyzed, even though the proposed method could be easily extended to an elliptic case with similar results.

The linear stability of circular DNKOs is studied under the assumption that the sail attitude coincides with its design value, so that both the cone angle ϕ and the parameter γ remain constant and equal to their nominal values ϕ_e and γ_e . Accordingly, during the flight the spacecraft is not subjected to any transverse acceleration, and the vertical component of the angular momentum h_z remains constant and equal to its nominal value $h_{z_e} \triangleq \rho_0^2 \omega_e$; see Eq. (1.26). The subscript e will now be used for the nominal equilibrium (unperturbed) value. The equations of motion (1.26) and (1.27) can be rewritten by assuming that both the radial and vertical spacecraft coordinate are given by the sum of their design value $\{\rho_e, z_e\}$ and a small perturbation term $\{\delta_\rho, \delta_z\}$, viz.

$$\rho = \rho_e + \delta_\rho, \quad z = z_e + \delta_z \quad (1.43)$$

with $\delta_\rho/\rho_e \ll 1$ and $\delta_z/z_e \ll 1$. Substituting Eq. (1.43) into Eqs. (1.26) and (1.27), neglecting the perturbations term of order greater than one, and subtracting the equilibrium solution, we obtain the dynamical equations describing the evolution of the perturbation components. If the temporal derivatives are transformed into derivatives with respect to a dimensionless time $\tilde{t} \triangleq \omega_e t$, the results after some calculations are

$$\delta''_{\rho} = a_{11}\delta_{\rho} + a_{12}\delta_z \quad (1.44)$$

$$\delta''_z = a_{21}\delta_{\rho} + a_{22}\delta_z \quad (1.45)$$

where the prime symbol denotes a derivative with respect to \tilde{t} . The coefficients a_{ij} are defined as

$$a_{11} \triangleq 3 \cos^3 \psi_e - 1 - 3(\omega_e^2/\omega_{K_e}^2) + f[\cos \phi_e - (1 + \eta) \cos \psi_e \cos(\phi_e + \psi_e)] \quad (1.46)$$

$$a_{12} \triangleq 3 \cos \psi_e \sin \psi_e - f[\sin \phi_e + (1 + \eta) \sin \psi_e \cos(\phi_e + \psi_e)] \quad (1.47)$$

$$a_{21} \triangleq 3 \cos \psi_e \sin \psi_e + f[\sin \phi_e - (1 + \eta) \cos \psi_e \sin(\phi_e + \psi_e)] \quad (1.48)$$

$$a_{22} \triangleq 3 \sin^2 \psi_e - 1 + f[\cos \phi_e - (1 + \eta) \sin \psi_e \sin(\phi_e + \psi_e)] \quad (1.49)$$

where f is given by

$$f = f(\psi_e, \omega_e^2/\omega_{K_e}^2) \triangleq \left(1 - \frac{\omega_e^2/\omega_{K_e}^2}{1 + \tan^2 \psi_e}\right) \sqrt{1 + \frac{\tan^2 \psi_e}{[(1 + \tan^2 \psi_e)/(\omega_e^2/\omega_{K_e}^2) - 1]^2}} \quad (1.50)$$

Note that all of the four coefficients a_{ij} depend on the pair $(\psi_e, \omega_e^2/\omega_{K_e}^2)$. Indeed, the nominal value of the cone angle ϕ_e is determined by the DNKO parameters through Eq. (1.24).

Using the Laplace transformation method, the characteristic equation associated with the system of differential Eqs. (1.44)–(1.45) can be written in the form

$$s^4 + bs^2 + c = 0 \quad (1.51)$$

where

$$b \triangleq -(a_{11} + a_{22}) \quad (1.52)$$

$$c \triangleq a_{12}a_{22} - a_{12}a_{21} \quad (1.53)$$

Since the characteristic equation is biquadratic, only a marginal linear stability is achievable. The latter requires all roots of Eq. (1.51) to be imaginary. Equivalently, the three conditions that guarantee a stable motion are $b > 0$, $c > 0$, and $b^2 - 4c > 0$. Clearly, such conditions must be combined with the propulsive system constraints, i.e., $\phi \leq \phi_{\max}$ and $f < 0$ (the sail cannot produce a thrust towards the Sun). When all

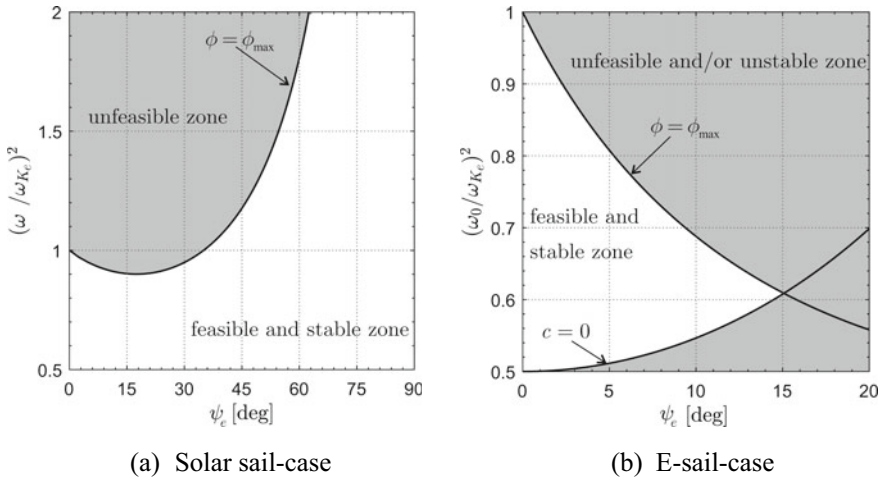


Fig. 1.18 Stability region of circular heliocentric DNKOs

the aforementioned conditions are combined, the resulting marginal stability regions for circular heliocentric DNKOs are shown in Fig. 1.18a for the solar sail, and in Fig. 1.18b for the E-sail case.

The stability conditions of practical interest are met for a solar sail-propelled spacecraft only, provided $\phi \leq \phi_{\max}$. Instead, for an E-sail-based spacecraft, the stability region is narrow and exists only for small elevation angles ($\psi_e < 20^\circ$). This suggests that a solar sail could constitute a better candidate for a future DNKO scientific mission.

1.3 Displaced Non-Keplerian Orbits in a Circular Restricted Three-Body Problem

Previous results have shown that one of the most promising utilizations of non-Keplerian trajectories is the generation of artificial orbits that are synchronous with the rotation of the observed planet. On the other hand, the generation of a DNKO with a vertical displacement above the ecliptic is very demanding in terms of the propulsive acceleration magnitude it requires, and equilibrium is possible only assuming large spacecraft–planet distances, so that the planetary gravity can be neglected.

The natural extension of the previous analysis is therefore obtained by including the planet’s gravity (thus moving from a two-body to a three-body heliocentric framework) in the mathematical model. With this in mind, a special application of the discussed model is the generation of an artificial equilibrium point when the non-Keplerian orbit belongs to the ecliptic (case of zero displacement). In particular,

L_1 -type equilibrium points will be considered, since they represent a vantage point for the scientific observation of the Sun, and for early solar warning missions. In this case, the spacecraft motion must be synchronous with the planetary revolution. This section is devoted to the analysis of such a special equilibrium condition in a three-body scenario. Section 1.3.1 discusses the mathematical model used to derive the requirements for the equilibrium maintenance. These results are then applied to some specific mission scenarios in Sect. 1.3.2. Finally, Sect. 1.3.3 is focused on the linear stability analysis of the equilibrium condition.

1.3.1 Mathematical Model

Assume now that the spacecraft S is moving under the gravitational forces exerted by the Sun and a planet. The analysis of the orbital motion of S can be simplified by means of two fundamental assumptions. First, the space craft total mass m_S is negligible with respect to the Sun's mass m_\odot and the planet's mass m_p , so that the motion of the two celestial bodies is unaffected by the presence of the vehicle. In the second place, the orbital eccentricity of the primaries is neglected, so that they are assumed to track two coplanar circular orbits around the center of mass of the system C , while maintaining a constant distance l . These assumptions define the classical circular restricted three-body problem (CR3BP) (Koon et al. 2011; Szebehely 1967).

In order to study the spacecraft motion, introduce a Cartesian reference frame $\tau(C; \hat{i}, \hat{j}, \hat{k})$. The unit vector \hat{i} points from the Sun to the planet, \hat{k} is perpendicular to the ecliptic in the direction of the planet's angular momentum vector, and \hat{j} completes the right-handed frame; see Fig. 1.19. Based on the coordinate axes definition, it can be verified that the reference frame τ rotates with a constant angular velocity

$\omega_p \triangleq \sqrt{G(m_p + m_\odot)/l^3}$ relative to an inertial frame.

Using the standard notation of a CR3BP, the total mass $m_\odot + m_p$ of the two primaries is taken as the reference mass, and the (constant) distance l is chosen as the reference length. Finally, the time t is expressed in dimensionless units by normalizing ω_p to 1. Accordingly, the planet dimensionless mass is $\mu \triangleq m_p/(m_\odot + m_p)$, while the dimensionless distance between C and the planet (or the Sun) is $1 - \mu$ (or μ); see Fig. 1.19. Bearing in mind that the angular velocity vector of the reference frame can be expressed as \hat{k} in dimensionless units, the motion of S is described by the following differential equation (Battin 1987)

$$\tilde{\mathbf{r}}'' + 2\hat{\mathbf{k}} \times \tilde{\mathbf{r}}' + \hat{\mathbf{k}} \times (\hat{\mathbf{k}} \times \tilde{\mathbf{r}}) + \frac{1 - \mu}{\tilde{r}_\Delta^3} \tilde{\mathbf{r}}_\Delta + \frac{\mu}{\tilde{r}_p^3} \tilde{\mathbf{r}}_p = \tilde{\mathbf{a}} \quad (1.54)$$

where the prime symbol denotes a derivative with respect to the dimensionless time ($t\omega_\oplus$), and the tilde superscripts are used to identify dimensionless quantities. In particular, $\tilde{\mathbf{r}}$, $\tilde{\mathbf{r}}_\odot$ and $\tilde{\mathbf{r}}_p$ are the dimensionless position vectors of S with respect to

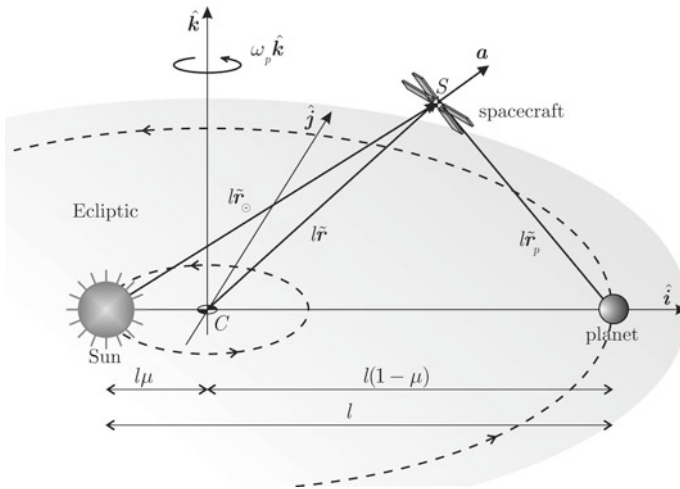


Fig. 1.19 Sketch of the CR3BP framework (Aliasi et al. 2011)

C , the Sun, and the planet, respectively, with $\tilde{\mathbf{r}}_\odot \triangleq \|\tilde{\mathbf{r}}_\odot\|$ and $\tilde{\mathbf{r}}_p \triangleq \|\tilde{\mathbf{r}}_p\|$. Finally, in Eq. (1.54), $\tilde{\mathbf{a}}$ denotes the dimensionless propulsive acceleration vector provided by the sail.

By geometrical considerations, it is possible to express the position vector and the planet–spacecraft vectorial distance as

$$\tilde{\mathbf{r}} = \tilde{\mathbf{r}}_\odot - \mu \hat{\mathbf{i}}, \tilde{\mathbf{r}}_p = \tilde{\mathbf{r}} - (1 - \mu) \hat{\mathbf{i}} = \tilde{\mathbf{r}}_\odot - \hat{\mathbf{i}} \quad (1.55)$$

which allows Eq. (1.54) to be rewritten as a function of a single dimensionless vector, viz.

$$\tilde{\mathbf{r}}_\odot'' + 2\hat{\mathbf{k}} \times \tilde{\mathbf{r}}_\odot' + \hat{\mathbf{k}} \times [\hat{\mathbf{k}} \times (\tilde{\mathbf{r}}_\odot - \mu \hat{\mathbf{i}})] + \frac{1-\mu}{\tilde{r}_\odot^3} \tilde{\mathbf{r}}_\odot + \frac{\mu}{\|\tilde{\mathbf{r}}_\odot - \hat{\mathbf{i}}\|^3} (\tilde{\mathbf{r}}_\odot - \hat{\mathbf{i}}) = \tilde{\mathbf{a}} \quad (1.56)$$

In order to simplify the notation and make the nomenclature consistent with that of the existing literature (Aliasi et al. 2011), the thrust models for the solar sail and the E-sail cases are expressed by means of the lightness number formulation. Accordingly, the dimensionless propulsive acceleration vector $\tilde{\mathbf{a}}$ generated by a solar sail can be written from Eq. (1.1) in dimensionless units as

$$\tilde{\mathbf{a}} = \beta \frac{1-\mu}{\tilde{r}_\odot^2} \frac{\hat{\mathbf{r}}_\odot \cdot \hat{\mathbf{n}}}{b_1 + b_2 + b_3} \{b_1 \hat{\mathbf{r}}_\odot + [b_2 (\hat{\mathbf{r}}_\odot \cdot \hat{\mathbf{n}}) + b_3] \hat{\mathbf{n}}\} \quad (1.57)$$

where $\hat{\mathbf{r}}_{\odot}$ is the Sun–spacecraft unit vector. Considering the E-sail case, the vector $\tilde{\mathbf{a}}$ can be expressed by means of an equivalent lightness number as

$$\tilde{\mathbf{a}} = \tau\beta \frac{1-\mu}{2\tilde{r}_{\odot}} [\hat{\mathbf{r}}_{\odot} + (\hat{\mathbf{r}}_{\odot} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] \quad (1.58)$$

where $\beta \triangleq a_c l^2 / (Gm_{\odot})$. When the propulsive system is chosen, Eq. (1.57) or (1.58) can be substituted in Eq. (1.56) to study the system dynamics.

An artificial equilibrium point (AEP) in a CR3BP is an equilibrium point in the rotating (synodic) reference frame τ . Such an AEP corresponds to a non-Keplerian orbit, when observed with respect to an inertial frame. The determination of AEPs in the CR3BP framework can be done with the aid of Eq. (1.56) by setting the derivatives of the Sun–spacecraft vector $\mathbf{e} \tilde{\mathbf{r}}_{\odot}$ equal to zero, i.e., $\tilde{\mathbf{r}}'_{\odot e} = \tilde{\mathbf{r}}''_{\odot e} = 0$, where the subscript e identifies a nominal (and unperturbed) condition.

1.3.2 Case Study

As already mentioned, the most interesting application of AEPs for solar observation is constituted of an L_1 -type AEP in the Sun–[Earth + Moon] CR3BP. In this case, the reference distance of the CR3BP is $l = r_{\oplus} = 1$ au, the angular velocity of the synodic frame is $\omega_{\oplus} = 2\pi$ rad/year, while $\mu = 3.0404 \times 10^{-6}$.

This case is illustrated in Fig. 1.20, which shows that the spacecraft is at an equilibrium position in the synodic reference frame, and lies on the Sun–Earth line at a (dimensionless) distance $\tilde{r}_{\odot 0} \in (0, 1)$ from the Sun, viz.

$$\tilde{\mathbf{r}}_{\odot e} = \tilde{r}_{\odot e} \hat{\mathbf{i}} \quad (1.59)$$

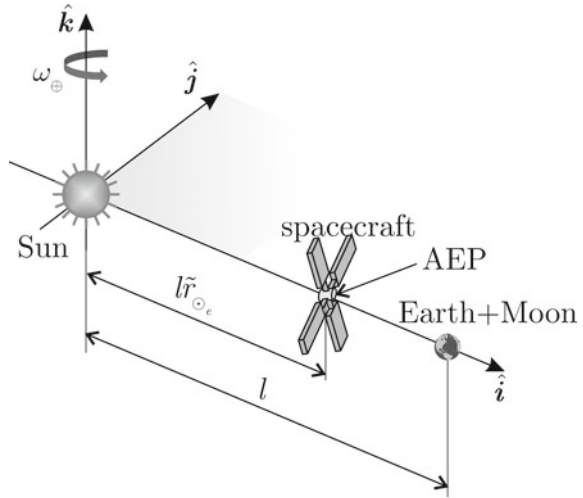
The maintenance of an L_1 -type AEP requires a constantly acting propulsive acceleration ($\tau = 1$ for the E-sail) directed along the $\hat{\mathbf{i}}$ -direction, viz.

$$\mathbf{a}_e = \beta_e \frac{1-\mu}{\tilde{r}_{\odot e}^{\eta}} \hat{\mathbf{i}} \quad (1.60)$$

so that the required thrust angle is constantly equal to zero, i.e., $\phi \equiv 0$. This condition can be obtained only with a Sun-facing attitude, i.e., $\alpha \equiv 0$. The required performance parameter of the sail can be given as a function of the desired Sun–spacecraft equilibrium distance as

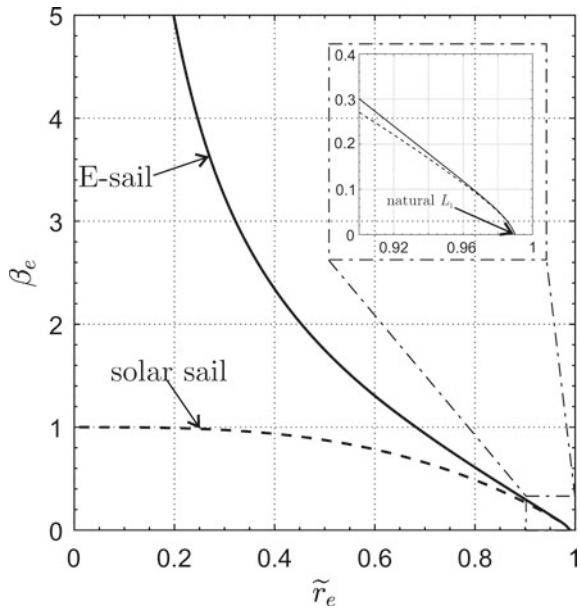
$$\beta_e = \frac{\mu \tilde{r}_{\odot e}^{\eta}}{1-\mu} \left[1 - \frac{1}{(1-\tilde{r}_{\odot e})^2} \right] + \tilde{r}_{\odot e}^{\eta-2} - \frac{\tilde{r}_{\odot e}^{\eta+1}}{1-\mu} \quad (1.61)$$

Fig. 1.20 Sketch of the L_1 -type AEP maintenance mission scenario



The results of Eq. (1.61) are summarized in Fig. 1.21. Note that realistic values of β_e can only maintain AEPs whose distance from the Earth is significantly smaller than 0.1 au. However, a spacecraft placed at an AEP with $e \tilde{r}_e = 0.980521$ could guarantee an early warning time of about 2 h in case of solar flares, thus doubling the performance of the ACE mission (Stone et al. 1998), which is currently orbiting around the (natural) equilibrium point L_1 .

Fig. 1.21 Required nominal lightness number as a function of Sun–AEP distance in the Sun–[Earth + Moon] CR3BP



1.3.3 Linear Stability Analysis

The analysis of the dynamical behavior of solar sail- or E-sail-based spacecraft in the vicinity of an L_1 -type AEP is now performed using the transformation

$$\tilde{\mathbf{r}} = \tilde{\mathbf{r}}_e + \delta\tilde{\mathbf{r}} \triangleq [\tilde{r}_{\odot e} - \mu + x, y, z]^T, \quad \dot{\tilde{\mathbf{r}}} = \delta\dot{\tilde{\mathbf{r}}} \triangleq [v_x, v_y, v_z]^T \quad (1.62)$$

where $\{x, y, z\} \ll 1$ and $\{v_x, v_y, v_z\} \ll 1$ are dimensionless perturbation terms. Introduce the state vector \mathbf{x} defined as

$$\mathbf{x} \triangleq [x, y, z, v_x, v_y, v_z]^T \quad (1.63)$$

of which the components are the position and velocity errors relative to the L_1 -type AEP; see Eq. (1.62). Substituting Eq. (1.62) into Eq. (1.54), subtracting the equilibrium solution (1.61) and neglecting the second-order perturbation terms, the spacecraft linearized dynamics may be written in a compact form as

$$\dot{\mathbf{x}} = \mathbb{A}\mathbf{x} \quad (1.64)$$

where

$$\mathbb{A} = \begin{bmatrix} \mathbb{O} & \mathbb{I} \\ \mathbb{C} & \mathbb{D} \end{bmatrix} \quad (1.65)$$

in which \mathbb{O} is a 3×3 zero matrix, \mathbb{I} is a 3×3 identity matrix, while matrices \mathbb{C} and \mathbb{D} are defined as

$$\mathbb{C} \triangleq \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix}, \quad \mathbb{D} \triangleq \begin{bmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (1.66)$$

with

$$c_1 \triangleq 1 + 2\frac{1-\mu}{\tilde{r}_{\odot e}^3} + 2\frac{\mu}{(1-\tilde{r}_{\odot e})^3} - \frac{\eta}{\tilde{r}_{\odot e}} \left[\frac{1-\mu}{\tilde{r}_{\odot e}} - \frac{\mu}{(1-\tilde{r}_{\odot e})^2} - \tilde{r}_{\odot e} + \mu \right] \quad (1.67)$$

$$c_2 \triangleq \frac{\mu}{\tilde{r}_{\odot e}} \left[1 - \frac{1}{(1-\tilde{r}_{\odot e})^3} \right] \quad (1.68)$$

$$c_3 \triangleq c_2 - 1 \quad (1.69)$$

The existence of one positive eigenvalue of matrix \mathbb{A} in Eq. (1.65) implies that the spacecraft (perturbed) motion around L_1 -type AEP is always unstable (Aliasi et al. 2012; Biggs et al. 2010). Therefore, to guarantee AEP maintenance, the spacecraft must be equipped with a suitable control system.

As is well known, the unstable direction of L_1 -type points is the radial direction (along the unit vector \hat{i}). Accordingly, a control system for AEP maintenance could generate a suitable variation of the lightness number β with respect to the nominal equilibrium value β_e , thus introducing a small variation $\delta\beta$ and adjusting the radial component of the propulsive acceleration. The lightness number variation could be achieved by exploiting electrochromic control devices (Aliasi et al. 2013b; Funase et al. 2011) for the solar sail case, or by adjusting the grid voltage for the E-sail case (Toivanen et al. 2013, 2017). Equation (1.64) may be written as

$$\dot{\mathbf{x}} = \mathbb{A}\mathbf{x} + \mathbb{B}\delta\beta \quad (1.70)$$

with

$$\mathbb{B} \triangleq \begin{bmatrix} 0 \\ 0 \\ 0 \\ (1 - \mu)/\tilde{r}_{\odot 0}^\eta \\ 0 \\ 0 \end{bmatrix} \quad (1.71)$$

Assume that a proportional-derivative control law is applied, viz.

$$\delta\beta = -\mathbb{K}\mathbf{x} \quad (1.72)$$

where \mathbb{K} is a row vector of gains. If only the radial direction is controlled, \mathbb{K} has the following structure

$$\mathbb{K} [k_1 \ 0 \ 0 \ k_2 \ 0 \ 0] \quad (1.73)$$

where k_1 (or k_2) is the gain relative to x (or v_x). When Eqs. (1.71), (1.72) and (1.73) are substituted into Eq. (1.70), it is found that the system stability depends on the eigenvalues of matrix \mathbb{C} , defined as

$$\mathbb{C} = \mathbb{A} - \mathbb{B}\mathbb{K} \quad (1.74)$$

which obviously depend on the propulsive system (through η) and the selected AEP (through $\tilde{r}_{\odot 0}$). This allows the control system designer to select values of k_1 and k_2 capable of guaranteeing stability. Detailed discussions on such a control strategy exist in the literature for the solar sail case (Niccolai et al. 2020b), for the E-sail case (Niccolai et al. 2020a), and also for a generalized sail (Aliasi et al. 2011).

1.4 Conclusions

The foregoing analysis has shown the potentiality of solar sails and electric solar wind sails as primary propulsive systems of a spacecraft tracking a heliocentric displaced non-Keplerian orbit. The performance requirements (in terms of propulsive acceleration magnitude) of orbits with significant displacements are very demanding, so that this advanced mission concept seems feasible only for a medium- or far-term technology level.

On the other hand, the generation of an artificial L_1 -type Lagrangian point in the Sun-[Earth + Moon] system constitutes a special case of non-Keplerian orbit with significantly lower performance requirements. Such a specific mission scenario would have a very interesting scientific application, since it would be ideal for a solar observation mission capable of providing an early warning in case of solar flares. In this case, the natural instability of such an equilibrium point could be counteracted by a suitable control system capable of adjusting the sail lightness number.

References

- Aliasi, G., G. Mengali, and A.A. Quarta. 2011. Artificial equilibrium points for a generalized sail in the circular restricted three-body problem. *Celestial Mechanics and Dynamical Astronomy* 110 (4): 343–368. <https://doi.org/10.1007/s10569-011-9366-y>.
- Aliasi, G., G. Mengali, and A.A. Quarta. 2012. Passive control feasibility of collinear equilibrium points with solar balloons. *Journal of Guidance, Control, and Dynamics* 35 (5): 1657–1661. <https://doi.org/10.2514/1.57393>.
- Aliasi, G., G. Mengali, and A.A. Quarta. 2013a. Artificial equilibrium points for an electric sail with constant attitude. *Journal of Spacecraft and Rockets* 50 (6): 1295–1298. <https://doi.org/10.2514/1.A32540>.
- Aliasi, G., G. Mengali, and A.A. Quarta. 2013b. Artificial lagrange points for solar sail with electrochromic material panels. *Journal of Guidance, Control, and Dynamics* 36 (5): 1544–1550. <https://doi.org/10.2514/1.58167>.
- Aliasi, G., G. Mengali, and A.A. Quarta. 2015. Artificial periodic orbits around L_1 -type equilibrium points for a generalized sail. *Journal of Guidance, Control, and Dynamics* 38 (9): 1847–1852. <https://doi.org/10.2514/1.G000904>.
- Ancona, E., and R.Y. Kezerashvili. 2017. Temperature restrictions for materials used in aerospace industry for the near-Sun orbits. *Acta Astronautica* 140: 565–569. <https://doi.org/10.1016/j.actastro.2017.09.002>.
- Baig, S. and C.R. McInnes. 2008. Artificial three-body equilibria for hybrid low-thrust propulsion. *Journal of Guidance, Control, and Dynamics* 31 (6): 1644–1655. <https://doi.org/10.2514/1.36125>.
- Bassetto, M. et al. 2018. Plasma brake approximate trajectory. Part II: relative motion. In: *4th IAA conference on university satellite missions and Cubesat workshop*, vol. 163, 249–259.
- Battin, R.H. 1987. An Introduction to the mathematics and methods of astrodynamics. *AIAA. Chap.* 8: 371–381.
- Betts, B., D.A. Spencer, J.M. Bellardo, et al. 2019. LightSail 2: controlled solar sail propulsion using a CubeSat. In: *70th international astronomical congress*. Washington (DC), USA.
- Biggs, J.D., and C.R. McInnes. 2010. Passive orbit control for space-based geo-engineering. *Journal of Guidance, Control and Dynamics* 33 (3): 1017–1020. <https://doi.org/10.2514/3.21211>.

- Bookless, J., and C.R. McInnes. 2006. Dynamics and control of displaced periodic orbits using solar-sail propulsion. *Journal of Guidance, Control, and Dynamics* 29 (3): 527–537. <https://doi.org/10.2514/1.15655>.
- Caruso, A., G. Mengali, and A.A. Quarta. 2019. Elliptic displaced or-bit approximation with equally spaced impulses. *Journal of Guidance, Control, and Dynamics* 42 (2): 411–415. <https://doi.org/10.2514/1.G003900>.
- Caruso, A. et al. 2020. Solar sail optimal control with solar irradiance fluctuations. In *Advances in space research*. In press. <https://doi.org/10.1016/j.asr.2020.05.037>.
- Ceriotti, M., C.R. McInnes, and B.L. Diedrich. 2011. The pole-sitter mission concept: an overview of recent developments and possible future applications. In: *62nd international astronomical congress*, vol. 3, 2543–2559. Cape Town, South Africa.
- Ceriotti, M., J. Heiligers, and C.R. McInnes. 2014. Trajectory and spacecraft design for a pole-sitter mission. *Journal of Spacecraft and Rockets* 51 (1): 311–326. <https://doi.org/10.2514/1.A32477>.
- Dachwald, B. 2004. Minimum transfer times for nonperfectly reflecting solar sailcraft. *Journal of Spacecraft and Rockets* 41 (4): 693–695. <https://doi.org/10.2514/1.6279>.
- Forward, R.L. 1984. Light-levitated geostationary cylindrical orbits using perforated light sails. *Journal of the Astronautical Sciences* 32: 221–226.
- Fu, B., E. Sperber, and F. Eke. 2016. Solar sail technology - A state of the art review. *Progress in Aerospace Sciences* 86: 1–19. <https://doi.org/10.1016/j.paerosci.2016.07.001>.
- Funase, R. et al. 2011. Fuel-free and oscillation-free attitude control of IKAROS solar sail spacecraft using reflectivity control device. In: *28th international symposium on space technology and science*. Okinawa, Japan.
- Funase, R. et al. 2012. IKAROS, a solar sail demonstrator and its application to Trojan asteroid exploration. In *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference*. Honolulu, HI, USA. <https://doi.org/10.2514/6.2012-1748>.
- Gong, S., and J. Li 2014a. Solar sail heliocentric elliptic displaced orbits. *Journal of Guidance, Control, and Dynamics* 37 (6): 2021–2025. <https://doi.org/10.2514/1.G000660>.
- Gong, S., and J. Li. 2014b. Spin-stabilized solar sail for displaced solar orbits. *Aerospace Science and Technology* 32 (1): 188–199. <https://doi.org/10.1016/j.ast.2013.10.002>.
- Heaton, A.F., and A.B. Artusio-Glimpse. 2015. An update to NASA reference sail thrust model. In *AIAA SPACE 2015 conference and exposition*. Pasadena, CA, USA. <https://doi.org/10.2514/6.2015-4506>.
- Heaton, A., N. Ahmad, and K. Miller. 2017. Near earth asteroid scout thrust and torque model. In *The 4th international symposium on solar sailing*, vol. 17055. Kyoto, Japan.
- Heiligers, J., and C.R. McInnes. 2014. New families of sun-centered non-Keplerian orbits over cylinders and spheres. *Celestial Mechanics and Dynamical Astronomy* 120 (2): 163–194. <https://doi.org/10.1007/s10569-014-9570-7>.
- Heiligers, J., M. Ceriotti, C.R. McInnes, et al. 2011. Displaced geostationary orbit design using hybrid sail propulsion. *Journal of Guidance, Control, and Dynamics* 34 (6): 1852–1866. <https://doi.org/10.2514/1.53807>.
- Huo, M., G. Mengali, and A.A. Quarta. 2018. Electric sail thrust model from a geometrical perspective. *Journal of Guidance, Control, and Dynamics* 41 (3): 734–740. <https://doi.org/10.2514/1.G003169>.
- Janhunen, P. 2004. Electric sail for spacecraft propulsion. *Journal of Propulsion and Power* 20 (4): 763–764. <https://doi.org/10.2514/1.8580>.
- Janhunen, P. 2010. Electrostatic plasma brake for deorbiting a satellite. *Journal of Propulsion and Power* 26 (2): 370–372. <https://doi.org/10.2514/1.47537>.
- Johnson, L. et al. 2011. NanoSail-D: a solar sail demonstration mission. *Acta Astronautica* 68 (5–6): 571–575. <https://doi.org/10.1016/j.actaastro.2010.02.008>.
- Kestilä, A., T. Tikka, P. Peitso, et al. 2013. Aalto-1 nanosatellite—technical description and mission objectives. *Geoscientific Instrumentation, Methods and Data Systems* 2: 121–130. <https://doi.org/10.5194/gi-2-121-2013>.

- Kezerashvili, R.Y. 2008. Solar sail interstellar travel: 1. Thickness of solar sail films. *Journal of the British Interplanetary Society* 61 (11): 430–439.
- Kezerashvili, R.Y. 2014. Advances in solar sailing. In M. Macdonald, ed. Springer Praxis. Chap. 3, 573–592. https://doi.org/10.1007/978-3-642-34907-2_36.
- Khurshid, O. et al. 2014. Accommodating the plasma brake experiment on-board the Aalto-1 satellite. *Proceedings of the Estonian Academy of Sciences* 63 (2S): 258–266. <https://doi.org/10.3176/proc.2014.2S.07>.
- Koon, W.S. et al. 2011. *Dynamical systems, the three-body problem and space mission design*. Marsden Books. ISBN: 978–0–615–24095–4.
- Lätt, S. et al. 2014. ESTCube-1 nanosatellite for electric solar wind sail in-orbit technology demonstration. *Proceedings of the Estonian Academy of Sciences* 63 (2S): 200–209. <https://doi.org/10.3176/proc.2014.2S.01>.
- Lücking, C.M., C. Colombo, and C.R. McInnes. 2012. Electrochromic orbit control for smart-dust devices. *Journal of Guidance, Control and Dynamics* 35 (5): 1548–1558. <https://doi.org/10.2514/1.55488>.
- Macdonald, M., R.J. McKay, M. Vasile, et al. 2011. Low-thrust-enabled highly-non-Keplerian orbits in support of future Mars exploration. *Journal of Guidance, Control, and Dynamics* 34(5): 1396–1411. <https://doi.org/10.2514/1.52602>.
- McInnes, C.R. 2011. Displaced non-Keplerian orbits using impulsive thrust. *Celestial Mechanics and Dynamical Astronomy* 110: 199–215. <https://doi.org/10.1007/s10569-011-9351-5>.
- McInnes, C.R. 1998. Dynamics, stability, and control of displaced non-Keplerian orbits. *Journal of Guidance, Control, and Dynamics* 21 (5): 799–805. <https://doi.org/10.2514/2.4309>.
- McInnes, C.R. 1999. *Solar sailing: technology, dynamics and mission applications*. Springer. ISBN: 978-1-85233-102-3. <https://doi.org/10.1007/978-1-4471-3992-8>.
- McInnes, C.R. 2003. Inverse solar sail trajectory problem. *Journal of Guidance, Control, and Dynamics* 26 (2): 369–371. <https://doi.org/10.2514/2.5057>.
- McKay, R.J., M. Macdonald, J.D. Biggs, et al. 2011. Survey of highly-non-Keplerian orbits with low-thrust propulsion. *Journal of Guidance, Control, and Dynamics* 34 (3): 645–666. <https://doi.org/10.2514/1.52133>.
- McNutt, L. et al. 2014. Near-earth asteroid (NEA) scout. In *AIAA SPACE 2014 conference and exposition*. Paper AIAA 2014–4435. San Diego (CA), 4–7 August. <https://doi.org/10.2514/6.2014-4435>.
- Mengali, G., and A.A. Quarta. 2007b. Tradeoff performance of hybrid low-thrust propulsion system. *Journal of Spacecraft and Rockets* 44 (6): 1263–1270. ISSN: 0022-4650. <https://doi.org/10.2514/1.30298>.
- Mengali, G., and A.A. Quarta. 2005. Optimal three-dimensional interplanetary rendezvous using nonideal solar sail. *Journal of Guidance, Control, and Dynamics* 28 (1): 173–177. ISSN: 0731–5090. <https://doi.org/10.2514/1.8325>.
- Mengali, G., and A.A. Quarta. 2007a. Optimal heliostationary missions of high-performance sailcraft. *Acta Astronautica* 60 (8–9): 676–683. ISSN: 0094–5765. <https://doi.org/10.1016/j.actastro.2006.07.018>.
- Mengali, G., and A.A. Quarta. 2007c. Trajectory design with hybrid low-thrust propulsion system. *Journal of Guidance, Control, and Dynamics* 30 (2): 419–426. <https://doi.org/10.2514/1.22433>.
- Mengali, G., and A.A. Quarta. 2009. Non-Keplerian orbits for electric sails. *Celestial Mechanics and Dynamical Astronomy* 105 (1): 179–195. <https://doi.org/10.1007/s10569-009-9200-y>.
- Mengali, G., and A.A. Quarta. 2016. Heliocentric trajectory analysis of sun-pointing smart dust with electrochromic control. *Advances in Space Research* 57 (4): 991–1001. <https://doi.org/10.1016/j.asr.2015.12.017>.
- Mengali, G., A.A. Quarta, and P. Janhunen. 2008. Electric sail performance analysis. *Journal of Spacecraft and Rockets* 45 (1): 122–129. ISSN: 0022–4650. <https://doi.org/10.2514/1.31769>.
- Mengali, G. et al. 2007a. Refined solar sail force model with mission application. *Journal of Guidance, Control, and Dynamics* 30 (2): 512–520. ISSN: 0731–5090. <https://doi.org/10.2514/1.24779>.

- Monk, P.M.S., R.J. Mortimer, and D.R. Rosseinsky. 2007b. *Electrochromism: fundamentals and applications*. Wiley-VCH.
- Mori, O., J. Mastumoto, T. Chujo, et al. 2019. Solar power sail mission of OKEANOS. *Astrodynamics* 4 (3): 233–248. <https://doi.org/10.1007/s42064-019-0067-8>.
- Mori, O. et al. 2010. Attitude control of IKAROS solar sail spacecraft and its flight results. In *61st international astronomical congress*. Paper IAC-10.C1.4.3. Prague, Czech Republic.
- Morimoto, M.Y., H. Yamakawa, and K. Uesugi. 2007. Artificial equilibrium points in the low-thrust restricted three-body problem. *Journal of Guidance, Control, and Dynamics* 30 (5): 1563–1568. <https://doi.org/10.2514/1.26771>.
- Niccolai, L., A.A. Quarta, and G. Mengali. 2017. Electric sail elliptic displaced orbits with advanced thrust model. *Acta Astronautica* 138: 503–511. <https://doi.org/10.1016/j.actaastro.2016.10.036>.
- Niccolai, L., A.A. Quarta, and G. Mengali. 2018. Electric sail-based displaced orbits with refined thrust model. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 232 (3): 423–432. <https://doi.org/10.1177/0954410016679195>.
- Niccolai, L. et al. 2017. Plasma brake approximate trajectory. Part I: geocentric motion. In *4th IAA conference on university satellite missions and Cubesat workshop*, vol. 163, 235–247.
- Niccolai, L. et al. 2019. Effects of optical parameter measurement uncertainties and solar irradiance fluctuations on solar sailing. In *Advances in Space Research*. In press. <https://doi.org/10.1016/j.asr.2019.11.037>.
- Niccolai, L. et al. 2020a. Artificial collinear lagrangian point maintenance with electric solar wind sail. In *IEEE Transactions on Aerospace and Electronic Systems*. In press. <https://doi.org/10.1109/TAES.2020.2990805>.
- Niccolai, L. et al. 2020b. Feedback control law of solar sail with variable surface reflectivity at Sun-Earth collinear equilibrium points. In *Aerospace Science and Technology* 106. <https://doi.org/10.1016/j.ast.2020.106144>.
- Nye, B., and E. Greeson. 2016. The Lightsail story, public outreach strategies & results. In *67th international astronomical congress*. Pasadena, CA, USA.
- Orsini, L., et al. 2018. Plasma brake model for preliminary mission analysis. *Acta Astronautica* 144: 297–304. <https://doi.org/10.1016/j.actaastro.2017.12.048>.
- Pan, X., et al. 2019. Nonlinear dynamics of displaced non-Keplerian orbits with low-thrust propulsion. *Communications in Nonlinear Science and Numerical Simulation* 66: 61–83. <https://doi.org/10.1016/j.cnsns.2018.06.005>.
- Pan, X. et al. 2020. Linearized relative motion and proximity control of E-sail-based displaced orbits. In *Aerospace Science and Technology*, vol. 99. <https://doi.org/10.1016/j.ast.2019.105574>.
- Pezent, J., R. Sood, and A. Heaton. 2019. High-fidelity contingency trajectory design and analysis for NASA's near-earth asteroid (NEA) Scout solar sail Mission. *Acta Astronautica* 159: 385–396. <https://doi.org/10.1016/j.actaastro.2019.03.050>.
- Quarta, A.A., G. Mengali, and L. Niccolai. 2020. Solar sail optimal transfer between heliostationary points. *Journal of Guidance, Control, and Dynamics* 43 (10): 1935–1942. <https://doi.org/10.2514/1.G005193>.
- Quarta, A.A., and G. Mengali. 2016. Minimum-time trajectories of electric sail with advanced thrust model. *Aerospace Science and Technology* 55: 419–430. <https://doi.org/10.1016/j.ast.2016.06.020>.
- Russell Lockett, T. et al. 2019. Near-Earth asteroid scout flight mission. *IEEE Aerospace and Electronic Systems Magazine* 35 (3): 20–29. <https://doi.org/10.1109/MAES.2019.2958729>.
- Simo, J. 2017. A comparative study of displaced non-Keplerian orbits with impulsive and continuous thrust. In: *27th AAS/AIAA space flight mechanics meeting*. San Antonio (TX), USA.
- Slavinskis, A. et al. 2015. ESTCube-1 in-orbit experience and lessons learned. *IEEE Aerospace and Electronic Systems Magazine* 30 (8): 12–22. <https://doi.org/10.1109/MAES.2015.150034>.
- Song, M., X. He, and D. He. 2016. Displaced orbits for solar sail equipped with reflectance control devices in Hill's restricted three-body problem with oblateness. *Astrophysics and Space Science* 361 (10). 10. 1007/s10509–016–2915–9.
- Stone, E.C. et al. 1998. The advanced composition explorer. *Space Science Reviews* 86 (1–4):1–22.

- Szebehely, V.G. 1967. Theory of orbits - The restricted problem of three bodies. Academic Press. ISBN: 978-0124124318.
- Toivanen, P.K., and P. Janhunen. 2013. Spin plane control and thrust vectoring of electric solar wind sail. *Journal of Propulsion and Power* 29 (1): 178–185. <https://doi.org/10.2514/1.B34330>.
- Toivanen, P.K., and P. Janhunen. 2017. Thrust vectoring of an electric solar wind sail with a realistic sail shape. *Acta Astronautica* 131: 145–151. <https://doi.org/10.1016/j.actaastro.2016.11.027>.
- Tsuda, Y. et al. 2011. Achievement of IKAROS - Japanese deep space solar sail demonstration mission. In *7th IAA symposium on realistic advanced scientific space*. Aosta, Italy. <https://doi.org/10.1016/j.actaastro.2012.03.032>.
- Vulpetti, G., C. Circi, and T. Pino. 2017. Coronal Mass ejection early-warning mission by solar-photon sailcraft. *Acta Astronautica* 140: 113–125. <https://doi.org/10.1016/j.actaastro.2017.07.042>.
- Vulpetti, G., L. Johnson, and G.L. Matloff. 2015. *Solar sails: a novel approach to interplanetary travel*. Chichester, UK: Springer-Praxis. ISBN: 978-1-4939-0940-7. <https://doi.org/10.1007/978-1-4939-0941-4>.
- Vulpetti, G. 2013. *Fast solar sailing: astrodynamics of special sailcraft trajectories*, Chap. 6, 165–254. Springer. ISBN: 978-94-007-4776-0. <https://doi.org/10.1007/978-94-007-4777-7>.
- Wright, J.L. 1992. *Space sailing*. Gordon and Breach, ed., 223–233 Philadelphia, PA: Gordon and Breach Science Publishers.
- Zola, D. et al. 2018. Photon momentum change of quasi-smooth solar sails. *Journal of the Optical Society of America A: Optics and Image Science, and Vision* 35 (8): 1261–1271. <https://doi.org/10.1364/JOSAA.35.001261>.

Chapter 2

Dynamics and Control of Electrostatic Flight



Marco B. Quadrelli, Michele Bechini, Joseph Wang, and Shota Kikuchi

Abstract We describe the principles of electrostatic flight in tenuous plasma around solar system bodies. The lack of an atmosphere, low gravity levels, and unknown surface soil properties pose a very difficult challenge for all forms of known locomotion at airless bodies. The environment near the surface of asteroids, comets, and the Moon is electrically charged due to the Sun's photoelectric bombardment and lofting dust, which follows the Sun's illumination as the body spins. If a body with high surface resistivity is exposed to solar wind and solar radiation, Sun-exposed areas and shadowed areas become differentially charged. Our work in this field is motivated by the E-Glider, i.e., the Electrostatic Glider, which provides an enabling capability for practical electrostatic flight at airless bodies, a solution applicable to many types of in situ missions, which leverages the natural environment. The E-Glider is a small spacecraft that uses, instead of avoids, the charged environment in the solar system for near fuel-less circumnavigation, allows in situ characterization of the plasmasphere of planetary bodies, and reduces the risk of landing on hazardous surfaces.

M. B. Quadrelli (✉)

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA

e-mail: marco.b.quadrelli@jpl.nasa.gov

M. Bechini

Department of Aerospace Science and Technology (DAER), Politecnico di Milano, Via La Masa 34, 20156 Milano, Italy

e-mail: michele.bechini@polimi.it

J. Wang

Department of Astronautical Engineering, University of Southern California, 3650 S. McClintock Avenue, Los Angeles, CA 90089, USA

e-mail: josephjw@usc.edu

S. Kikuchi

Japan Aerospace Exploration Agency (JAXA), Kanagawa, Japan

e-mail: kikuchi.shota@perc.it-chiba.ac.jp

Planetary Exploration Research Center, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,

https://doi.org/10.1007/978-3-030-97913-3_2

2.1 Introduction

Small airless bodies in the solar system (small asteroids and comets) represent the next frontier in deep space exploration. Recent studies have demonstrated the important role played by small airless bodies in the origin and history of the solar system (Quadrelli et al. 2017b). Understanding small airless bodies also contributes directly to research addressing the characteristics of the solar system that led to the origin of life. The National Research Council has designated technologies for exploring small bodies as a high priority for NASA because of their destination potential for both scientific discovery and human spaceflight (which would also likely require precursor robotic missions) (Council 2011).

Currently, our knowledge of small bodies is mostly obtained from remote sensing. While remote sensing is very useful in providing information from a distance, an in-depth knowledge requires proximity and in situ measurements. Small satellites (SmallSat) (including CubeSats and nanosatellites) can enable a wide range of proximity missions around small bodies.

In the last few years, nanosatellites have found numerous interesting applications in commercial and scientific missions. Their role has switched from the old concept of a cheap and highly reliable small technology demonstrator to one of the most adopted technologies in space application, especially for Earth observation. The novelty in the application of nanosatellites lies in interplanetary missions and high-level scientific missions. To make these types of missions feasible for a nanosatellite, a “push” in the development of new advanced technologies is required. In particular, the main fields in which the biggest effort should be employed are propulsion (state-of-the-art solutions are chemical and electric thrusters (Páscoa et al. 2018), but two promising technologies are solar sails and field-emission electric propulsors), communication (low-power deep space systems), and navigation and control (autonomous navigation and high-accuracy pointing). If properly developed, these technologies can lead to the beginning of a new era of space exploration based on less expensive but more versatile spacecraft with new operational capabilities.

However, mobility around small bodies is highly challenging (Quadrelli et al. 2017b). Gravitational acceleration produced by small bodies is very small, typically on the order of the milli-G order of magnitude (Quadrelli et al. 2017b). The shape of asteroids/comets is typically extremely irregular, and the mass distribution is typically nonuniform. Hence, the gravity field around small bodies is typically highly complex (Scheeres 1994; Scheeres et al. 2006).

The National Research Council (USA) stated that the development of new technologies for small-body mobility should be of high priority for NASA (Council 2011). Moreover, recent observations have demonstrated the relevance of small bodies from an astrobiological point of view (Quadrelli et al. 2017b), making their exploration extremely intriguing. In situ analysis of small bodies like asteroids and comets is limited by the knowledge of the surface terrain since all the current robotics and human systems rely on the interaction between the system itself and the main body surface. Several studies have revealed that the surface of small bodies can show

extremely different landscapes composed of a thick layer of fine regolith in some cases or by big boulders in other cases (Han 2015; Scheeres et al. 2006). Recently, some advanced vehicle concepts capable of operating in extreme conditions like the ones depicted above have been proposed, for example, the DuAxel vehicle (Nesnas et al. 2012) which will be capable of operating in extremely challenging surface conditions, or the recently assembled small helicopter capable of flying in Mars' atmosphere (Withrow-Maser et al. 2020), avoiding the problems related to the surface conformation. Despite the extremely fascinating capabilities of these vehicles, they are still not suitable for applications on a small airless body due to the extremely challenging conditions for mobility.

The environment presented by small bodies is extremely challenging (Quadrelli et al. 2017b). Due to the extremely irregular shape presented by some asteroids, the gravity field can be highly irregular, especially on the surfaces (Scheeres 1994; Scheeres et al. 2006). The combined effects of the irregularity of the gravity field and the low intensity of the gravitational acceleration produced by these bodies (milli-G order of magnitude Quadrelli et al. 2017b) make the environment highly perturbed (Scheeres 2012). The effects of solar radiation pressure (SRP) have been proven to have a strong impact on vehicle dynamics (Scheeres 1994, 1999). As a result of this highly perturbed environment, escape velocities from these bodies are particularly low (Scheeres and Marzari 2002; Scheeres 2007). This must be carefully taken into account both for landers and orbiters, and for these reasons, vehicles that operate in micro-gravity are different from planetary vehicles, requiring special precautions in the design phase. Mobility in this environment is currently achieved by using hoppers, grippers, and hybrid systems (Seeni et al. 2010; Quadrelli et al. 2012). No other solutions are present at this time (Quadrelli et al. 2017b). Another promising idea is to take advantage of the environment near the airless body, developing the vehicle named Electrostatic Glider (E-Glider) capable of exploiting the naturally charged particle environment near the surface to produce lift (Quadrelli et al. 2017b).

As will be discussed later, for small satellites, the effects from solar radiation pressure and electrostatic interactions between a charged airless body and a charged spacecraft can become comparable to that of a gravitation field, and can thus have a strong impact on the vehicle dynamics (Scheeres 1994, 1999; Cui and Wang 2019). The combined effects from a small and irregular gravitational field and the perturbations from solar radiation pressure and Coulomb force make the dynamic environment around small bodies highly complex (Scheeres 2012). As a result, vehicles that operate in such a highly perturbed micro-gravity environment need to satisfy a set of mobility requirements different from standard planetary spacecraft. Various propulsion options and mobility concepts have been previously proposed for small-body missions, such as solar sails, electro-spray thrusters, hoppers, grippers, and hybrid systems (Seeni et al. 2010; Quadrelli et al. 2012).

The E-Glider concept, developed at NASA's Jet Propulsion Laboratory (Quadrelli et al. 2017b), is a new technology that can offer significantly more advantageous propulsion and navigation capabilities for proximity operations around small airless bodies using SmallSats.

This concept makes it possible to have a “closer look” at the asteroid surface without touching it and thus avoiding all the problems related to interacting with an extremely uncertain surface.

The E-Glider concept utilizes the electrostatic force between spacecraft and the naturally charged environment in proximity of an airless body for mobility. Without a global magnetic field and an atmosphere, small airless bodies are directly exposed to solar radiation and space plasma and thus are electrically charged by the ambient plasma and the emissions of photoelectrons and secondary electrons. A spacecraft around a small airless body is also electrically charged by the same process. By manipulating the charging state of a “glider” and thus the Coulomb force, an E-Glider may achieve complex orbital maneuvers beyond the capabilities of other mobility options.

We note that utilizing the electrostatic interaction between charged bodies had been considered for possible space applications in recent years, such as for docking, formation flying, collision avoidance, and attitude control (Quadrelli et al. 2017b; Schaub et al. 2004; Aslanov and Schaub 2019; King et al. 2002). However, these applications differ from the E-Glider concept because all of them rely on the Coulomb forces artificially generated between two (or more) spacecraft, while the E-Glider interacts directly with the electrostatic field around small bodies.

The E-Glider vehicle concept is bioinspired by small spiders (named *gossamer or ballooning spiders*) that produce charged threads that are mutually repelled (creating a sort of “hot-air balloon” made by thin threads) due to the presence of an electric charge, which also interacts with the Earth’s static atmospheric electric field, generating a lift component on the spider itself (Morley and Robert 2018). This *ballooning* is effective also in absence of convection or aerodynamics effects. Thus, behaving like a gossamer spider, the E-Glider transforms the problem of the spacecraft charging into an advantage being capable of orbiting and maneuvering due to electrostatic interactions with the environment (see Fig. 2.1).

This paper proposes the two distinct types of operations for an E-glider, namely, electrostatic hovering and electrostatic orbiting. The basic strategy of electrostatic hovering is to create artificial equilibrium points by inducing repulsive electrostatic force. These artificial equilibrium points are present not only on the nightside but also on the dayside, unlike natural equilibrium points, as illustrated in Fig. 2.2. Therefore, the proposed method can potentially achieve fuel-free hovering on the dayside without experiencing an eclipse. On the other hand, the utilization of electrostatic force offers advantages for orbiting operations as well. This paper identifies a new class of periodic orbits around asteroids using electrostatic force, which is called *electrostatic periodic orbits*. In contrast to the natural terminator orbits, these orbits are displaced from the terminator plane in the direction of the Sun, as depicted in Fig. 2.2, enabling the observation of the sunlit side of an asteroid. Besides, the electrostatic periodic orbits are Sun-synchronous, thereby ensuring constant illumination from the Sun. Another advantage of electrostatic orbiting is that it only requires a small amount of power; for example, some electrostatic periodic orbits consume only a few watts of electricity.

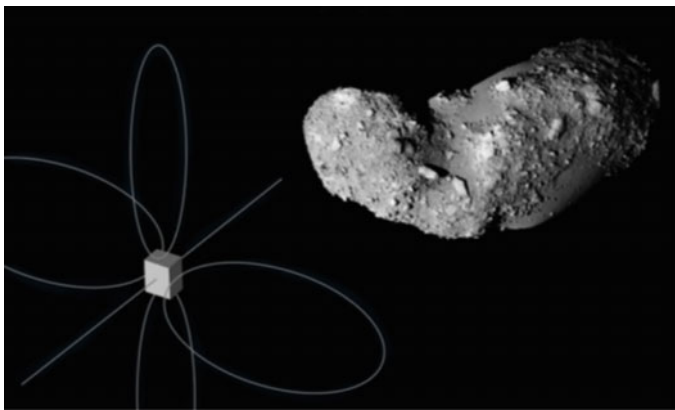
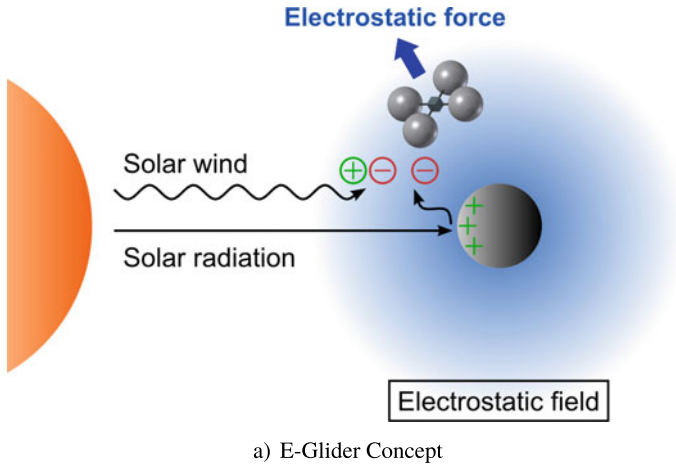


Fig. 2.1 E-Glider concept (a) and artistic concept (b) of an E-Glider approaching 25143 Itokawa (Corradino 2018)

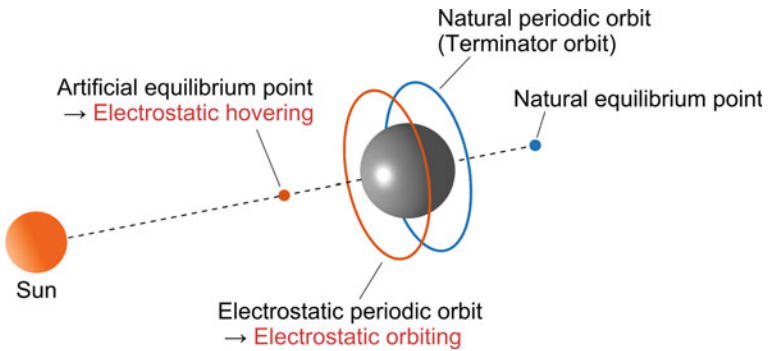


Fig. 2.2 Electrostatic hovering and electrostatic orbiting

As mentioned above, electrostatic hovering and electrostatic orbiting methods using an E-glider allow dayside operation without requiring any fuel. Therefore, the proposed methods are advantageous for mass budget, optical observation, solar power generation, and thermal design. By virtue of these characteristics, the E-Glider enables asteroid missions with lower cost and higher scientific value. For instance, an E-Glider could serve as a daughter spacecraft (secondary spacecraft deployed by a bigger spacecraft that acts as the mother spacecraft) for close observation of an asteroid. In addition to these practical advantages, this study is also intriguing in that completely new aspects of astrodynamics are revealed.

Many science objectives can be addressed by the E-Glider at small bodies, such as determining surface mechanical properties, searching for in situ resources, and understanding and simulating human activities in a low-gravity environment, among many others. Thanks to recent advances in miniaturization, several science-grade instruments are becoming available for implementation on small vehicles such as CubeSats. Some of these instruments which could be suitable for use on the E-Glider are (Kobrick et al. 2014) quadrupole ion trap spectrometers (2.5 kg, with isotopic accuracy $< 1\%$), snow and water imaging spectrometers (with high throughput, low polarization, high uniformity, in the 350–1700-nm spectral range), advanced infrared photodetectors (thermal sensitivity of 0.2°), high-resolution visible cameras (used for science, optical navigation, and autonomous navigation demonstration), and micro-seismometers.

To alter the charge level of the electrodes of the E-Glider, the methods described in Quadrelli et al. (2017a) have been considered. A “classic” charge ejection system that works by emitting beams of electrons (Evlanov et al. 2013) or ions (Masek and Cohen 1978) could be used to generate and control surface charging (Lai 1989), but it must be noticed that the emission of only positive ions leads to negative potential of the spacecraft of the order of kV (Lai 1989; Masek and Cohen 1978). Because of the potential bias due to the differential charging caused by the returning particles and to the uncertainties in the definition of the return current patterns (Quadrelli et al. 2017a), charge control systems based on monoenergetic beams are not commonly used. Moreover, for the E-Glider concept, the system for generating ion beams can be bulky. Another possible method is the use of electron field emission devices (Iwata et al. 2012; Khan et al. 2013). Currently, these devices are limited to only electron emission, and they need to be hard-mounted and coupled to the surfaces to be charged. These devices can be used to imitate photoelectron emission (Quadrelli et al. 2017a) due to the low potential reached. Proper selection of the surface material could act as another simple way to generate differential charges on surfaces. Several studies on the electrical properties of the materials are available in the literature (e.g., Plis et al. 2018; Czepiela et al. 2000; Mizera 1983). This method is not well suited for the E-Glider concept, even if this is a simple and passive method, because the differential charge is strongly dependent on the environmental conditions, which can be highly variable in the scenarios that will be explored by the E-Glider. In conclusion, the last method considered to induce different potentials on different surfaces is the employment of direct biasing devices like batteries and solar cells or small Van De Graaff generators (Peck 2005). A small Van De Graaff generator seems

to be the most promising and reliable charge source to be applied to the E-Glider concept due to the possibility of generating high surface potentials (Quadrelli et al. 2017a), even if drawbacks like the presence of moving parts or the necessity of being always powered by batteries or solar cells are still present. These methods must be capable of reacting in a short time to a fast-changing external environment that can be sensed by using classic Langmuir probes mounted onboard and that could act also as electrodes to control the E-Glider dynamics.

This paper presents a unique dynamical framework behind the complex environment around an asteroid involving the interaction between irregular gravitational force, SRP force, and electrostatic force. Consequently, this research expands the possibility of flight mechanics in space. This paper concludes that electrostatic flight using an E-Glider is useful for asteroid missions and exhibits unique and valuable dynamic characteristics. Figure 2.3 shows the elements of the concept of operations for modeling and simulation of E-Glider operations: (a) the relevant dust and charge environment is modeled with high-fidelity physics codes; (b) the coupled orbital and attitude dynamics can now be modeled in this environment; (c) the local plasma conditions, combined with the E-Glider dynamics and local charge levels can now enable the process of electrostatic inflation; (d) electrostatic maneuvering is now possible; (e) circumnavigation and small-body sampling is enabled based on electrostatic hovering and orbiting; and (f) leading to new airless body science that was not possible before.

The chapter's layout is organized as follows. In Sect. 2.2, we discuss the challenges presented by the environment near an airless body. Then in the Sects. 2.3–2.7, a description of the kinematics and kinetics of electrostatic flight is presented. We consider a system composed by a small airless asteroid and an E-Glider. These sections include the derivation of the equations of motion and of attitude dynamics. The gravitational and the SRP effects equations are derived for an extended spacecraft, and the gravitational field is defined both for a spherical body and for an ellipsoidal one. The core of the dynamical modeling is the spacecraft–plasma electrostatic interaction model and the model used to define the plasma around an airless body. These topics are discussed in Sect. 2.8, where both an analytical and a numerical model of the electric field are presented, and in Sect. 2.9, where the equations for the electrostatic effects are derived. The definition and the analysis of a new class of orbits named *electrostatic periodic orbits* are carried out in Sect. 2.10 by using the Nitter model for the electrostatic field definition. In the same section, the effects of an ellipsoidal asteroid on the E-Glider orbital dynamics are evaluated. By using the same model for the electrostatic field and for the spacecraft, the attitude stability of a double-dipole spacecraft is addressed in Sect. 2.11. These two analyses are merged to investigate the coupled orbital attitude stability in Sect. 2.12. The other mode to conduct the dayside operation of the E-Glider, the fixed hovering with respect to the Sun, is investigated for a point-charge spacecraft in Sect. 2.13 by using the particle-in-cell (PIC)-described plasma field. In Sect. 2.14, a preliminary control law to switch a single-dipole and a double-dipole E-Glider from a hovering condition to another is defined and investigated, by including also a preliminary electrostatic attitude control. In regards to the power required and the potential reached by the electrodes,

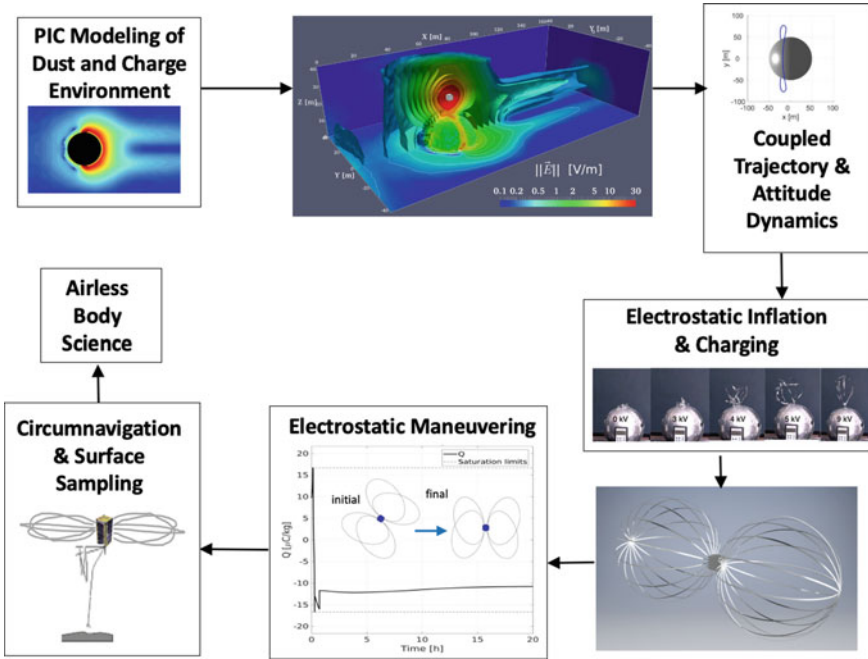


Fig. 2.3 Block diagram of the E-Glider concept of operations

Table 2.1 E-Glider concept models

S/C model	Electrodes geometry model	Plasma model	Sections
Point-charge S/C	Sphere	PIC	Section 2.13
Single-dipole S/C	Sphere	PIC	Section 2.14
Double-dipole S/C	Sphere	Nitter	Sections 2.10, 2.11, 2.12, 2.15
Double-dipole S/C	Sphere	PIC	Sections 2.14, 2.15
Double-dipole S/C	Wires (hoops)	PIC	Section 2.15

several analyses are conducted considering both the Nitter and the PIC plasma model for both the orbital and the hovering case for several electrodes geometries to identify the most promising one, and by comparing the results obtained and reported in Sect. 2.15. In conclusion, the main outcomes of this research are summarized in in Sect. 2.16 with possible future developments on the E-Glider concept.

To have an easier overview of the spacecraft models described in this work, they are summarized in Table 2.1 with the sections in which they are discussed.

2.2 Environmental Challenges at Small Bodies

The physics at airless bodies is dominated by four physical fields (Kobrick et al. 2014; Quadrelli et al. 2015): (a) microgravity, responsible for locomotion; (b) cohesion forces, which can dominate particle interactions through van der Waals forces; (c) solar radiation, which is constantly acting; and (d) electrostatics, which is strongest at the terminator where it can lead to significant dust transport. The highly irregular shapes of many asteroids and other small bodies lead to unique modeling and dynamics challenges. In contrast to the gravitational fields of spherical and ellipsoidal bodies, those produced by near-earth objects (NEOs) are frequently much more complex. The gravitational fields of these irregular bodies exhibit high levels of variation at both the surface and locations near the bodies. These gravitational fields are often orders of magnitude weaker than the Earth's. In addition to exhibiting irregular shapes, the gravitational fields produced by small bodies often have milli-G or micro-G orders of magnitude. As a result, escape velocities from these bodies are exceptionally low and must be carefully considered when maneuvering landers or spacecraft. Another consequence of these low gravitational magnitudes is that the rotational period, sometimes as fast as a fraction of a minute, may impact the motion of the spacecraft's motion. It may be possible to take advantage of this behavior to aid in motion between surface locations on a small body. This could potentially be achieved by applying an impulse to the lander such that it hops away from the surface without an orbital velocity component while the small body continues to rotate. This maneuver would lead to a change in position when gravity pulls the lander back to the surface. As the topics examined illustrate, it is necessary to understand the impacts of both small gravitational magnitudes and irregular gravitational field shapes to ensure successful spacecraft interactions with small bodies. The environment near the surface of airless bodies (asteroids, comets, moons) is electrically charged due to interactions with solar wind plasma and ultraviolet (UV) radiation. Charged dust is ever present, in the form of dusty plasma (Vladimirov 2005). Comets have a gas tail and a second electrostatic tail. This environment is also largely unexplored. Electrostatically levitating dust grains have been hypothesized to exist above tens of meters above the dayside surface (Hartzell 2012; Lee 1996). If a body with high surface resistivity is exposed to the solar wind and solar radiation, sun-exposed areas and shadowed areas become differentially charged. Charging on the dayside surface is dominated by photoelectrons emitted due to solar UV radiation that create a positive surface potential, while the shadowed side accumulates electrons and acquires a negative surface potential. Recent work Renno and Kok (2008), Stubbs et al. (2006) shows that on the Moon, soft solar X-rays with wavelengths smaller than 25 Angstroms can remove electrons with energies of 500 to 1500 eV from the surface and create cm-scale electric fields which may reach levels of 50–150 kV/m. The spokes in Saturn's rings are most likely clouds of particles electrostatically levitated from the surfaces of larger bodies in the rings, and electrostatic dust transport processes have been proposed on the surface of Mercury (Ip 1986) and comets (Mendis et al. 1981). Asteroid electric charge has never been measured, but simple estimates

predict that an electric potential difference of 1 kV can be attained on the dark side compared to the sunlit side, which becomes slightly positively charged by photoelectron emission. These differences are enhanced further at the terminator (the day/night boundary) when fields could reach 100–300 kV/m (Aplin et al. 2011) (with results obtained by simulation). Millimeter-size particles can be most easily lifted from the surface of Itokawa (Hartzell 2012). As these particles are lifted, they dislodge smaller particles that are harder to lift due to their strong cohesive forces. Once separated from the surface, grains can either travel on ballistic trajectories, escape from the asteroid, or levitate. During these migrations, the larger particles can get trapped in topographic lows, as observed in Miyamoto (2007). As a surface element on a resistive asteroid rotates in and out of view of the Sun, electrostatic levitation may agitate its uppermost particulate layer. Larger levitated particles remaining gravitationally bound to the asteroid are redistributed across its surface following local electrostatic and gravity gradients. Consequently, the study of levitating dust is relevant in that it provides some insight into the plasma environment and confirms the possibility of levitation. An intriguing example from nature discussed in Gorham (2013) refers to existing observations and the physics of spider silk in the presence of the Earth’s static atmospheric electric field (–120 V/m negative) to indicate a potentially important role for electrostatic forces in the flight of gossamer spiders. A compelling example is analyzed in detail, motivated by the observed “unaccountable rapidity” in the launching of such spiders from the vessel H.M.S. Beagle, recorded by Charles Darwin during his famous voyage, on a day without wind, and far away from the shore. It is believed that such spiders can emit threads that are either preloaded with a static electric charge so that the presence of this charge will lead both to mutual repulsion among the emitted threads, and an additional overall induced electrostatic force on the spider, providing a component of lift that is independent of convection or aerodynamic effects. The E-Glider biomorphically behaves like one of these spiders, greatly favored by the charged environment, in absence of aerodynamics and convection, and in the microgravity fields at small bodies.

2.3 Kinematics

The definition of the notation and the reference frames used is necessary to correctly understand the dynamic models and the equations of motion explained in the following paragraphs. The notation used to express the vectors and the matrices in the different reference frames is the following:

- ${}^a\mathbf{x}$ stands for a vector or a tensor \mathbf{x} expressed in the a -frame
- ${}^a\omega_{bc}$ stands for the angular velocity of c frame with respect to b frame expressed in the a frame
- ${}^a\mathbf{R}_b$ stands for the rotation matrix (or tensor) \mathbf{R} which converts ${}^b\mathbf{x}$ into ${}^a\mathbf{x}$, thus ${}^a\mathbf{x} = {}^a\mathbf{R}_b {}^b\mathbf{x}$

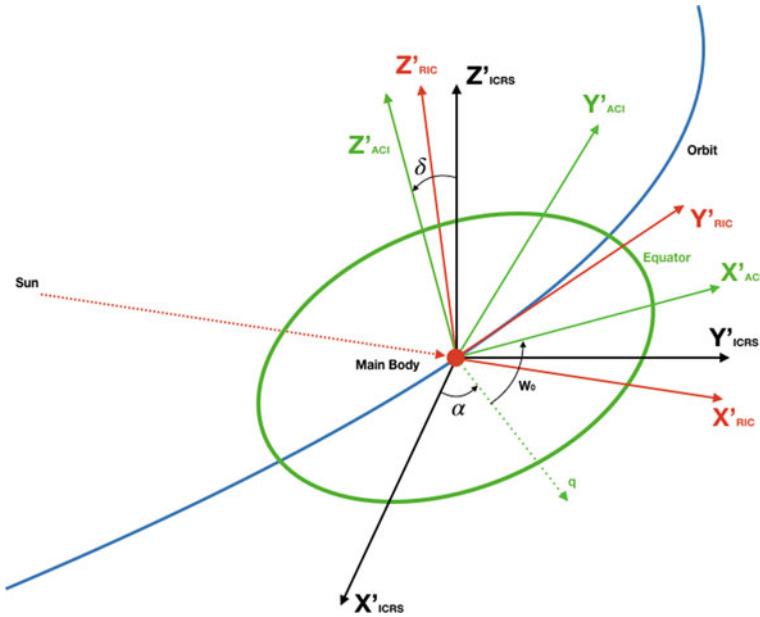


Fig. 2.4 Main body-centered reference frames: asteroid-centered inertial (ACI) and radial/in-track/cross-track (RIC)

2.3.1 Main Body-Centered Reference Frames

- Radial/in-track/cross-track (RIC) frame: r

The radial/in-track/cross-track (RIC) reference frame is the reference frame used to write the translation equations of motion. This is a non-inertial reference frame.

The RIC frame is defined as (see Fig. 2.4):

- O = origin at the main body center of mass
- X = axis directed away from the solar system barycenter (e.g., along the radial direction)
- Y = axis lying on the orbital plane and completing the right-handed orthogonal frame (e.g., along the in-track direction)
- Z = axis parallel to the orbital angular momentum vector (e.g., along the cross-track direction)

The RIC can be derived from the perifocal reference frame by translating it from the barycenter of the solar system to the main body center of mass and then by applying a rotation equal to the true anomaly of the main body.

$${}^r\mathbf{R}_p = [\Theta(t)]_3 \quad (2.1)$$

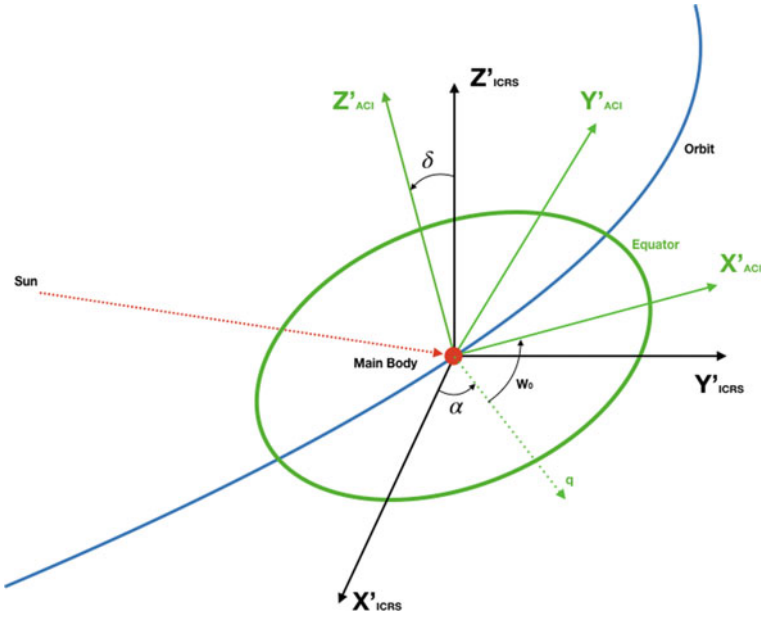


Fig. 2.5 ACI reference frame

- Asteroid-centered inertial (ACI) frame: a
The Asteroid-Centered Inertial (ACI) reference frame helps to define the attitude and the rotations of both the spacecraft and the Main Body. This reference frame can be assumed to be inertial when the attitude refers to it (the non-inertial components are due to translation).

This reference frame can be defined as (see Fig. 2.5):

- O = origin at the main body center of mass
- X = axis lying on the equator and pointing towards the prime meridian at the reference epoch
- Y = axis lying on the equator and completing the right-handed orthogonal frame
- Z = axis directed as the main body rotation angular momentum vector

The ACI reference frame can be derived by translating the international celestial reference frame on the main body and then rotating it as

$${}^a\mathbf{R}_i = [W_0]_3[\delta]_1[\alpha]_3 \tag{2.2}$$

where W_0 is the position of the prime meridian at a given epoch, δ is the declination of the positive pole, and α is the right ascension of the positive pole (see Fig. 2.5).

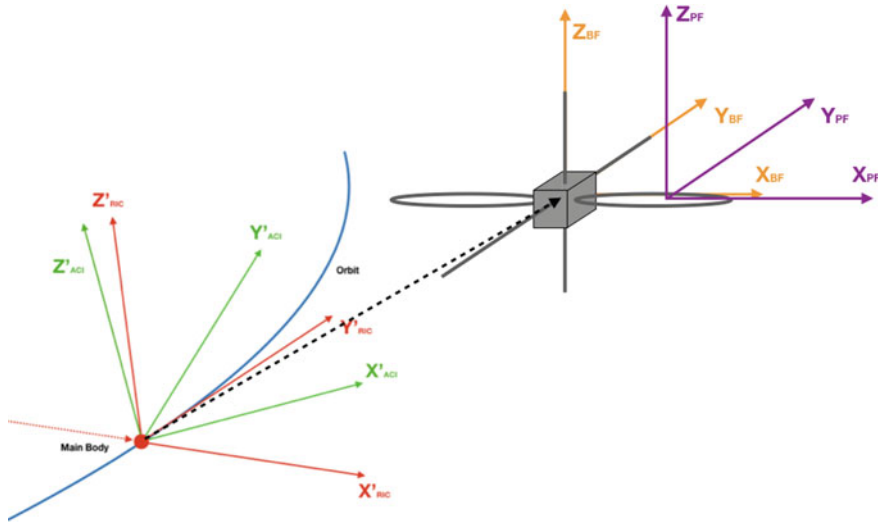


Fig. 2.6 Spacecraft-centered reference frames: body-fixed (BF) and post-fixed (PF)

2.3.2 Spacecraft-Centered Reference Frames

- Body-fixed (BF) frame: b
The body-fixed (BF) reference frame is rigidly “attached” to the spacecraft. The propagation of the attitude equations of motion is carried out in this reference frame. We assumed the attitude and the rotation rates considered as expressed in the BF frame and related to the ACI frame.

The definition of the BF reference frame is (see Fig. 2.6) as follows:

- O = origin at the spacecraft center of mass
- Axes = defined by the geometry of the spacecraft (usually oriented towards the principal axes of inertia)

The attitude quaternion ${}^b q_a$ defines the orientation with respect to the ACI frame.

2.4 Linearized Equations of Motion for Translational Motion

In the ensuing paragraph, we derive the equations of motion both for the case of the Clohessy–Wiltshire formulation, suitable for a main body with almost circular orbits, and for the case of the more accurate “full dynamic” formulation, which can be applied also to asteroids that are on orbits with high eccentricity.

The linearized equations of motion are obtained by considering the main body as a target and the E-Glider as the chaser, while the Sun is the third body in the full model case. The derivation of the equations of motion is in the RIC reference frame; thus, the distance of the target from the Sun and both the angular velocity and the angular acceleration with respect to the Sun are needed to derive the correct formulation. We assume the case of “proximity flight” (Scheeres and Marzari 2002) in the derivation of the equations of motion; thus, the distance of the spacecraft from the Sun and the distance of the main body from the Sun are comparable, and the distance of the spacecraft from the main body is much less than the previous two distances. By knowing the Keplerian parameters of the main body orbit around the Sun and $\theta = \theta(t)$, the equations applied in the computation of the distance d of the main body from the Sun, and the instantaneous asteroid orbital angular velocity and angular acceleration $\dot{\theta}$ and $\ddot{\theta}$ are

$$d = \frac{h_a^2}{\mu_s} \cdot \frac{1}{1 + e_a \cos \theta} = \frac{P_a}{1 + e_a \cos \theta} \quad (2.3)$$

$$\dot{\theta} = \frac{h_a^2}{d^2} = \frac{\sqrt{P_a \mu_s}}{d^2} \quad (2.4)$$

$$\ddot{\theta} = -2 \sqrt{\frac{\mu_s}{P_a}} \cdot \frac{e_a \sin \theta \dot{\theta}}{d} = -2 \frac{e_a \dot{\theta}^2 \sin \theta}{1 + e_a \cos \theta} \quad (2.5)$$

The angular velocity vector of the main body about the Sun is $\mathbf{\Omega}$, and it has a constant direction taken to be in the Z-direction in the perifocal (PQW) frame; thus, $\mathbf{\Omega}$ is also the angular velocity of the RIC frame with respect to the PQW frame. The magnitude of $\mathbf{\Omega}$ follows Eq. (2.4). Moreover, ${}^r\mathbf{\Omega} = (0, 0, \dot{\theta})$ and ${}^r\dot{\mathbf{\Omega}} = (0, 0, \ddot{\theta})$ can be easily verified. We need the spacecraft acceleration in the RIC frame to write the translational equation of motion in this reference frame. Let \mathbf{r} be the position vector of the spacecraft with respect to the main body, while \mathbf{D}_s and \mathbf{D}_a are the position vectors of the spacecraft with respect to the Sun and of the main body with respect to the Sun, respectively, as shown in Fig. 2.7. By knowing that $\mathbf{D}_s = \mathbf{D}_a + \mathbf{r}$, the absolute acceleration can be obtained after some mathematical steps as (Curtis 2010):

$$\ddot{\mathbf{D}}_s = \ddot{\mathbf{D}}_a + \mathbf{a} + \dot{\mathbf{\Omega}} \times \mathbf{r} + \mathbf{\Omega} \times \mathbf{\Omega} \times \mathbf{r} + 2\mathbf{\Omega} \times \mathbf{v} \quad (2.6)$$

where \mathbf{r} , \mathbf{v} , and \mathbf{a} are in the RIC frame. $\dot{\mathbf{\Omega}} \times \mathbf{r}$ is the term related to the angular acceleration of the frame, while $\mathbf{\Omega} \times \mathbf{\Omega} \times \mathbf{r}$ and $2\mathbf{\Omega} \times \mathbf{v}$ are the centrifugal term and the Coriolis accelerations, respectively. By solving Eq. (2.6) for the relative acceleration \mathbf{a} in the RIC frame and by knowing that $\ddot{\mathbf{r}} = \ddot{\mathbf{D}}_s - \ddot{\mathbf{D}}_a$, it results that

$$\mathbf{a} = {}^p\ddot{\mathbf{r}} - \dot{\mathbf{\Omega}} \times \mathbf{r} - \mathbf{\Omega} \times \mathbf{\Omega} \times \mathbf{r} - 2\mathbf{\Omega} \times \mathbf{v} \quad (2.7)$$

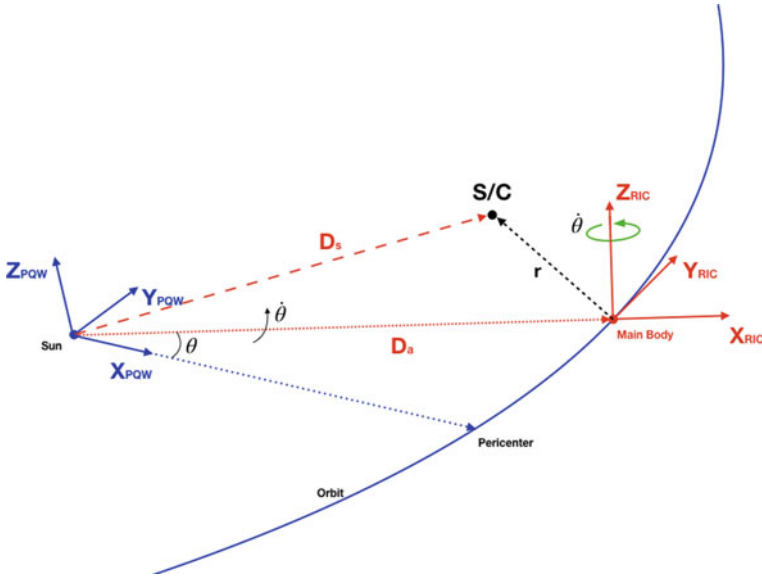


Fig. 2.7 Full model reference frames and vectors

The vector ${}^p\ddot{\mathbf{r}}$ measured in the inertial frame must be computed to define the relative acceleration in the RIC frame (the comoving one). To do that, a linearized model that is valid since $r \ll D_s, D_a$ can be used. By recalling that $\ddot{\mathbf{D}}_a = -\frac{\mu_s}{D_a^3} \mathbf{D}_a$, the equations of motion of the chaser relative to the target measured in the PQR inertial frame can be derived after some mathematical steps, assuming negligible the higher-order terms, as

$$\ddot{\mathbf{r}} = -\frac{\mu_s}{D_a^3} \left[\mathbf{r} - \frac{3}{D_a^2} (\mathbf{D}_a \cdot \mathbf{r}) \mathbf{D}_a \right] \quad (2.8)$$

By expressing \mathbf{r} and \mathbf{D}_a in the comoving RIC frame and by substituting the result into Eq. (2.7), the equation of motion for the translation for the full model in the RIC frame can be obtained by introducing the term $\frac{{}^r\mathbf{f}}{M}$, which gives the effects of the active forces acting on the spacecraft. Hence,

$${}^r\mathbf{a} = \frac{{}^r\mathbf{f}}{M} - \frac{\mu_s}{d^3} \begin{bmatrix} -2r_x \\ +r_y \\ +r_z \end{bmatrix} + \ddot{\theta} \begin{bmatrix} r_y \\ -r_x \\ 0 \end{bmatrix} + \dot{\theta}^2 \begin{bmatrix} +r_x \\ +r_y \\ 0 \end{bmatrix} + 2\dot{\theta} \begin{bmatrix} +v_y \\ -v_x \\ 0 \end{bmatrix} \quad (2.9)$$

$${}^r\dot{\mathbf{r}} = {}^r\mathbf{v} \quad (2.10)$$

For the purposes of this work, the main body gravitational perturbation effects, the Solar radiation pressure force, and the spacecraft electrostatic interactions with the

plasma field are considered. The effects of the magnetic field \mathbf{B} of both the asteroid and the Sun have been neglected in the following analyses by assumptions.

The Clohessy–Wiltshire formulation can be derived by assuming the target (the main body) to be on a circular unperturbed orbit around the third body (the Sun). This strong assumption allows to consider the mean motion of the main body as constant, hence $\dot{\theta} = N = \text{const}$. Moreover, $e = 0$ for a circular orbit; thus, the angular momentum can be written as $h_a = \sqrt{\mu_s D_a}$. By using these relations and by knowing that since $N = \text{const}$, the term related to $\dot{\Omega}$ is null, and the first cardinal equation in the RIC frame (Eq. (2.9)) can be rewritten by using the Clohessy–Wiltshire approximation as follows:

$${}^r \mathbf{a} = \frac{{}^r \mathbf{f}}{M} + 2N \begin{bmatrix} +v_y \\ -v_x \\ 0 \end{bmatrix} + N^2 \begin{bmatrix} +3r_x \\ 0 \\ -r_z \end{bmatrix} \quad (2.11)$$

$${}^r \dot{\mathbf{r}} = {}^r \mathbf{v} \quad (2.12)$$

2.5 Attitude Dynamics

The equation of motion for the attitude is the classical Euler’s equation used for attitude propagation. This equation can be derived from the momentum equation written in an inertial reference frame. The Euler’s equation in the BF frame is

$${}^b \mathbf{J} {}^b \dot{\boldsymbol{\omega}}_{ab} = {}^b \mathbf{J} {}^b \boldsymbol{\omega}_{ab} \times {}^b \boldsymbol{\omega}_{ab} + {}^b \mathbf{T} \quad (2.13)$$

${}^b \boldsymbol{\omega}_{ab}$ is the angular rate of the BF frame (thus the spacecraft angular rate) with respect to the inertial reference frame (ACI) expressed in the BF reference frame. The term ${}^b \mathbf{T}$ collects all the active torques applied to the spacecraft expressed in the BF reference frame. The active torques are given by the gravity field of the main body, the solar radiation pressure torque, and the electrostatic effects. The attitude dynamics equation must be completed by adding also the attitude kinematics. We used the quaternion representation in this work; thus, the quaternion kinematics equation can be written as

$${}^b \dot{\mathbf{q}}_a = \frac{1}{2} \begin{bmatrix} {}^b \boldsymbol{\omega}_{ab} \\ 0 \end{bmatrix} \otimes {}^b \mathbf{q}_a \quad (2.14)$$

Here, ${}^b \mathbf{q}_a$ is the attitude quaternion from the ACI to BF reference frame, and the symbol \otimes stands for the quaternion products. The quaternion must be normalized after each integration step in order to avoid divergences.

2.6 Gravitational Forces and Moments

Two different gravity field models have been considered, the classical and simple point mass gravity model and a more complex and accurate model based on the spherical harmonics expansion model.

2.6.1 Point Mass Gravity Model

The point mass gravity model is suitable for bodies with a spherical/symmetric mass distribution. If we can assume that the mass of the main body is concentrated in the center of gravity of the body itself, then the point mass gravity model is valid. This model is singularity-free, and its computational cost is extremely low, such that the point mass gravity model can be used for feasibility studies and first approximation analysis. \mathbf{f}_g , the gravitational force acting on the spacecraft, can be simply derived by knowing that the gravitational force is conservative. Thus, the gravitational acceleration is $\mathbf{a}_g = \nabla U$, where U is the gravitational potential; thus, it results to be

$$\mathbf{a}_g = -\frac{\mu}{r^3} \mathbf{r} \quad (2.15)$$

The formulation for the gravity gradient tensor \mathbf{G}_g is the following (Gottlieb 1993):

$$\mathbf{G}_g = \nabla \mathbf{a}_g = -\frac{\mu}{r^3} \left(\mathbf{I} - \frac{3}{r^2} \mathbf{r} \otimes \mathbf{r} \right) \quad (2.16)$$

Notice that the symbol \otimes stands for the outer product.

2.6.2 Spherical Harmonics Model

The expansion in spherical harmonics of the gravitational field is a commonly used method to compute the gravitational potential U_g . This method offers the possibility to compute the tangential components of the gravitational and to achieve an accuracy level higher than the point mass model without introducing an excessively high computational load (e.g., as the FE MASCON method). For the analysis of an E-Glider, the terminator region can be of particular interest; thus, a singularity-free method is mandatory for the computation of the gravity field also at the poles. There are several singularity-free methods; the one used here is the method developed by Pines in 1973 (Pines 1973) due to its accuracy and fast computational capabilities. A modified recursion formula is needed for the computation of the Legendre polynomials to also achieve the stability required for high-order gravitational models since the one

originally proposed by Pines was unstable for high n (Eckman et al. 2011; Lundberg 1988). Hence, the gravitational potential can be rewritten as

$$U_g = \sum_{n=0}^{\infty} \rho_n \sum_{m=0}^n A_{n,m}(u) D_{n,m}(s, t) \quad (2.17)$$

The definition of ρ_n is Pines (1973) (see Eq. (2.26)) and $D_{n,m}(s, t)$ is a mass coefficient function. Equation (2.17) must be differentiated in Pines' reference frame in order to compute the gravitational acceleration \mathbf{a}_g ; thus, the equations for the gravitational acceleration and for the gravity gradient tensor are

$$\mathbf{a}_g = \nabla U_g = a_1 \hat{\mathbf{i}} + a_2 \hat{\mathbf{j}} + a_3 \hat{\mathbf{k}} + a_4 \hat{\mathbf{r}} \quad (2.18)$$

$$\mathbf{G}_g = \nabla \mathbf{a}_g = \frac{\partial}{\partial \mathbf{r}} \left(a_1 \hat{\mathbf{i}} + a_2 \hat{\mathbf{j}} + a_3 \hat{\mathbf{k}} + a_4 \hat{\mathbf{r}} \right) \quad (2.19)$$

The equations to compute both the coefficients of Eq. (2.18) and the derived coefficients of Eq. (2.19) simply and efficiently are available in Pines (1973).

2.6.3 Forces and Torques on an Extended Body

The computation of both the gravitational forces and the gravitational torques acting on the spacecraft is mandatory, independently from the model assumed. The local acceleration $\mathbf{a}_g(\mathbf{r})$ and the local gravity gradient $\mathbf{G}_g(\mathbf{r})$ components can be computed as explained in the previous section.

In the analysis, the spacecraft can be considered or as an extended single body (hence with a single mass and a single inertia), or as an ensemble of parts (multibody approach) with their own mass localized at the center of gravity of the part itself. A linear model can be assumed for the case of a satellite modeled as an extended body. In this case, the vector $\boldsymbol{\rho} = \mathbf{r} - \mathbf{r}_0$, in which \mathbf{r}_0 is the position of the center of mass, gives the position of a point with respect to the center of mass of the spacecraft; thus, the approximated equation for the gravitational acceleration is

$$\mathbf{a}_g(\mathbf{r}) = \mathbf{a}_g(\mathbf{r}_0) + \mathbf{G}_g(\mathbf{r}_0)\boldsymbol{\rho} \quad (2.20)$$

By integrating Eq. (2.20) and by remembering that $\int_{S/C} \boldsymbol{\rho} dM$ is the first moment of mass about the center of mass itself, hence null by definition, the gravitational force results to be

$$\mathbf{f}_g = \mathbf{a}_g(\mathbf{r}_0)M \quad (2.21)$$

In the case of a spacecraft modeled as an ensemble of parts, the total gravitational force acting on the whole spacecraft can be computed as the summation of the forces acting on each part constituting the spacecraft; thus, by referring with the index i to the i -th part, the gravitational force is

$$\mathbf{f}_g = \sum_i \left(\int_{i\text{-th part}} \mathbf{a}_g(\mathbf{r}_i) dM_i \right) = \sum_i (\mathbf{a}_g(\mathbf{r}_i) M_i) \quad (2.22)$$

where \mathbf{r}_i and M_i are the position of the center of mass and the mass of the i -th part, respectively. This method has a stronger impact on the computational load with respect to the previous one, but it allows us to achieve a higher level of precision for highly extended bodies (where the linearized model of the first case is no more valid).

For the case of a single extended body, by using the very same assumptions of the previous paragraph, the equation for the gravity torque is

$$\mathbf{T}_g = \int_{S/C} \boldsymbol{\rho} \times (\mathbf{a}_g(\mathbf{r}_0) + \mathbf{G}_g(\mathbf{r}_0)\boldsymbol{\rho}) dM \quad (2.23)$$

By remembering that, once again, $\int_{S/C} \boldsymbol{\rho} dM$ is the first moment of mass about the center of mass itself, hence null by definition, we can obtain

$$\mathbf{T}_g = \int_{S/C} \boldsymbol{\rho} \times \mathbf{G}_g(\mathbf{r}_0) dM \quad (2.24)$$

The total torque for the case of a spacecraft made by several parts can be expressed as the sum of the torques given by the gravitational forces acting on each part i , resulting in

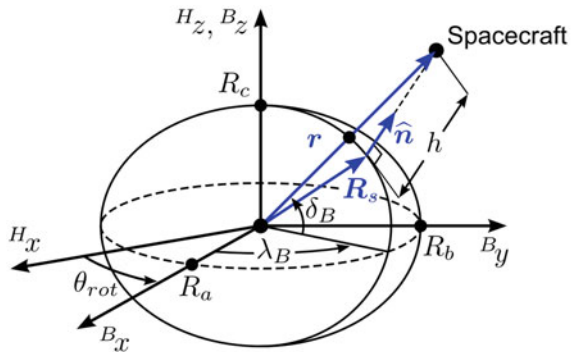
$$\mathbf{T}_g = \sum_i \left(\int_{i\text{-th part}} \boldsymbol{\rho}_i \times \mathbf{a}_g(\mathbf{r}_i) dM_i \right) = \sum_i (\boldsymbol{\rho}_i \times \mathbf{a}_g(\mathbf{r}_i) M_i) \quad (2.25)$$

where $\boldsymbol{\rho}_i$ is the position of a point of the i -th part with respect to the center of mass of the i -th part itself. As before, this method is both more precise and more expensive from the computational point of view.

2.6.4 Irregular Asteroid Model

The asteroid is modeled as a homogeneous triaxial ellipsoid with semimajor axes R_a , R_b , and R_c ($R_a \geq R_b \geq R_c$). The mean radius of the asteroid is given as $R = 50$ m, which satisfies $R^3 = R_a R_b R_c$, and the axis ratio is taken as a variable in later subsections. The asteroid is rotating uniformly about the shortest axis with the rotation period of $T_{rot} = 8$ hr, and the rotation axis is assumed to be perpendicular to

Fig. 2.8 Reference frames around an ellipsoidal asteroid



the ecliptic plane (Kryszyńska et al. 2007). Then the asteroid body-fixed coordinate can be defined as shown in Fig. 2.8. Here, a left superscript “ H ” represents the Hill coordinate system, and “ B ” represents the asteroid body-fixed frame. The $^H z$ axis and $^B z$ are identical because of the assumption regarding the rotation axis. Henceforth, the position of a spacecraft is expressed in terms of the Hill coordinate as $^H \mathbf{r} = [x, y, z]^T$ and in terms of the asteroid body-fixed frame as $^B \mathbf{r} = [x_B, y_B, z_B]^T$.

Let $^H C_B$ denote the rotational transformation matrix from the asteroid body-fixed coordinate to the Hill coordinate system. Then the coordinate transformation for an arbitrary state vector \mathbf{u} is expressed as $^H \mathbf{u} = ^H C_B ^B \mathbf{u}$, where $^H C_B$ is given by the equation below.

$$^H C_B = \begin{bmatrix} \cos \theta_{rot} & -\sin \theta_{rot} & 0 \\ \sin \theta_{rot} & \cos \theta_{rot} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.26)$$

Here, θ_{rot} is the rotation phase of an asteroid and given as a function of time by the following equation:

$$\theta_{rot} = \frac{2\pi}{T_{rot}} t \quad (2.27)$$

2.6.5 Irregular Gravitational Field

The gravitational potential of an asteroid is calculated based on a triaxial ellipsoid model. The gravitational coefficients C_{mn} of its spherical harmonics expansion up to the fourth order are defined by the following equations (Scheeres 2012):

$$\begin{aligned}
C_{20} &= \frac{1}{10R_a^2} \{2R_c^2 - (R_a^2 + R_b^2)\} \\
C_{22} &= \frac{1}{20R_a^2} (R_a^2 - R_b^2) \\
C_{40} &= \frac{15}{7} (C_{20}^2 + 2C_{22}^2) \\
C_{42} &= \frac{5}{7} C_{20} C_{22} \\
C_{44} &= \frac{5}{28} C_{22}^2
\end{aligned} \tag{2.28}$$

Using these coefficients, the gravitational potential is given by the equation below.

$$\begin{aligned}
U_G = \frac{\mu}{r} & \left[1 + \left(\frac{R}{r} \right)^2 \left\{ \frac{1}{2} C_{20} (3 \sin^2 \delta_B - 1) + 3 C_{22} \cos^2 \delta_B \cos 2\lambda_B \right\} \right. \\
& + \left(\frac{R}{r} \right)^4 \left\{ \frac{1}{8} C_{40} (35 \sin^4 \delta_B - 30 \sin^2 \delta_B + 3) \right. \\
& \left. \left. + \frac{15}{2} C_{42} \cos^2 \delta_B (7 \sin^2 \delta_B - 1) \cos 2\lambda_B + 105 C_{44} \cos^4 \delta_B \cos 4\lambda_B \right\} \right]
\end{aligned} \tag{2.29}$$

where δ_B and λ_B denote the latitude and longitude, respectively, defined in terms of the asteroid body-fixed frame. The relation between (δ_B, λ_B) and the position of a spacecraft can be expressed as follows:

$$\begin{aligned}
\delta_B &= \sin^{-1} \left(\frac{z_B}{r} \right) \\
\lambda_B &= \tan^{-1} \left(\frac{y_B}{x_B} \right)
\end{aligned} \tag{2.30}$$

Then from Eq. (2.29), the gravitational acceleration from an ellipsoidal asteroid can be obtained.

2.6.6 Gravitational Torque

The gravitational torque \mathbf{T}_G can be expressed as follows by applying the Taylor series expansion (Hughes 1986):

$$\mathbf{T}_G \simeq \frac{3\mu}{r^5} \mathbf{r} \times \mathbf{I} \mathbf{r} \tag{2.31}$$

It is important to note that both the gravity gradient torque and the electrostatic torque are dependent on the position of a spacecraft with respect to a small body, and therefore the orbital motion of a spacecraft exerts an influence on its attitude motion.

2.7 Solar Radiation Forces and Moments

The exchange of momentum between the photons and a surface results in solar radiation pressure. Each source of electromagnetic radiation has an effect on a solid surface, but the pressure given by the solar radiation is predominant at 1 AU (Lyon 2004); thus, the other terms are negligible in this analysis. In a macro-model approach, the incident radiation on a surface can be either absorbed, specularly reflected, or diffusely reflected (by assuming no transmission of radiation through the spacecraft). In the following, dA is the surface area with normal $\hat{\mathbf{n}}$ subjected to the incident flux Φ . The incident flux has an inclination α with respect to the normal $\hat{\mathbf{n}}$. The versor $\hat{\mathbf{s}}$ points towards the origin of the radiation. The summation of the three forces given by the absorption, the specular reflection, and the diffusive reflection of the incoming radiation gives the total resultant force over a flat surface dA :

$$d\mathbf{f}_p = d\mathbf{f}_{ps} + d\mathbf{f}_{pd} + d\mathbf{f}_{pa} = -\frac{\Phi}{c} dA \cos\alpha \left[\left(2C_{ps}\cos\alpha + \frac{2}{3}C_{pd} \right) \hat{\mathbf{n}} + (C_{pd} + C_{pa}) \hat{\mathbf{s}} \right] \quad (2.32)$$

The two models suitable for the computation of the solar radiation pressure effects are the cannonball model and the backward ray-casting model.

2.7.1 Cannonball Model

The cannonball model is a classical simplified approach to compute the resultant force and torque over the external surfaces of a spacecraft. The spacecraft can be approximated by a sphere of equivalent external area and with constant thermo-optical properties as an assumption of this model. The incoming flux of electromagnetic radiation is incident over the cross-sectional area, such that the total force results to be

$$\mathbf{f}_p = -\frac{\Phi}{c} \pi R^2 \left(C_{ps} + \frac{13}{9} C_{pd} + C_{pa} \right) \hat{\mathbf{s}} \quad (2.33)$$

With this model, the determination of the resulting torque is not possible (indeed $\mathbf{T}_p = 0$).

2.7.2 *Backward Ray-Casting Model*

This method relies on the generation of rays on the surfaces of the spacecraft. The propagation of these rays in the backward direction allows us to check if there are intersections and shadowing between surfaces. The spacecraft surfaces must be approximated as an ensemble of large arrays made by small “facets” to apply this method. Each facet acts as a source for a single ray; thus, if the facets are small enough, the total force can be computed as a summation instead of solving the integral over the entire surface. If $\hat{\mathbf{s}} \cdot \hat{\mathbf{n}} < 0$, the facet is certainly in shadow; hence, the ray and the facet can be discarded immediately from the computation.

A ray-casting intersection algorithm is needed to evaluate the rays that intercept one surface before reaching another one. The contribution must be discarded after the first intersection. The degree of precision depends on the number of facets used. The higher the number of facets, the higher the precision, but also the higher the computational cost. If thin wires or, more in general, thin features are present, the aliasing may arise (as in the case of forward ray casting), but the error introduced is usually not significant.

2.8 Plasma and Charging Interactions Around Small Asteroids

Without a global magnetic field, airless bodies such as asteroids are exposed directly to solar radiation, and space plasmas are thus electrically charged. To calculate the electrostatic force applied to an E-Glider, the plasma environment around an asteroid and the interactions between the E-Glider, asteroid, and plasma must be modeled appropriately. This section discusses such interactions and relevant modeling studies.

2.8.1 *Plasma and Charging Environments Around Small Asteroids*

Asteroids have a wide range of size. Typical near-earth asteroids (NEAs) have a size distribution from 1 m to ~ 32 km. Most asteroids are irregular shaped and show significant macroporosity (Clark et al. 2002). Many asteroids are covered by a regolith layer, similar to the Moon. An asteroid is a dielectric object. While few direct measurements of asteroids’ surface properties are currently available, one expects that the conductivity and dielectric constant of an asteroid’s surface would be similar to that of the lunar surface, which are estimated to be 10^{-14} S/m - 10^{-9} S/m and 2–10, respectively (Olhoeft and Strangway 1975).

Solar wind (SW) is a mesothermal plasma (the directed plasma flow speed is larger than ion thermal speed but less than electron thermal speed). While the solar wind

plasma parameters can undergo substantial change, the parameters under the average solar wind condition at 1 AU relevant to this study are plasma density $n_0 \sim 5 \text{ cm}^{-3}$; solar wind flow speed $V_{sw} \sim 400 \text{ km/s}$; and solar wind electron temperature $T_e \sim 10 \text{ eV}$. Based on these parameters, the solar wind Debye length is $\lambda_D \sim 10.5 \text{ m}$, the ambient solar wind proton flux density is $\Gamma_{i0} \sim n_0 V_{sw} \simeq 2 \times 10^8 \text{ cm}^{-2} \text{ s}^{-1}$ (current density $J_{i0} \sim 0.32 \mu\text{A/m}^2$), and the ambient solar wind electron flux density is $\Gamma_{e0} \sim n_0 \sqrt{kT_e/m_e} \simeq 6.6 \times 10^8 \text{ cm}^{-2} \text{ s}^{-1}$ (current density $J_{e0} \sim 1 \mu\text{A/m}^2$). Since the mean-free-path in the solar wind is typically on the order of 1 km or 10 s, the plasma flow around small asteroids (size smaller than 1 km) can be considered collisionless.

Photoelectron emissions occurs at a sunlit surface. The photoelectron temperature is $T_{ph} \sim 2.2 \text{ eV}$. Under normal sunlight incidence at 1 AU, the photoelectron flux density is $\Gamma_{ph0} \sim n_{ph0} \sqrt{kT_{ph}/m_e} \simeq 39.8 \times 10^8 \text{ cm}^{-2} \text{ s}^{-1}$ (current density $J_{ph0} \sim 6.4 \mu\text{A/m}^2$), the number density is $n_{ph0} \sim 64 \text{ cm}^{-3}$, and the photoelectron Debye length is $\lambda_{D,ph0} \sim 1.38 \text{ m}$. The photoelectron number density at the asteroid surface as a function of the local sun elevation angle α is $n_{ph}(\alpha) = n_{ph0} \sin \alpha$. For surfaces with a small Sun incidence angle, the photoelectron Debye length would be significantly larger than $\lambda_{D,ph0}$, and the photoelectron flux density significantly smaller than Γ_{ph0} due to reduced photoelectron emission.

An asteroid in a mesothermal solar wind flow forms a plasma wake behind it in which only mobile electrons can penetrate (Fig. 2.14 Wang and Hastings 1992; Wang and Hu 2018). Since only the electrons can impinge on the dark side of the body, the wake side surface will charge negatively until the local electric field is strong enough to repel all impingement electrons. Depending on the solar wind condition, the potential on the dark side can thus reach negative values of tens to several hundreds of volts (Fig. 2.15) (Lee 1996; Mendis et al. 1981). On the sunlit side, while both ions and electrons can strike the surface, surface charging is mostly dominated by photoelectron emission. The emitted photoelectrons cause the surface to accumulate positive charges until its potential is high enough to impair photoelectron emission itself. The potential of the sunlit surface is therefore on the order of the photoelectron temperature, i.e., a few volts. One notes that the combined effects from the plasma flow and localized sunlit/shadow region on the asteroid surface can generate a complex plasma flow field around asteroids and differential charging on the asteroid surface.

Many studies have been carried out to investigate the interactions between asteroids and solar wind plasma and the dynamics of charged dust grains (Lee 1996; Nitter et al. 1998; Colwell et al. 2005; Han and Wang 2019; Yu et al. 2019). In this section, both analytical modeling and numerical simulations will be presented. Nitter in Nitter et al. (1998) carried out an analytical derivation of the mono-dimensional plasma sheath around an asteroid. From Nitter et al. (1998), we first derived a multi-sheath model that has been used also for preliminary study on the E-Glider concept (Quadrelli et al. 2017a, b; Kikuchi 2017). This derived model was obtained by relaxing some hypotheses (Hartzell 2012) and by including the effects of drifting electrons (Jeong 2008). The analytical approach is not capable of solving the equa-

tions for nonelementary cases unless strong assumptions and simplifications are taken into account (Corradino 2018). Hence, numerical simulations based on fully kinetic particle-in-cell (PIC) simulations are also carried out to obtain the electric field around small asteroids.

2.8.2 Analytical Modeling of the Plasma and the Electric Field

We first present a simplified analytical analysis. The relationship between particle densities and the electrostatic potential is described by Poisson's equation. Given that the asteroid has a spherical shape and the particle distribution is symmetrical about the subsolar line, the electrostatic potential is expressed as a function of the altitude h and the solar incident angle θ defined by the equations below.

$$\begin{aligned} h &= r - R \\ \theta &= \cos^{-1}\left(-\frac{x}{r}\right) \end{aligned} \quad (2.34)$$

Here, $R = D/2$ is the radius of an asteroid. Then the electrostatic potential around the asteroid is modeled using the following Poisson equation, which is expressed as a second-order differential equation in terms of h (Nitter et al. 1998; Jeong 2008; Hartzell 2012):

$$\frac{\partial^2 \phi(h, \theta)}{\partial h^2} = -\frac{e}{\varepsilon_0}(n_i - n_e - n_p) \quad (2.35)$$

where e is the elementary charge; ε_0 is the vacuum permittivity; and n_i , n_e , and n_p are number densities of the solar wind ions, the solar wind electrons, and the photoelectrons, respectively. Assuming that the solar wind ions are modeled as a mono-energetic beam, and that the solar wind electrons and the photoelectrons follow Maxwellian distributions, n_i , n_e , and n_p are given by analytical expressions, as presented in Jeong (2008). Based on this assumption, $\phi(h, \theta)$ can be solved numerically from Eq. (2.36).

Given that there is the direct relationship between (h, θ) and the position vector \mathbf{r} , the electrostatic potential can also be expressed in the Cartesian coordinate as $\phi(\mathbf{r})$. Therefore, the electrostatic force acting on a spacecraft with the charge Q is calculated from the equation below.

$$\mathbf{F}_E = Q \cdot \mathbf{E}(\mathbf{r}) = -Q \frac{\partial \phi(\mathbf{r})}{\partial \mathbf{r}} \quad (2.36)$$

Here, $\mathbf{E}(\mathbf{r})$ denotes the local electrostatic field. Although the Poisson's equation is decreased to a one-dimensional differential equation as shown in Eq. (2.36), this electrostatic force model can represent three-dimensional variation by numeri-

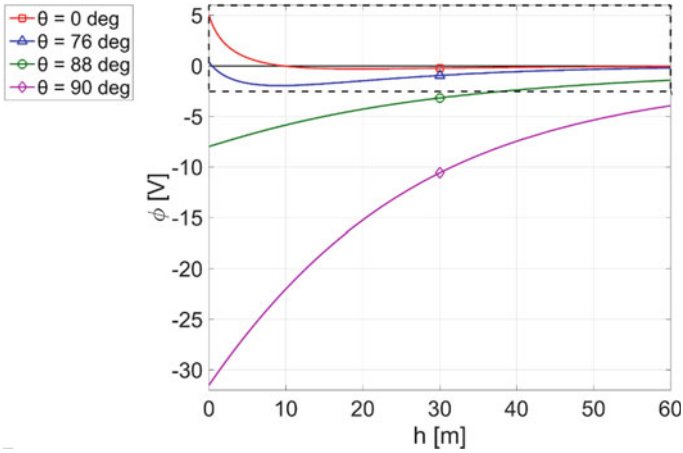


Fig. 2.9 Electrostatic potential profiles for different solar incident angles

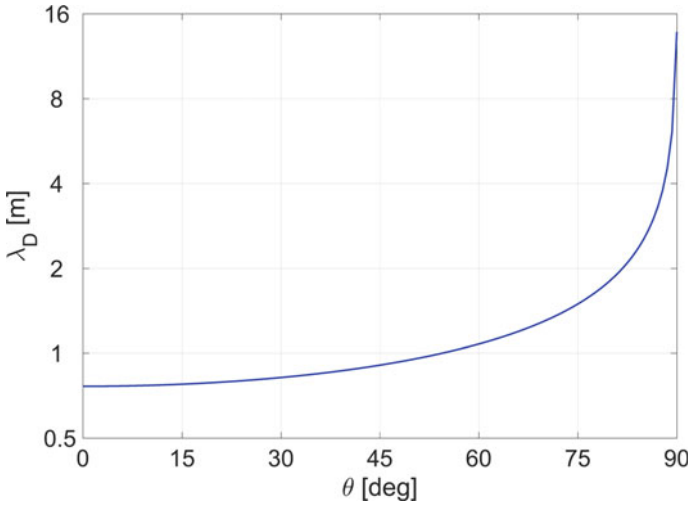


Fig. 2.10 Relationship between the solar incident angle and the Debye length

cally approximating the electrostatic field in longitudinal and latitudinal directions (Hartzell 2012).

Figure 2.9 depicts the electrostatic potential profiles computed from Eq. (2.35) for several different solar incident angles. It can be observed that the surface of the asteroid is positively charged when the solar incident angle is small (i.e., near the subsolar region), while the surface is negatively charged when the solar incident angle is large (i.e., near the terminator region). The enlarged view in Fig. 2.9 also shows that non-monotonic sheath profiles appear in some cases, which implies that the plasma structure around an asteroid is complex. Figure 2.10 shows the relationship between

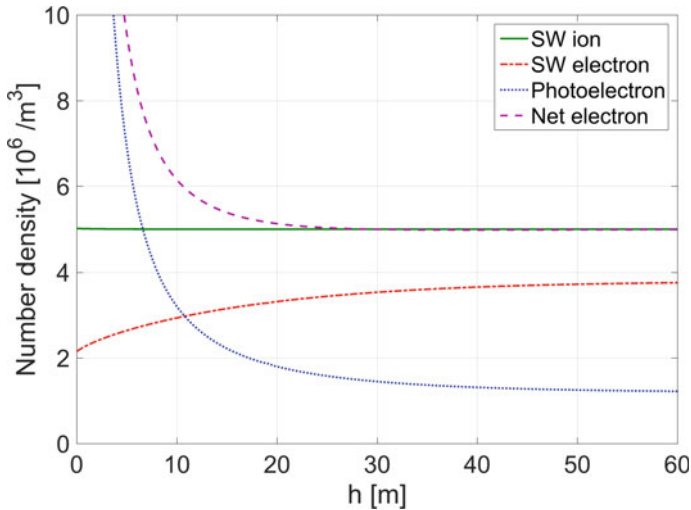


Fig. 2.11 Number densities of charged particles

the solar incident angle and the Debye length λ_D , which is defined in Nitter et al. (1998). The Debye length is an indication of how far from the asteroid surface the electrostatic effects can exert influence. Thus, Fig. 2.10 implies that the electrostatic force obtained in the terminator region is stronger than that of the subsolar region.

Figure 2.11 illustrates the distributions of the charged particles along the subsolar line ($\theta = 0$ deg). One of the remarkable features is that the number density of photoelectrons is considerably large near the surface. This dense photoelectron layer on the dayside involves a strong screening effect. Consequently, the Debye length near the subsolar point is comparatively short, as shown in Fig. 2.10. It is also shown in Fig. 2.11 that when the altitude increases, the density of the net electrons approaches that of solar wind ions, which indicates that the plasma is in a quasi-neutral state.

Figure 2.12 illustrates a contour map of the electrostatic potential around the asteroid, which is expressed in the x - y plane (please note that the symbol $[S]$ is used in the figures of this chapter to mark the direction of the Sun). The broken line in the enlarged view corresponds to the potential level of zero volts, and it is evident that the dayside region close to the surface has positive potential; on the other hand, there exists a strong negative potential region on the nightside and around the terminator. Note that the wake streams of the solar wind behind the asteroid are not considered in this simulation, which can exert a strong influence on the nightside electrostatic potential (Han et al. 2016a; Yu et al. 2016); however, the current model is regarded to be valid for the analysis of plasma structures on the dayside and in the terminator region, which is our major interest as mentioned in the introduction. The visual representation of the electrostatic field is displayed in Fig. 2.13. The direction and magnitude of the electrostatic field are expressed by the arrows and their colors,

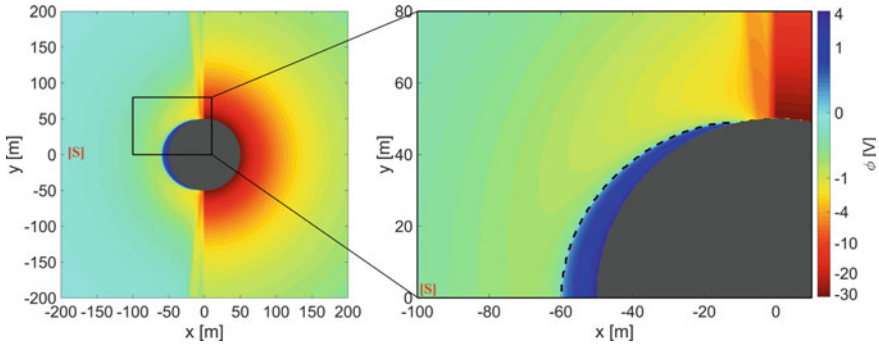


Fig. 2.12 Electrostatic potential around the asteroid

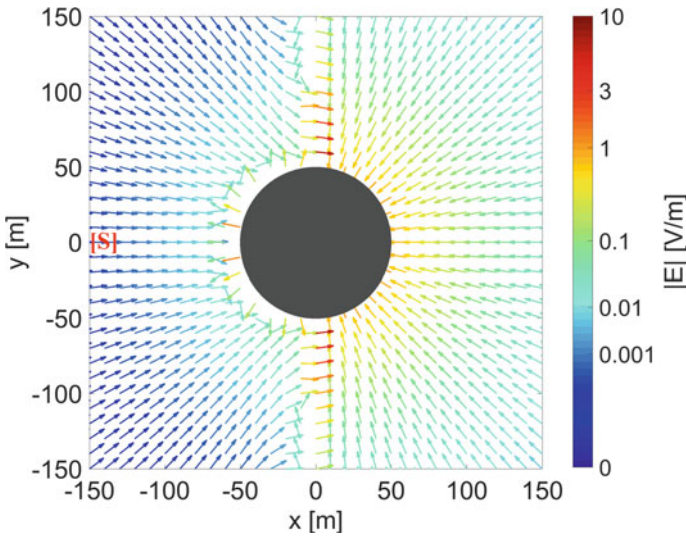


Fig. 2.13 Electrostatic field vector components around the asteroid

respectively. This figure is useful to understand the behavior of electrostatic force acting on an E-Glider and to make effective use of it.

These unique characteristics regarding the plasma environment around an asteroid, which are provided in Figs. 2.9, 2.10, 2.11, 2.12 and 2.13, have been revealed in previous studies as well and are consistent with them (Nitter et al. 1998; Poppe 2011; Han et al. 2016a; Yu et al. 2016). From these observations, it can be concluded that the plasma model used in this study is valid and can be applied for E-Glider simulations.

Besse and Rubin presented in Besse and Rubin (1980) a very clear and simple model of dipole charging, with an insight on the definition of the photoelectron sheath for a sphere. Besse and Rubin (1980) shows the possible issues with a trapping region

for photoelectrons on the sunward side of a charged sphere, and the model used in Besse and Rubin (1980) would be of interest in future studies. For instance, a cross-comparison of the analytical model described in this section coupled with the model presented in Besse and Rubin (1980) with the more accurate but time-consuming numerical simulations using the *USC-IFEPIC* (presented in the next section) will be carried out for E-Glider charging in plasma. Once calibrated, the analytical method can be used to quickly estimate the effects of the photoelectron emission from the E-Glider, making it possible to easily include the effects of spacecraft material properties in the analyses presented in this work. As the spacecraft material selection has not been finalized yet, detailed material properties were not included in E-Glider charging calculations in this paper but will be assessed in future study.

2.8.3 PIC Simulations of the Plasma and Electric Field

As discussed in Section 8.1, the plasma environment around small asteroid is that of a collisionless, mesothermal plasma flow. Furthermore, since the photoelectrons dominate the charging of the sunlit surface and the solar wind electrons dominate the charging of the wake side surface, the detailed dynamics of both the solar wind electrons and the photoelectrons play an important role in asteroid charging. Thus, full kinetic numerical simulations are usually required to correctly model the plasma environment and electric field around a small asteroid.

The collisionless nature of plasma flow around small asteroids renders the particle-in-cell (PIC) method (Birdsall and Langdon 1991), which solves plasma particle trajectory, space charge, and the Poisson equation self-consistently, as the preferred modeling method. To resolve the electron dynamics and the photoelectron sheath correctly, we apply a full particle PIC simulation model, which uses macro-particles to represent both electrons and ions. To maintain the correct mesothermal velocity ratio, full particle simulations must also be carried out using the correct ion to electron mass ratio.

In this study, the simulation model used to resolve the plasma environment and electric field is a recently developed immersed-finite-element particle-in-cell code, *USC-IFEPIC*, Han et al. (2016a, b). The *USC-IFEPIC* code is a three-dimensional (3D) full particle electrostatic PIC code designed to simulate plasma interactions involving complex boundary conditions. All plasma species (solar wind protons and electrons, photoelectrons, and secondary electrons) are represented by macro-particles. The electric potential Φ , the space charge, and the trajectories of each macro-particle are solved self-consistently from Poisson's equation and Newton's second law:

$$\nabla \cdot (\epsilon \nabla \Phi) = -e(n_i - n_e - n_{ph} - n_{se}), \quad m \frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (2.37)$$

where ϵ is the relative permittivity, n_i , n_e , n_{ph} , and n_{se} are the number densities of the solar wind ion, solar wind electron, photoelectron, and secondary electron, respectively. In this study, the contribution of secondary electrons emitted from the asteroid surface is ignored because the secondary density is orders of magnitude smaller than that of the other species (Han 2015; Lee 1996; Whipple 1981). Application of the E-Glider is for small asteroids. As the size of most small asteroids is smaller than the ion gyro-radius, the effect of the interplanetary magnetic field is also ignored.

In plasma charging studies, the object surface is typically treated as the boundary to the ambient plasma, and surface charging is handled through a boundary condition coupled with a current balance calculation. Such an approach, which is used in standard spacecraft charging software, is not always sufficient for asteroid charging because asteroids are dielectric objects. The asteroid capacitance is often not trivial, and the combined effects from plasma flow and localized sunlit/shadow regions also generate complex differential charging on the surface. In the USC-IFEPIC code, the asteroid is considered as part of the simulation domain with material conductivity explicitly included. The relative permittivity of the asteroid surface is taken to be similar to that of the lunar regolith, $\epsilon \sim 4$. The electric field is solved for both inside and outside of the asteroid; and asteroid charging is calculated directly from local charge deposition at the surface (Han et al. 2016b).

Full particle PIC simulations using the real ion-to-electron mass ratio are computationally expensive. A critical aspect in this modeling study is to resolve the electric field accurately for complex asteroid shapes while still maintaining an efficient computing speed. USC-IFEPIC applies a novel field solution algorithm, the non-homogeneous interface flux jump immersed-finite-element PIC (IFE-PIC) algorithm (Han et al. 2016a, b), to solve the electric field. In this algorithm, the boundary is treated as an interface between two mediums. The solution mesh can be generated regardless of the location of the interface. Poisson's equation is solved using a finite element (FE) method with a bases function designed to resolve the discontinuity of the electric field flux at the interface (Han et al. 2016b). This approach allows one to use a Cartesian-based mesh to solve the electric field in the presence of complex boundaries with the same accuracy as a body-fitting mesh FE solver (Kafafy et al. 2005; Kafafy and Wang 2006). This approach also preserves the standard particle-search and particle-mesh interpolation in PIC, thus maintaining the standard PIC computation speed (Wang et al. 2006).

The USC-IFEPIC model was validated against the one-dimensional (1D) analytical solutions of Nitter et al. (1998), Jeong (2008) in Ref. Han (2015) and was previously applied to simulate lunar surface charging (Han et al. 2018), asteroid charging (Han and Wang 2019), and charged dust dynamics around small asteroids (Yu et al. 2019). Figures 2.14 and 2.15 show a typical set of asteroid-plasma interaction simulations using USC-IFEPIC. The asteroid is taken to be a spherical object that has a rock core and an outer layer of dust grains. The radius of the rock core is about 12.62 m, the outer dust layer thickness is $d_{layer} \simeq 1.38$ m, and the total asteroid radius is $r_A = 14$ m. The asteroid size is similar to that of the near-Earth asteroids 1998 KY26, 2004 FH, 367943 Duende, and 2014 RC.

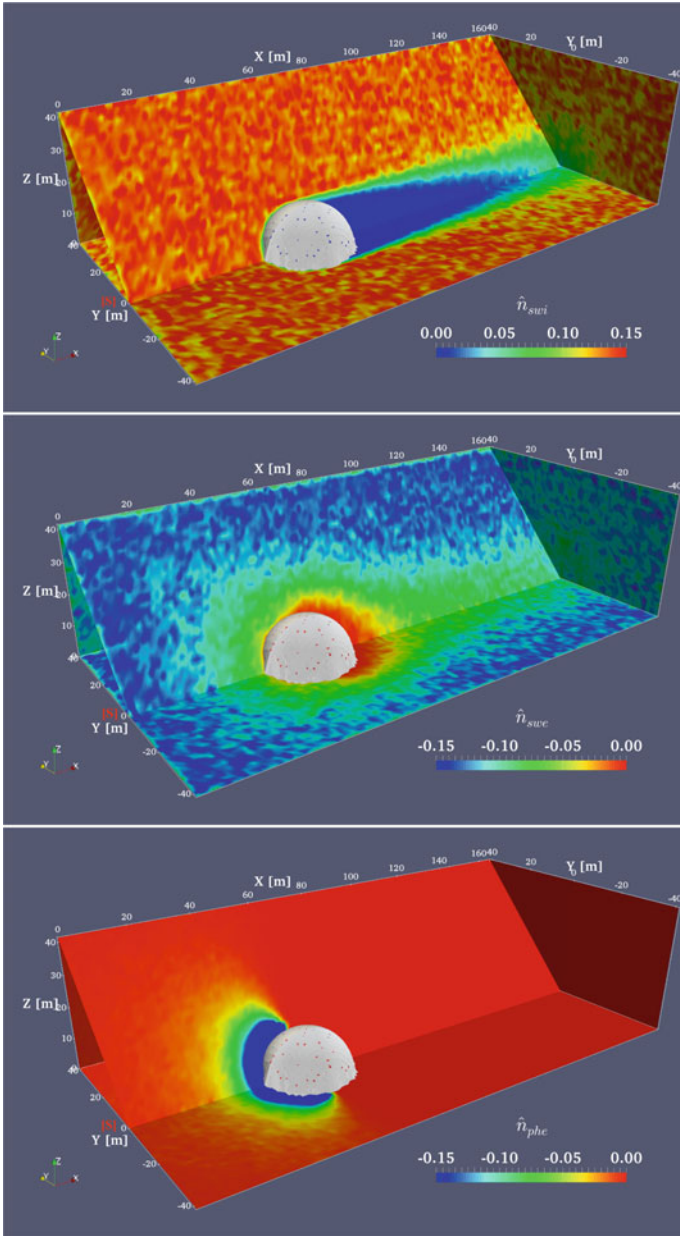


Fig. 2.14 USC-IFEPIC simulations of solar wind–asteroid interactions. From top to bottom: solar wind ions, solar wind electrons, and photoelectrons

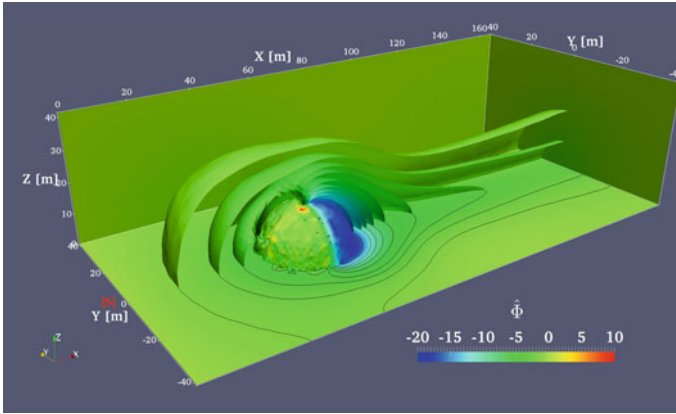


Fig. 2.15 USC-IFEPIC simulations of solar wind–asteroid interactions: electrostatic potential

The species density distribution (Fig. 2.14) shows very clearly the presence of a marked and well-defined plasma wake. The general rarefaction in proximity of the asteroid surface can be attributed to the overall negative net charge acquired by the body. The photoelectrons show a typical diffusion pattern from the sunlit side, and their density rapidly decays to zero with increasing heights. While the species densities of course determine the whole plasma environment and electrostatic fields, *per se* they only come into play in the E-Glider model when calculating current collection and power expenditure. The electrostatic potential, on the other hand, directly influences the spacecraft dynamics, and it is therefore perhaps the most important result of the simulations.

As shown in Fig. 2.15, the near-surface field resembles that of a dipole, which is consistent with the differential charging phenomenon, while the far field decays radially (except in the wake region). The surface potential is in the range of -20 V on the dark side and slightly negative on the sunlit side. This relatively low and negative potential on the sunlit side indicates that, at least for this size of the asteroid, the increased solar wind electron flux is sufficient to offset the positive charge generation caused by the photoelectrons. The potential profile obtained from the simulation also shows that the Nitter model implemented in the past (Nitter et al. 1998; Hartzell 2012) does not provide accurate estimates, especially for small asteroids, being derived from an infinite planar surface 1D model. The main drawbacks of this model are the inability to correctly capture the wake and its underestimation of the radial decay (Fig. 2.16). Both inaccuracies lead to an excessively optimistic and nonconservative estimate of the electric fields.

The electrostatic field intensity, of which the PIC-derived data is shown in Fig. 2.17, offers virtually the same data as the potential, but in a more easily readable form. The electrostatic force on the E-Glider is then calculated by interpolating the electric field obtained from USC-IFEPIC.

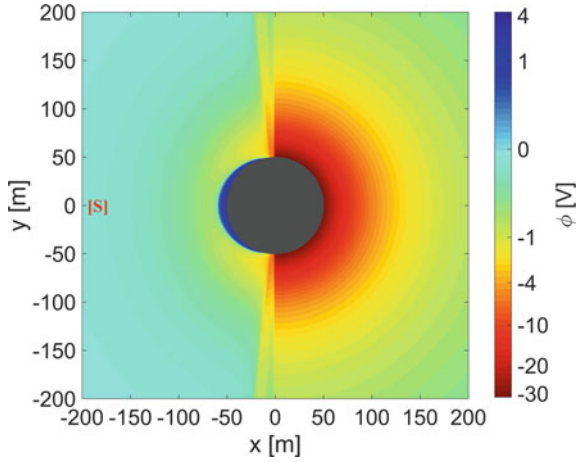


Fig. 2.16 Obsolete electrostatic potential calculated with the Nitter model

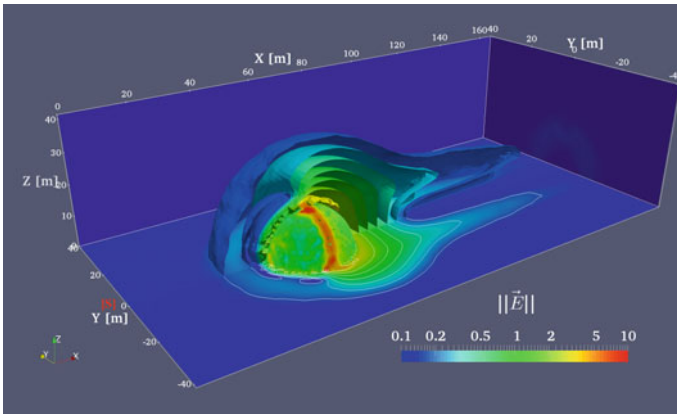


Fig. 2.17 Electrostatic field magnitude

2.9 Electrostatic Forces and Moments

The electrostatic acceleration can be expressed as

$$\mathbf{a}_E = \frac{\mathbf{F}_E}{M} = \frac{Q}{M} \frac{\partial \phi(\mathbf{r})}{\partial \mathbf{r}} \tag{2.38}$$

where \mathbf{F}_E is the electrostatic force, Q is the charge of an E-Glider, and ϕ is the electrostatic potential. Finally, the SRP acceleration is obtained from the following equation:

$$\mathbf{a}_{SRP} = a_{SRP} \cdot \hat{\mathbf{d}} = \frac{(1 + \zeta)P_0 A \hat{\mathbf{d}}}{d^2 M} \quad (2.39)$$

where $P_0 \simeq 1 \times 10^{17}$ kg.m/s² is the solar flux constant; d is the distance from the Sun expressed in AU; and $\zeta \equiv C_s + 2/3C_d$ is the reflectivity of the surface of a spacecraft. This model, which is the so-called cannonball model, assumes that a spacecraft has a spherical shape and that the SRP force consists of only a radial component. It should be also noted that the effect of the solar eclipse is not included in this SRP model.

The electrostatic torque applied to a spacecraft is defined by the equation below.

$$\mathbf{T}_E = \int \boldsymbol{\rho} \times d\mathbf{F}_E(\mathbf{R}) \quad (2.40)$$

Here, \mathbf{R} denotes the relative position vector of a mass element with respect to the center of mass of the small body. \mathbf{R} can be expressed as follows:

$$\mathbf{R} = \mathbf{r} + \boldsymbol{\rho} \quad (2.41)$$

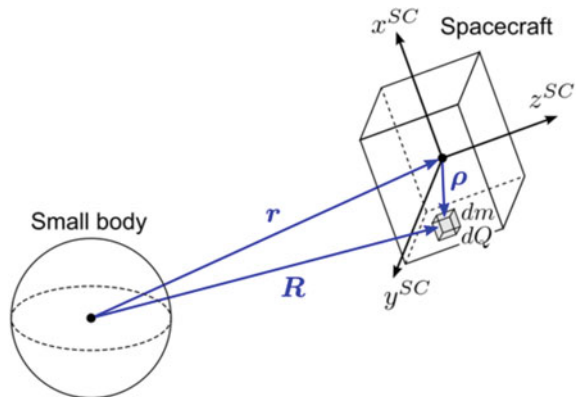
where \mathbf{r} is the relative position vector of the center of mass of the spacecraft relative to that of the small body, and $\boldsymbol{\rho}$ is the relative position vector of the element relative to the center of mass of the spacecraft. Considering the vector form of the Taylor series, the following equation can be derived when $\boldsymbol{\rho} \ll \mathbf{r}$:

$$f(\mathbf{R}) = f(\mathbf{r} + \boldsymbol{\rho}) \simeq f(\mathbf{r}) + \nabla f(\mathbf{r}) \cdot \boldsymbol{\rho} \quad (2.42)$$

where $\nabla = \partial/\partial \mathbf{r}$ (see Fig. 2.18).

The electrostatic potential ϕ_E is expressed as a function of the radial distance r and the solar incident angle α . Therefore, the partial derivative of the electrostatic potential is given by the equation below.

Fig. 2.18 Mass element of the spacecraft modeled as a rigid body



$$\frac{\partial \phi_E}{\partial \mathbf{r}} = \frac{\partial \phi_E}{\partial r} \mathbf{e}_r + \frac{1}{r} \frac{\partial \phi_E}{\partial \alpha} \mathbf{e}_\alpha \quad (2.43)$$

where \mathbf{e}_r is a unit vector along the radial direction, and \mathbf{e}_α is a unit vector along a transverse direction that exists in the same plane as the subsolar line and \mathbf{e}_r . According to the definitions, \mathbf{e}_r and \mathbf{e}_α are expressed as follows:

$$\begin{aligned} \mathbf{e}_r &= -\frac{\mathbf{r}}{r} \\ \mathbf{e}_\alpha &= -\frac{\cos \alpha}{r \sin \alpha} \mathbf{r} - \frac{1}{\sin \alpha} \hat{\mathbf{d}} \end{aligned} \quad (2.44)$$

Note that $\hat{\mathbf{d}}$ is a unit vector pointing from the Sun to a small body, which satisfies $\mathbf{r} \cdot \hat{\mathbf{d}} = r \cos \alpha$. Then substitution of Eq. (2.44) into Eq. (2.43) yields the following equation:

$$\begin{aligned} \frac{\partial \phi_E}{\partial \mathbf{r}} &= \left(\frac{1}{r} \frac{\partial \phi_E}{\partial r} + \frac{\cos \alpha}{r^2 \sin \alpha} \frac{\partial \phi_E}{\partial \alpha} \right) \mathbf{r} - \frac{1}{r \sin \alpha} \frac{\partial \phi_E}{\partial \alpha} \hat{\mathbf{d}} \\ &\equiv f_1 \mathbf{r} + f_2 \hat{\mathbf{d}} \end{aligned} \quad (2.45)$$

where f_1 and f_2 are functions of r and α , which are calculated numerically based on the electrostatic potential model. In the same manner, the derivatives of f_1 and f_2 can be expressed as

$$\frac{\partial f_1}{\partial \mathbf{r}} \equiv -g_1 \mathbf{r} - g_2 \hat{\mathbf{d}}, \quad \frac{\partial f_2}{\partial \mathbf{r}} \equiv -g_2 \mathbf{r} - g_3 \hat{\mathbf{d}} \quad (2.46)$$

where g_1 , g_2 , and g_3 are also computed numerically. Consequently, the electrostatic torque can be obtained from Eqs. (2.40)–(2.46) as follows.

$$\begin{aligned} \mathbf{T}_E &= -\int \boldsymbol{\rho} \times \nabla \phi_E(\mathbf{R}) dQ \\ &= -\int \boldsymbol{\rho} \times (f_1(\mathbf{R})\mathbf{R} + f_2(\mathbf{R})\hat{\mathbf{d}}) dQ \\ &\simeq -\int \boldsymbol{\rho} \times \{(f_1(\mathbf{r}) + \nabla f_1(\mathbf{r}) \cdot \boldsymbol{\rho})(\mathbf{r} + \boldsymbol{\rho}) + (f_2(\mathbf{r}) + \nabla f_2(\mathbf{r}) \cdot \boldsymbol{\rho})\hat{\mathbf{d}}\} dQ \\ &= -\int \boldsymbol{\rho} \times \{(f_1 + g_1 \mathbf{r} \cdot \boldsymbol{\rho} - g_2 \hat{\mathbf{d}} \cdot \boldsymbol{\rho})(\mathbf{r} + \boldsymbol{\rho}) + (f_2 + g_2 \mathbf{r} \cdot \boldsymbol{\rho} - g_3 \hat{\mathbf{d}} \cdot \boldsymbol{\rho})\hat{\mathbf{d}}\} dQ \\ &\simeq g_1 \mathbf{r} \times \mathbf{J} \mathbf{r} + g_2 (\mathbf{r} \times \mathbf{J} \hat{\mathbf{d}} + \hat{\mathbf{d}} \times \mathbf{J} \mathbf{r}) + g_3 \hat{\mathbf{d}} \times \mathbf{J} \hat{\mathbf{d}} \end{aligned} \quad (2.47)$$

where \mathbf{J} is the tensor defined by the equation below.

$$\mathbf{J} \equiv \int (|\boldsymbol{\rho}|^2 \mathbf{E} - \boldsymbol{\rho} \boldsymbol{\rho}^T) dQ \quad (2.48)$$

When it is assumed that the mass distribution and the charge distribution of a spacecraft are identical, the following relationship holds.

$$\frac{dQ(\boldsymbol{\rho})}{Q} = \frac{dm(\boldsymbol{\rho})}{m} \quad (2.49)$$

Based on this assumption, the electrostatic torque in Eq. (2.47) can be rewritten as follows:

$$\mathbf{T}_E = \frac{Q}{m} \{g_1 \mathbf{r} \times \mathbf{I} \mathbf{r} + g_2 (\mathbf{r} \times \hat{\mathbf{I}} \hat{\mathbf{d}} + \hat{\mathbf{d}} \times \mathbf{I} \mathbf{r}) + g_3 \hat{\mathbf{d}} \times \hat{\mathbf{I}} \hat{\mathbf{d}}\} \quad (2.50)$$

where \mathbf{I} is the moment of inertia tensor defined by the equation below.

$$\mathbf{I} \equiv \int (|\boldsymbol{\rho}|^2 \mathbf{E} - \boldsymbol{\rho} \boldsymbol{\rho}^T) dm \quad (2.51)$$

The detailed electrostatic torque model derived in this subsection is used for numerical simulations.

Figure 2.12 shows that the gradient of the electrostatic potential is broadly directed to/from the center of the small body. Based on this observation, it can be approximated that the electrostatic force applied to a spacecraft has only a radial component, yielding the equation below.

$$\frac{\partial \phi_E}{\partial \alpha} \equiv 0 \quad (2.52)$$

Based on this approximation, the following equations hold.

$$g_1 = \Gamma, \quad g_2 = g_3 = 0 \quad (2.53)$$

Here, Γ can be numerically computed from the equation below.

$$\Gamma(r) \equiv \frac{1}{r^3} \frac{\partial \phi_E}{\partial r} - \frac{1}{r^2} \frac{\partial^2 \phi_E}{\partial r^2} \quad (2.54)$$

From Eqs. (2.50), (2.52), and (2.53), the electrostatic torque can be approximated by the following simplified form:

$$\mathbf{T}_E \simeq \frac{Q\Gamma}{m} \mathbf{r} \times \mathbf{I} \mathbf{r} \quad (2.55)$$

Comparing Eqs. (2.31) and (2.55), it is evident that the electrostatic torque is expressed in the same form as the gravity gradient torque. This similarity enables analytical analyses for the attitude motion of an E-Glider by extending conventional analysis methods. The simplified electrostatic torque model derived in this subsection is used for analytical studies.

2.9.1 Finite Element Electrostatic Force Modeling

To include the plasma wake effects is one of the biggest issues in modeling the electrostatic field, and all the analytical models seem to be inadequate for this purpose. For this reason, the electrostatic field and the electrostatic potential are external input data. These data represent the electrostatic field and potential sampled on a 3D Cartesian mesh of points. The data (both for the electrostatic potential and the electrostatic field) are given in the RIC reference frame, and they are time-invariant (Corradino 2018) by assumption. The time-invariant nature of the data is a strong assumption, but otherwise, the implementation of a dynamic model for the data will result in a high computational load, and it will imply also the need for storage capabilities for a huge amount of data.

The computation of both the force \mathbf{f}_e and the torque \mathbf{T}_e exerted on the spacecraft is possible by starting from the input files containing the values of the electrostatic field of the asteroid under analysis.

We can consider the spacecraft as an extended body characterized by a net charge and first and second moment of charge, or as an ensemble of parts, each one characterized by a net charge localized in the center of charge of the part (which could be not coincident with the center of mass of the part itself). If the assumption of a unique extended body is valid for the spacecraft, the electric field is linear by assumption (as done for the gravity field in a previous section); thus, by defining ρ , the electric field can be expressed as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}(\mathbf{r}_0) + \mathbf{G}_e(\mathbf{r}_0)\rho \quad (2.56)$$

in which $\mathbf{G}_e(\mathbf{r}_0)$ is the electrostatic field gradient computed in the center of mass of the spacecraft. The derivation of an analytical expression is not possible for the electrostatic field gradient; thus, it is numerically computed starting from the data of the electrostatic field. The tensor $\mathbf{G}_e(\mathbf{r}_0)$ is obtained by taking the vector of the gradient of the electrostatic field along the three directions shaped column-wise into a matrix. Hence, the electrostatic force is

$$\begin{aligned} \mathbf{f}_e &= \int_{S/C} [\mathbf{E}(\mathbf{r}_0) + \mathbf{G}_e(\mathbf{r}_0)\rho] dq = \mathbf{E}(\mathbf{r}_0) \int_{S/C} dq + \mathbf{G}_e(\mathbf{r}_0) \int_{S/C} \rho dq \\ &= \mathbf{E}(\mathbf{r}_0)q + \mathbf{G}_e(\mathbf{r}_0)\mathbf{S}_q \end{aligned} \quad (2.57)$$

The left integral cannot be canceled out because, as specified before, the center of charge may not coincide with the center of mass of the spacecraft. In analogy with the mass-related cases, this term is the first moment of charge \mathbf{S}_q about the center of mass. This term is negligible if the center of mass and the center of charge are coincident. If the spacecraft is an ensemble of i parts, the summation of the force acting on each part is the total force given by the electrostatic field. By naming \mathbf{r}_i the position of the center of mass of each part, the total electrostatic force is

$$\mathbf{f}_e = \sum_i \left(\int_{i\text{-thpart}} \mathbf{E}(\mathbf{r}_i) dq_i \right) = \sum_i \left(\mathbf{E}(\mathbf{r}_i) \int_{i\text{-thpart}} dq_i \right) = \sum_i \left(\mathbf{E}(\mathbf{r}_i) q_i \right) \quad (2.58)$$

As for the previous case, this method allows us to achieve a higher accuracy (especially in those cases in which the spacecraft is very extended and thus the linear approximation is no more valid) but has as a drawback higher computational cost.

The torque due to the electric field on an extended spacecraft with respect to the center of mass of the spacecraft itself is

$$\mathbf{T}_e = \int_{S/C} \boldsymbol{\rho} \times \mathbf{E}(\mathbf{r}) dq \quad (2.59)$$

By considering the spacecraft as a single extended body, the linear approximation for the local variation of the electric field is valid; thus, Eq. (2.59) can be rewritten as

$$\begin{aligned} \mathbf{T}_e &= \int_{S/C} \boldsymbol{\rho} \times [\mathbf{E}(\mathbf{r}_0) + \mathbf{G}_e(\mathbf{r}_0)] dq \\ &= \int_{S/C} \boldsymbol{\rho} dq \times \mathbf{E}(\mathbf{r}_0) + \int_{S/C} \boldsymbol{\rho} \times \mathbf{G}_e(\mathbf{r}_0) \boldsymbol{\rho} dq \\ &= \mathbf{S}_q \times \mathbf{E}(\mathbf{r}_0) + \mathbf{T}_{eG} \end{aligned} \quad (2.60)$$

where \mathbf{T}_{eG} involves the computation of the second moment of charge \mathbf{I}_q , which is similar to the computation of the mass inertia tensor \mathbf{J} , with charges instead of masses. If the spacecraft is made of parts, the transport theorem can be applied to translate all the contributions of each part to a reference point, and then the summation of these contributions gives the second moment of charge of the entire spacecraft. \mathbf{T}_{eG} can be computed as the gravitational torque; thus,

$$\mathbf{T}_{eG} = \begin{bmatrix} G_{e,yz}(I_{q,zz} - I_{q,yy}) + G_{e,xz}I_{q,xy} - G_{e,xy}I_{q,xz} + I_{q,yz}(G_{e,zz} - G_{e,yy}) \\ G_{e,xz}(I_{q,xx} - I_{q,zz}) - G_{e,yz}I_{q,xy} + G_{e,xy}I_{q,yz} + I_{q,xz}(G_{e,xx} - G_{e,zz}) \\ G_{e,xy}(I_{q,yy} - I_{q,xx}) + G_{e,yz}I_{q,xz} - G_{e,xz}I_{q,yz} + I_{q,xy}(G_{e,yy} - G_{e,xx}) \end{bmatrix} \quad (2.61)$$

with $\mathbf{G}_e(\mathbf{r}_0)$ computed in the center of mass. If the spacecraft is an ensemble of parts by assumption, the total torque acting on the spacecraft is the summation of the torques generated by the electrostatic field on each part i . By calling \mathbf{r}_i the position of the center of mass of the i -th part, Eq. (2.59) becomes

$$\mathbf{T}_e = \sum_i (\boldsymbol{\rho}_i \times \mathbf{E}(\mathbf{r}_i) q_i) \quad (2.62)$$

This last formulation offers a more accurate estimation of the torque given by the electrostatic field, but the computational load is higher, as in the previous cases.

2.10 Electrostatic Orbiting and its Stability

Dayside equilibrium points can be created by inducing electrostatic force; however, electrostatic hovering at such an equilibrium point will consume a large amount of power. For this reason, this section proposes the electrostatic orbiting method as an alternative strategy for an E-Glider operation and identifies a new class of periodic orbits around asteroids called *electrostatic periodic orbits*.

2.10.1 Orbit Design Methodology

Electrostatic periodic orbits are designed by using the symmetry inherent in the equations of motion, Eq. (2.9), which can be expressed as follows (Broschart et al. 2009; Hénon 1969):

$$(t, x, y, z) \rightarrow (-t, x, -y, z) \quad (2.63)$$

If the set of variables on the left-hand side of Eq. (2.63) satisfy Eq. (2.9), then that on the right-hand side also satisfies the equation. This symmetry is known to hold for the circular restricted three-body problems subject to SRP, and it holds for an E-Glider system as well because the electrostatic potential is assumed to have symmetry about the x axis. Because of the symmetry, if an initial position on the x - z plane is given as $\mathbf{r} = [x_0, 0, z_0]^T$ and an initial velocity perpendicular to this plane is given as $\dot{\mathbf{r}} = [0, \dot{y}_0, 0]^T$, then the spacecraft trajectories obtained through forward and backward propagation are symmetrical to each other about the x - z plane (Fig. 2.19a). Thus, when a spacecraft perpendicularly intersects the x - z plane again, a periodic orbit solution is obtained as a closed continuous trajectory (Fig. 2.19b).

Then a set of initial conditions, with three degrees of freedom, are expressed as (x_0, z_0, \dot{y}_0) . On the other hand, terminal constraints, $\dot{x} = \dot{z} = 0$, must be satisfied when a trajectory intersects the x - z plane after half a period. Consequently, an electrostatic periodic orbit solution, which is obtained by numerical computation, has one degree of freedom. To systematically analyze the solution space, an initial altitude h_0 and an initial phase ψ_0 , which are alternative parameters for describing the initial position in place of x_0 and z_0 , are introduced as follows:

$$\begin{aligned} h_0 &= \sqrt{x_0^2 + z_0^2} - R \\ \psi_0 &= \tan^{-1} \left(-\frac{z_0}{x_0} \right) \end{aligned} \quad (2.64)$$

Note that $\psi_0 = 0$ and 90° correspond to the subsolar point and the terminator point, respectively. Among the three initial variables (h_0, ψ_0, \dot{y}_0) , an initial altitude h_0 is designated as a free parameter to search for periodic orbit solutions.

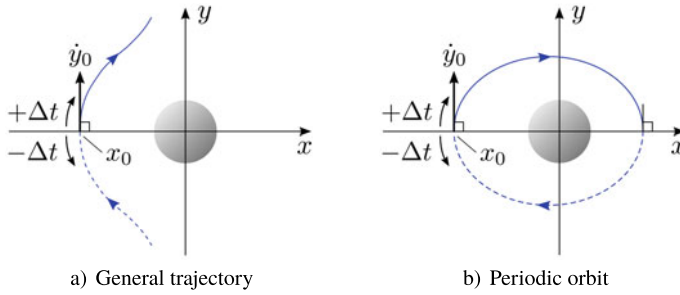


Fig. 2.19 Orbit design methodology of electrostatic periodic orbits

2.10.2 Electrostatic Periodic Orbit

Figure 2.20 provides an example of a natural periodic orbit (i.e., $Q = 0$), which is commonly referred to as a terminator orbit, and Fig. 2.21 provides examples of electrostatic periodic orbits for two different charge levels. These orbits are obtained with $h_0 = 15$ m. Note that these figures are expressed in the Hill coordinate, and thus, the negative direction of the x axis corresponds to the direction of the Sun. According to the definition of the coordinate system, these orbits can also be classified as Sun-synchronous orbits that do not experience an eclipse. Here, the periods of the orbits shown in Figs. 2.20, 2.21a, and b are $T = 3.5, 4.9,$ and 7.3 hr, respectively.

Figure 2.20 shows that the orbital plane is displaced from the terminator plane in the anti-Sun direction due to the effect of SRP. This observation indicates that this natural periodic orbit is located on the nightside of the asteroid; thus, it is not suitable for optical observations. This is the primary drawback of terminator orbits around asteroids. By contrast, as shown in Fig. 2.21, electrostatic periodic orbits are located on the dayside. Therefore, these orbits offer a significant advantage for optical observations. Moreover, these orbits are Sun-synchronous and achieve constant illumination from the Sun, which is advantageous for solar power generation and thermal design. Broadly speaking, when the magnitude of a charge increases, an orbit achieves larger displacement from the terminator plane in the direction of the Sun, as depicted in Fig. 2.21a and b. Another important fact is that these orbits are accomplished by inducing negative charging, thereby requiring only a small amount of power, as will be pointed out in Sect. 2.15.3.

Figure 2.22 illustrates the history of the magnitudes of forces acting on a spacecraft during one orbital period in the orbit provided in Fig. 2.21a. The magnitude of the electrostatic force was computed as $1\text{--}10 \mu\text{ N}$ in this simulation. While the electrostatic force is weaker than the gravitational force, it has the same order of magnitude as that of the SRP force. This result indicates that an electrostatic periodic orbit with displacement in the direction of the Sun can be achieved without fully compensating for the gravitational force, leading to energy-efficient operation compared with electrostatic hovering.

2.10.3 Evolution of Periodic Orbit Solutions

Shape transitions of electrostatic periodic orbits are depicted in Fig. 2.23. These orbits are computed for different charge values, $-200 \mu\text{C} \leq Q \leq -3 \mu\text{C}$, and a constant initial altitude, $h_0 = 15 \text{ m}$, by applying the numerical continuation method (Seydel 2009). The vertical axis represents an initial phase ψ_0 obtained as a result of the numerical calculation. This figure shows intriguing structures of both the entire solution space and orbital shapes themselves. The orbit that is expressed as the diamond marker at $Q = 0$ corresponds to the natural periodic orbit, which is also shown in Fig. 2.20, and it has an initial phase larger than 90° . By contrast, all of the electrostatic periodic orbit solutions depicted in this figure are obtained with initial phases smaller than 90° . It can be inferred from this result that these electrostatic periodic orbits are placed on the dayside, unlike natural terminator orbits. Interestingly, bifurcation appears in the region with a comparatively small magnitude of charge, and it involves several different orbit solutions with exactly the same charge value. As already mentioned, an orbit with a larger magnitude of a charge appears to have larger displacement from the terminator plane.

Figure 2.24 illustrates electrostatic periodic orbits computed for different initial altitudes, $10 \text{ m} \leq h_0 \leq 80 \text{ m}$, and a constant charge, $Q = -50 \mu\text{C}$. The vertical axis represents an initial velocity \dot{y}_0 obtained as a result of the numerical calculation. As observed from the figure, a higher initial altitude does not necessarily result in

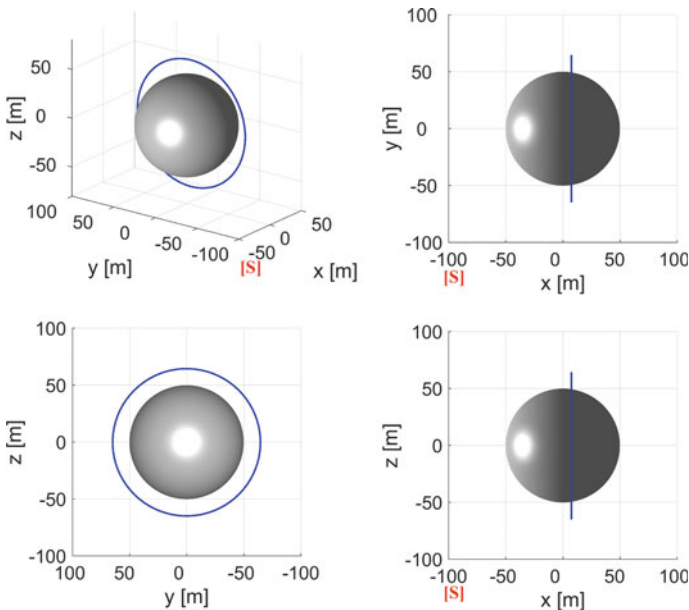


Fig. 2.20 Natural periodic orbit

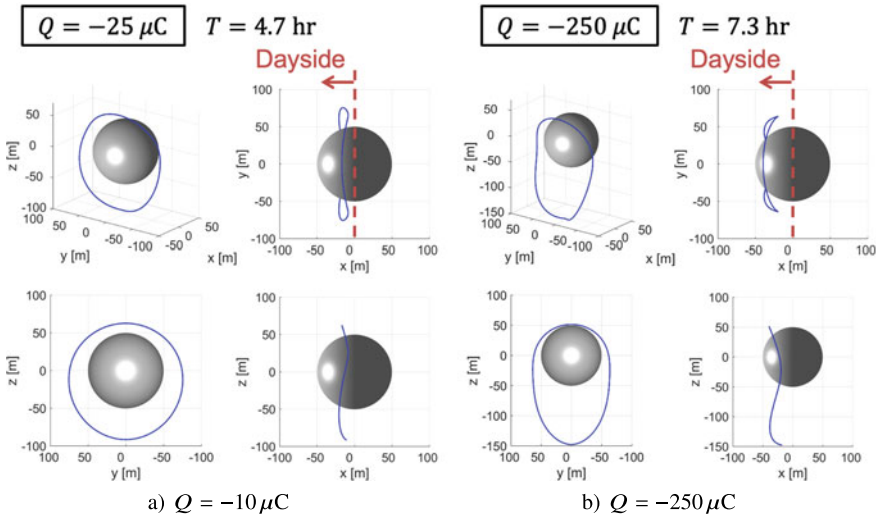


Fig. 2.21 Electrostatic periodic orbits

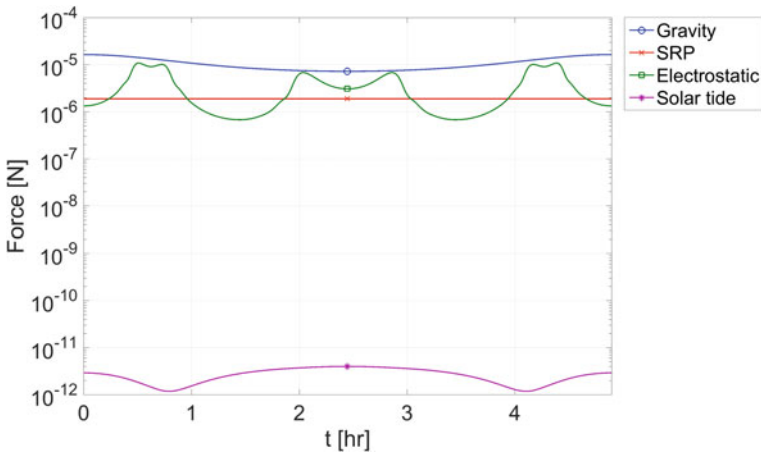


Fig. 2.22 Forces acting on the spacecraft during one orbital period

a larger periodic orbit. Moreover, the orbits on the left side and the right side are almost symmetric to each other about the $x-z$ plane. This result implies that the size of an electrostatic periodic orbit is limited by the charge level because electrostatic force cannot exert influence on the motion of a spacecraft at a high altitude.

It is to be noted that Figs. 2.23 and 2.24 merely show examples of electrostatic periodic orbit families but not the entire orbit solutions. There probably exist other orbit families that are not presented in this study, because multiple equilibrium points are present in this system.

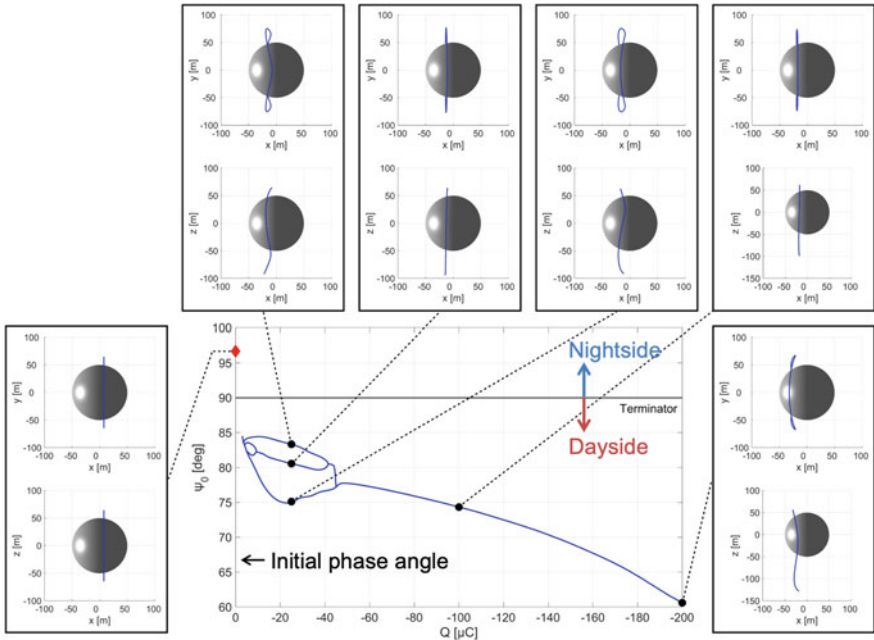


Fig. 2.23 Electrostatic periodic orbit solutions for different charge levels

2.10.4 Effects of Shape Irregularity

Analysis results that have been presented in previous sections are based on the spherical asteroid model. However, asteroids have irregular shapes in general, and the motion of a spacecraft around such an asteroid can be strongly perturbed due to its irregular gravitational field. Moreover, considering an E-glider system, irregularly shaped asteroids form irregular electrostatic fields around them, posing an additional perturbation on the spacecraft. This section evaluates the effects of these perturbations on spacecraft dynamics.

2.10.4.1 Irregular Electrostatic Field

Electrostatic potential around an asteroid is obtained from Eq. (2.35) as a function of the altitude h and the solar incident angle θ . In the case of a spherical asteroid, there exists an explicit relationship between (h, θ) and the position vector \mathbf{r} , as presented in Eq. (2.34). On the other hand, in the case of an ellipsoidal asteroid, there is no such explicit expression because the position vector and the normal vector to the surface are not parallel, as shown in Fig. 2.8. Therefore, this subsection derives the implicit relationship (h, θ) and the position vector \mathbf{r} to compute the electrostatic potential around an ellipsoid.

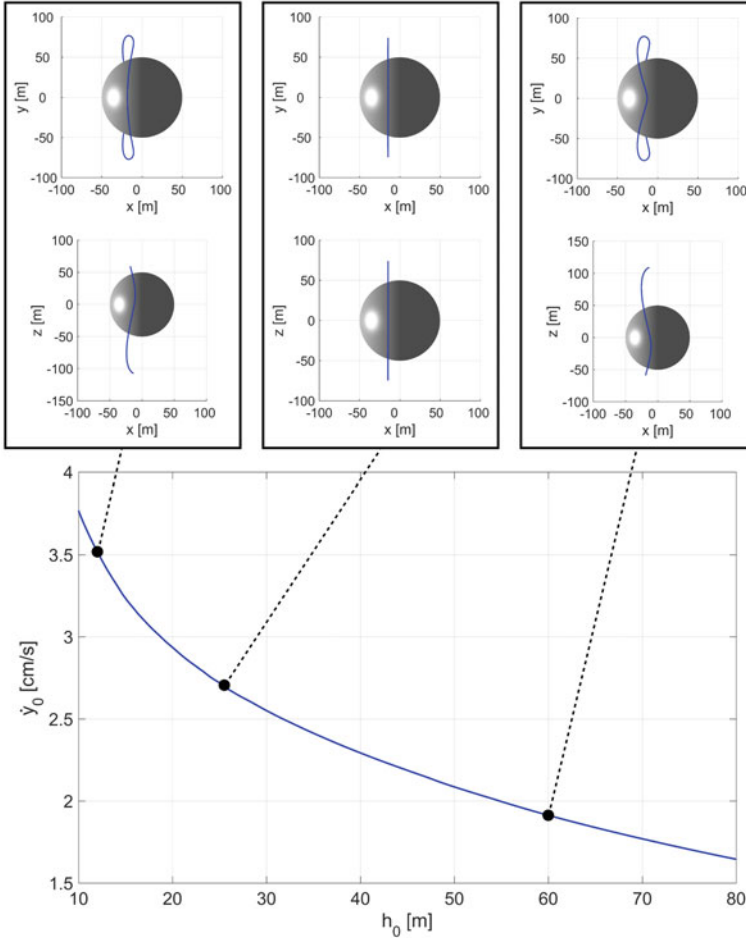


Fig. 2.24 Electrostatic periodic orbit solutions for different initial distances

An arbitrary position on the surface of the ellipsoid is defined as

$${}^B \mathbf{R}_s = [x_s, y_s, z_s]^T \tag{2.65}$$

Then these position variables satisfy the equation below.

$$f(x_s, y_s, z_s) = \frac{x_s^2}{R_a^2} + \frac{y_s^2}{R_b^2} + \frac{z_s^2}{R_c^2} - 1 = 0 \tag{2.66}$$

The normal vector with respect the surface of the ellipsoid at (x_s, y_s, z_s) can be derived as follows:

$${}^B \mathbf{n} = \left[\frac{\partial f}{\partial x_s}, \frac{\partial f}{\partial y_s}, \frac{\partial f}{\partial z_s} \right]^T = \left[\frac{2x_s}{R_a^2}, \frac{2y_s}{R_b^2}, \frac{2z_s}{R_c^2} \right]^T \quad (2.67)$$

$${}^B \hat{\mathbf{n}} = \frac{{}^B \mathbf{n}}{|{}^B \mathbf{n}|}$$

where $\hat{\mathbf{n}}$ represents a unit normal vector. The position vector can be expressed by the following equation:

$${}^B \mathbf{r} = {}^B \mathbf{R}_s + h \cdot {}^B \hat{\mathbf{n}} \quad (2.68)$$

The coordinate transformation from the asteroid body-fixed coordinate to the Hill coordinate yields the equations below.

$${}^H \mathbf{r} = {}^H \mathbf{C}_B {}^B \mathbf{r}, \quad {}^H \hat{\mathbf{n}} = {}^H \mathbf{C}_B {}^B \hat{\mathbf{n}} \quad (2.69)$$

Finally, the solar incident angle can be calculated as

$$\theta = \cos^{-1} (-{}^H \hat{\mathbf{n}} \cdot {}^H \hat{\mathbf{d}}) \quad (2.70)$$

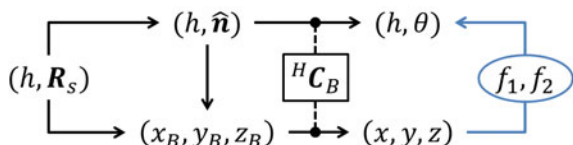
where ${}^H \hat{\mathbf{d}} = [1, 0, 0]^T$. Based on these equations, h and θ can be calculated implicitly, which can be expressed as follows:

$$\begin{aligned} h &= f_1(t, x, y, z) \\ \theta &= f_2(t, x, y, z) \end{aligned} \quad (2.71)$$

Note that these implicit functions are time-dependent because the coordinate transformation ${}^H \mathbf{C}_B$ is a function of the asteroid rotation phase θ_{rot} . Once h and θ are obtained from Eq. (2.71), the electrostatic potential around an ellipsoidal asteroid can be computed based on Eq. (2.35). The calculation process described above is presented in Fig. 2.25.

Figure 2.26 provides simulation results of the electrostatic potential around an asteroid modeled as a triaxial ellipsoid with an axis ratio of $R_a : R_b : R_c = 2.0 : 1.5 : 1.0$. The electrostatic potentials were calculated for four different rotation angles. These figures demonstrate that a time-varying irregular electrostatic field has been successfully simulated based on the proposed method. It appears that the structure of the electrostatic potential changes dynamically in accordance with the rotation phase of the asteroid. Moreover, this analysis method is performed by mapping an electrostatic potential from a spherical coordinate to an ellipsoidal coordinate based

Fig. 2.25 Process of calculating the altitude and the solar incident angle for an ellipsoid



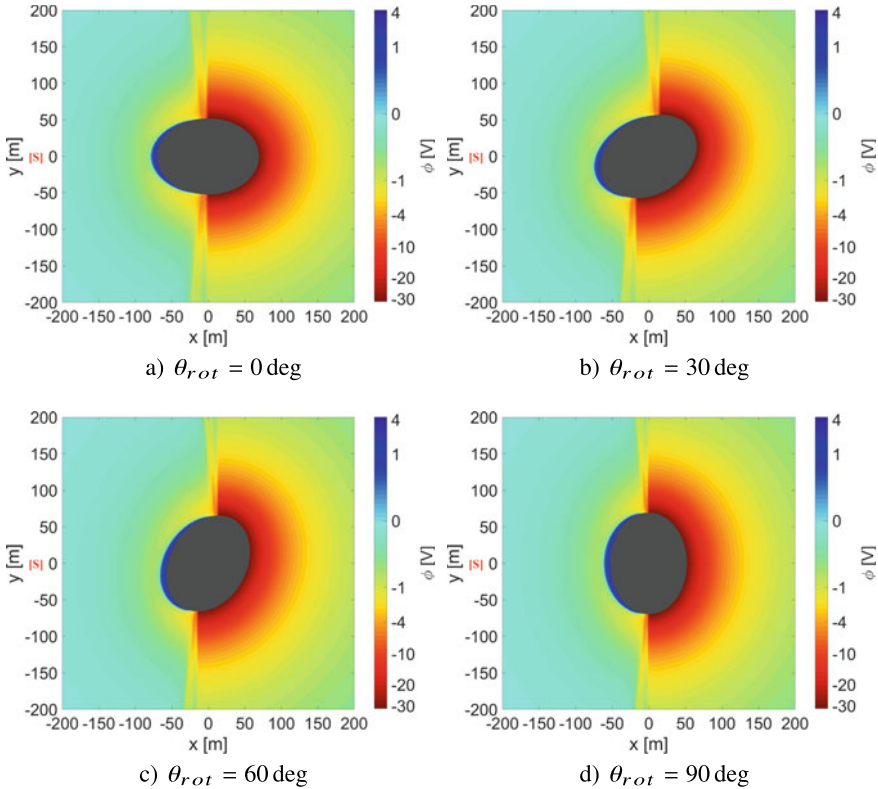


Fig. 2.26 Electrostatic potential around the ellipsoidal asteroid

on the geometrical relationship between them; therefore, it can emulate the time-varying behavior with relatively low computational cost compared with the classical particle-in-cell method.

2.10.4.2 Orbital Motion around an Irregularly Shaped Asteroid

Figure 2.27 provides the simulation results of the orbital motion of an E-glider around an ellipsoidal asteroid. The initial position and velocity used in these simulations are that of the periodic orbit solution around a spherical asteroid which is depicted in Fig. 2.21a. The equation of motion is the same as the one used for a spherical asteroid, but the gravitational potential U_G and the electrostatic potential ϕ are replaced by the models incorporating irregularly shaped effects, as discussed in Sects. 2.6.5 and 2.10.4.1. Note that the directions of asteroids illustrated in Fig. 2.27 merely show the initial states of them, and the asteroids are rotating with respect to the Hill coordinate.

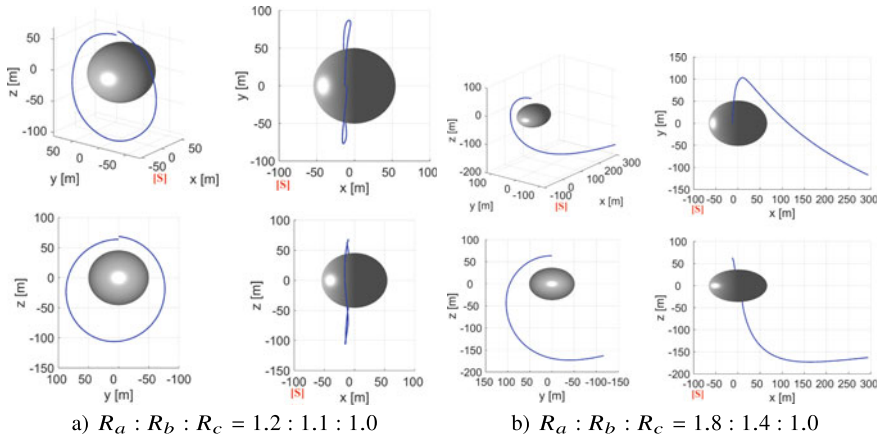


Fig. 2.27 Orbital motion around an irregularly shaped asteroid

Figure 2.27a shows the case for an asteroid with a relatively small oblateness, such as Benu and Ryugu (1999 JU3) (Nolan et al. 2013; Bellerose and Yano 2010). Although the simulated orbit is perturbed from the reference orbit, the position of the spacecraft after one period is close to the initial position. This result demonstrates that electrostatic orbits obtained for a spherical asteroid can serve as good approximations around a nearly spherical asteroid. On the other hand, the simulation result for an asteroid with a highly irregular shape, such as Itokawa (Fujiwara et al. 2006), is depicted in Fig. 2.27b. It is evident that the spacecraft escapes from the asteroid and is pushed away in the anti-Sun direction by the SRP.

In conclusion, it has been demonstrated that the motion around an irregularly shaped asteroid is perturbed because of the irregular gravitational and electrostatic field effects, and the perturbations might cause escape or collision in the worst-case scenario. This problem can be solved with two different approaches. The first one is to redesign a reference orbit by taking into account the effects of the shape irregularity. The other approach would be the implementation of feedback control of the electrostatic force. The magnitude of forces acting on the spacecraft orbiting around the ellipsoidal asteroid is presented in Fig. 2.28, which corresponds to the simulation provided in Fig. 2.27a. Here, the labels “J2” and “J4” represent the higher-order gravitational forces due to the $J_2 (= -C_{20})$ and $J_4 (= -C_{40})$ terms, respectively. As observed from this figure, the electrostatic force is stronger than the higher-order gravity and the SRP force. This result implies the perturbations can potentially be compensated for by applying the feedback control of electrostatic force via the spacecraft charge.

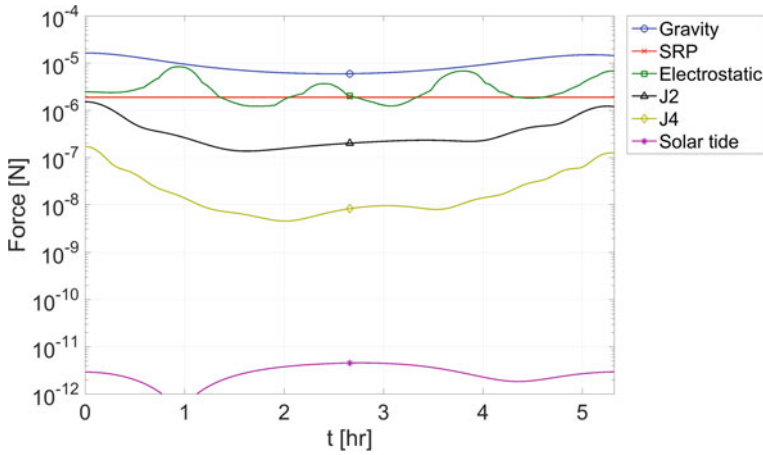


Fig. 2.28 Forces acting on the spacecraft orbiting around an irregularly shaped asteroid

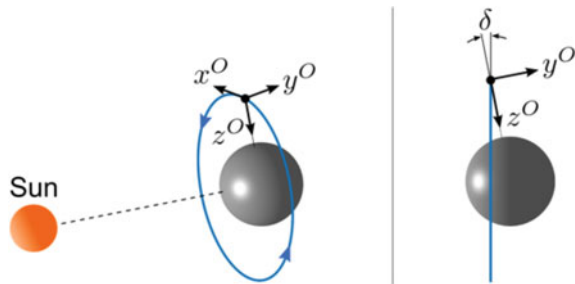
2.11 Attitude Stability

2.11.1 Linearized Euler Equation

The attitude motion of a spacecraft significantly depends on the orbit around a small body. This chapter assumes that a spacecraft is orbiting in a circular electrostatic periodic orbit with a slight displacement in the Sun’s direction. Then the orbital coordinate system can be defined as shown in Fig. 2.29.

The origin is at the center of the spacecraft; the z axis points in the direction of the center of the small body; the y axis is perpendicular to both the z axis and the velocity vector of the spacecraft; and the x axis completes a right-handed Cartesian coordinate system. Note that the direction of the x axis is identical to that of the velocity vector of the spacecraft when the orbit is circular. In terms of the orbital coordinate, the attitude of the spacecraft can be expressed by Euler angles (ϕ, θ, ψ) , considering a 2–1–3 rotation sequence from the orbital coordinate to the body-fixed

Fig. 2.29 Orbital coordinate system



coordinate. Note that the x , y , and z axes are called the roll, pitch, and yaw axes, respectively.

The angular velocity vector is expressed in the spacecraft-fixed frame as

$${}^{SC}\boldsymbol{\omega}_{SC/I} = {}^{SC}\boldsymbol{\omega}_{SC/O} + {}^{SC}\mathbf{C}_O^O \boldsymbol{\omega}_{O/I} \quad (2.72)$$

where ${}^{SC}\mathbf{C}_O$ denotes the rotational transformation from the orbital coordinate to the spacecraft-fixed coordinate given by the equation below.

$${}^{SC}\mathbf{C}_O = \mathbf{R}_z(\psi)\mathbf{R}_x(\phi)\mathbf{R}_y(\theta) \quad (2.73)$$

Considering a 2–1–3 rotation sequence, the angular velocity vector of the spacecraft-fixed frame relative to the orbital coordinate can be calculated from the following equation (Hughes 1986):

$${}^{SC}\boldsymbol{\omega}_{SC/O} = \begin{bmatrix} \dot{\phi} - \dot{\psi} \sin \theta \\ \dot{\theta} \cos \phi + \dot{\psi} \sin \phi \cos \theta \\ -\dot{\theta} \sin \phi + \dot{\psi} \cos \phi \cos \theta \end{bmatrix} \quad (2.74)$$

Assuming that the orbit offset angle illustrated in Fig. 2.29 is sufficiently small ($\delta \ll 1$) and that the mean motion of a small body is negligible compared with that of a spacecraft orbit ($N \ll n$), the angular velocity vector of the orbital coordinate relative to the inertial coordinate can be calculated from the following equation:

$${}^O\boldsymbol{\omega}_{O/I} \simeq -n \begin{bmatrix} 0 \\ \cos \delta \\ \sin \delta \end{bmatrix} \simeq -n \begin{bmatrix} 0 \\ 1 \\ \delta \end{bmatrix} \quad (2.75)$$

where n represents the mean motion of an orbit around a small body. Under the approximation that the mean motion of an electrostatic periodic orbit is identical to that of a Keplerian orbit with a radius of r , the following relationship is obtained.

$$n \simeq \sqrt{\frac{\mu}{r^3}} \quad (2.76)$$

The position vector \mathbf{r} and the moment of inertia tensor \mathbf{I} are expressed in the spacecraft-fixed frame as

$${}^{SC}\mathbf{r} = {}^{SC}\mathbf{C}_O^O \mathbf{r} = {}^{SC}\mathbf{C}_O \begin{bmatrix} 0 \\ 0 \\ -r \end{bmatrix} \quad (2.77)$$

$${}^{SC}\mathbf{I} = \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix} \quad (2.78)$$

Let the Euler angles and their derivatives be given as follows:

$$\phi = \bar{\phi}, \theta = \psi = \dot{\phi} = \dot{\theta} = \dot{\psi} = 0 \quad (2.79)$$

Here, $\bar{\phi}$ is a roll angle that satisfies $\bar{\phi} \ll 1$. Then by substituting Eqs. (2.72)–(2.79) into the Euler equation, it can be easily observed the equations of pitch motion about the y axis and yaw motion about the z axis are satisfied, which means that these motions are in equilibrium. The remaining equation about the roll axis must satisfy the equation below to achieve the equilibrium attitude.

$$\begin{aligned} (I_z - I_y)n^2(\bar{\phi} - \delta) + (I_z - I_y) \left(\frac{3\mu}{r^5} + \frac{Q\Gamma}{m} \right) r^2 \bar{\phi} &= 0 \\ \Leftrightarrow (I_z - I_y)n^2\{(4 + \xi)\bar{\phi} - \delta\} &= 0 \end{aligned} \quad (2.80)$$

Here, ξ is a nondimensional scalar value that represents the effect of electrostatic torque and is defined by the equation below.

$$\xi \equiv \frac{Qr^5\Gamma(r)}{m\mu} \quad (2.81)$$

By solving Eq. (2.80), the equilibrium roll angle is obtained from the next equation.

$$\bar{\phi} = \frac{\delta}{4 + \xi} \quad (2.82)$$

If the equilibrium attitude given by Eq. (2.79) is stable, an E-Glider system can achieve passive stabilization using gravity gradient torque and electrostatic torque. Given that the attitude of a spacecraft has a small deviation from its equilibrium state, the Euler angles can be expressed as $(\phi + \bar{\phi}, \theta, \psi)$, where $\phi, \theta, \psi \ll 1$. On the basis of this assumption, Eqs. (2.73) and (2.74) are approximated by the equation below:

$${}^{SC}\mathbf{C}_O \simeq \begin{bmatrix} 1 & \psi & -\theta \\ -\psi & 1 & \phi + \bar{\phi} \\ \theta & -(\phi + \bar{\phi}) & 1 \end{bmatrix}, \quad {}^{SC}\boldsymbol{\omega}_{SC/O} \simeq \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \quad (2.83)$$

Consequently, the following linearized Euler equations can be derived:

$$\begin{aligned} I_x \ddot{\phi} - (I_x - I_y + I_z)n\dot{\psi} + (4 + \xi)(I_y - I_z)n^2\phi &= 0 \\ I_y \ddot{\theta}(3 + \xi)(I_x - I_z)n^2\theta &= 0 \\ I_z \ddot{\psi} + (I_x - I_y + I_z)n\dot{\psi} - (I_x - I_y)n^2\psi &= 0 \end{aligned} \quad (2.84)$$

2.11.2 Stability Conditions

This subsection derives stability conditions of the attitude motion of an E-Glider based on the linearized equations of motion. Let the inertia ratio parameters be defined as

$$\sigma_1 = \frac{I_y - I_z}{I_x}, \quad \sigma_2 = \frac{I_x - I_z}{I_y}, \quad \sigma_3 = \frac{I_y - I_x}{I_z} \quad (2.85)$$

By using these parameters, Eq. (2.84) can be rewritten as the equation below.

$$\begin{aligned} \ddot{\phi} - (1 - \sigma_1)n\dot{\psi} + (4 + \xi)\sigma_1 n^2 \phi &= 0 \\ \ddot{\theta} + (3 + \xi)\sigma_2 n^2 \theta &= 0 \\ \ddot{\psi} + (1 - \sigma_3)n\dot{\phi} + \sigma_3 n^2 \psi &= 0 \end{aligned} \quad (2.86)$$

These linearized equations show that the roll motion and the yaw motion are coupled with each other, whereas the pitch motion is independent.

When the eigenvalue of this system is expressed as λ , the characteristic equation regarding the pitch motion is given by the equation below.

$$\lambda^2 + (3 + \xi)\sigma_2 n^2 = 0 \quad (2.87)$$

In the same manner, the characteristic equation regarding the roll and yaw motions is derived from the first and the third equations of Eq. (2.86), as expressed by the equation below.

$$\lambda^4 + \{1 + (3 + \xi)\sigma_1 + \sigma_1\sigma_3\}n^2\lambda^2 + (4 + \xi)\sigma_1\sigma_3 n^4 = 0 \quad (2.88)$$

Given the form of the characteristic equations presented in Eqs. (2.87) and (2.88), the pitch motion is stable when the eigenvalues are a conjugate pair of pure imaginary values, and the roll–yaw motion is stable when the eigenvalues have two conjugate pairs of pure imaginary values. Accordingly, the stability condition of the pitch motion is given by the following inequality, considering the relationship between the inertia ratio parameters.

$$(3 + \xi)\sigma_2 > 0 \Leftrightarrow (3 + \xi)(\sigma_1 - \sigma_3) > 0 \quad (2.89)$$

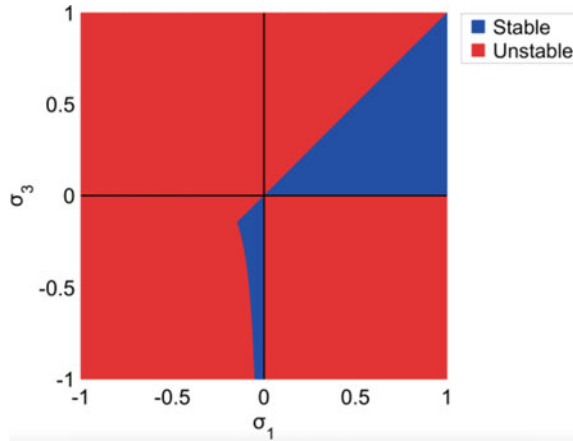
The stability conditions of the roll–yaw motion are given by the three inequalities below.

$$\begin{aligned} (4 + \xi)\sigma_1\sigma_3 &> 0 \\ 1 + (3 + \xi)\sigma_1 + \sigma_1\sigma_3 &> 0 \\ \{1 + (3 + \xi)\sigma_1 + \sigma_1\sigma_3\}^2 - 4(4 + \xi)\sigma_1\sigma_3 &> 0 \end{aligned} \quad (2.90)$$

Table 2.2 Classification of the stability domain based on the ξ value

Category	Condition
Class 1	$0 < \xi$
Class 2	$-3 < \xi < 0$
Class 3	$-4 < \xi < -3$
Class 4	$\xi < -4$

Fig. 2.30 The stability domain of the attitude motion subject to gravity gradient torque



Consequently, when the four inequalities given by Eqs. (2.89) and (2.90) are satisfied, the attitude motion of a spacecraft subject to the gravity gradient torque and the electrostatic torque exhibits stability. It is to be noted that these inequalities are expressed as the conditions on σ_1 and σ_3 . Another important fact is that the stability condition is dominated by the nondimensional parameter ξ . According to the definition given by Eq. (2.81), this parameter is an index of the magnitude of the electrostatic torque relative to the gravity gradient torque, and $\xi = 0$ corresponds to the classical attitude problem, in which the attitude motion is influenced solely by the gravity gradient torque (Hughes 1986).

In accordance with the ξ value, the stability behavior of the attitude motion of a spacecraft can be categorized into four types, as indicated in Table 2.2. Figure 2.30 shows the stability domain expressed in the σ_1 – σ_3 plane when the electrostatic torque is not considered. The blue region represents the cases where the spacecraft achieves stable libration around an equilibrium state. On the other hand, Fig. 2.31 illustrates the stability domains for the cases where the electrostatic is induced by using an E-Glider system. This figure shows examples of the four different types of stability diagrams, which are categorized based on Table 2.2. Comparing Figs. 2.30, 2.31a, and b, the class 1 system has a smaller stable region than that of the conventional system, while the class 2 system has a larger stable region. Intriguingly, the stability domain changes drastically in the class 3 and class 4 systems, as depicted in Fig. 2.31c

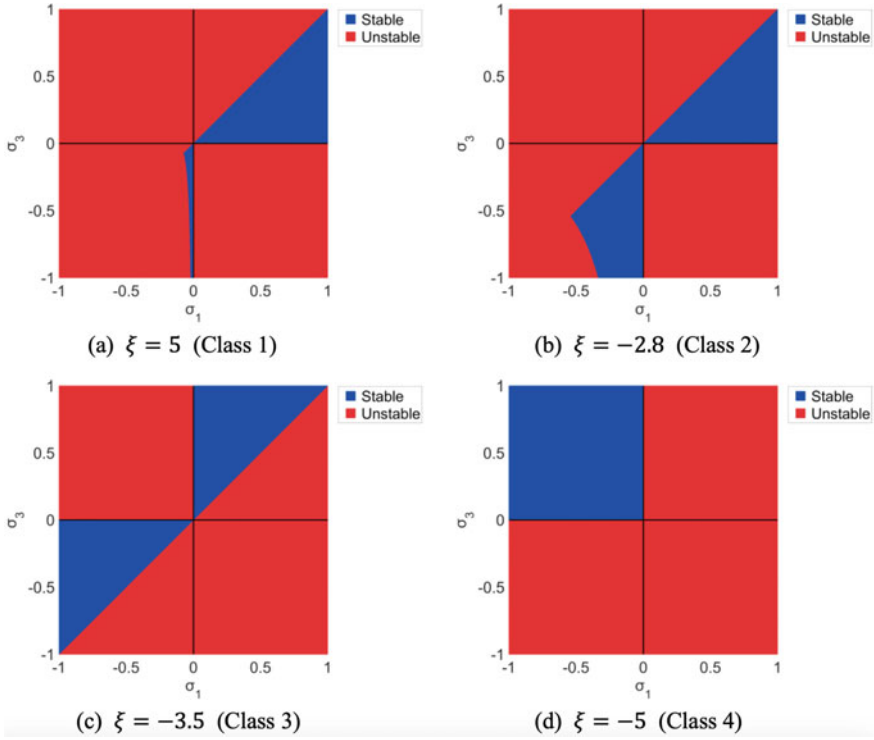


Fig. 2.31 The stability domain of the attitude motion subject to gravity gradient and electrostatic torques

and d. These results imply that the use of electrostatic torque enables a spacecraft to stabilize its attitude even when the spacecraft is originally unstable under the effect of the gravity gradient torque.

Since the attitude stability behavior of an E-Glider significantly depends on the ξ value, it is of great importance to analyze the possible range of ξ . As given by Eq. (2.87), the ξ value is a function of the charge Q and the distance r . Therefore, a contour map of ξ can be created in the $Q - h$ plane, as illustrated in Fig. 2.32. Here, h is the altitude, and the orbit offset angle is specified as $\sigma = 10^\circ$. This figure indicates that ξ varies widely enough to allow all of the possible four stability types. Based on this contour map and Table 2.2, the categorization of the attitude stability can be expressed in the $Q - h$ plane, as shown in Fig. 2.33. Each category is displayed in four different colors. This color map enables easy identification of the attitude stability behavior of a spacecraft orbiting in the corresponding electrostatic periodic orbit. A category transition appears at the altitude of approximately $h = 18$ m, which corresponds to the minimum-potential altitude of a non-monotonic sheath profile.

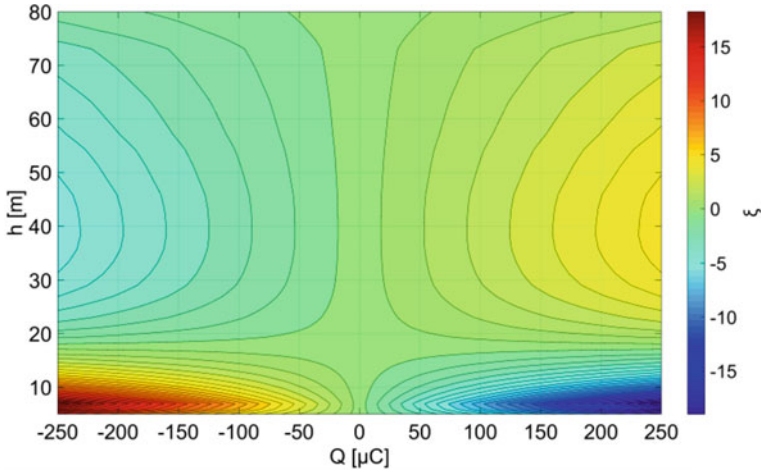


Fig. 2.32 Contour map of the ξ value

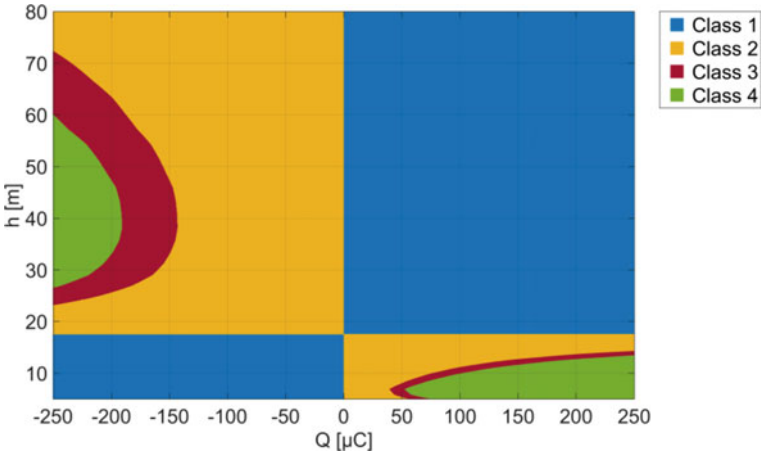


Fig. 2.33 Classification of the attitude stability based on the ξ value

2.11.3 Pitch Motion and Phase Diagram

Consider the situation where the offset of an orbital plane from the terminator plane is sufficiently small, and thus $\delta \simeq 0$ holds. In this case, if the roll and yaw motions are initially in equilibrium, these motions and the pitch motion do not influence each other. The independent pitch motion can be analyzed from the governing equations by substituting $\phi = \psi = \dot{\psi} = \dot{\phi} = 0$. This condition yields the equation below.

$${}^{SC}\mathbf{C}_O = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}, \quad {}^{SC}\boldsymbol{\omega}_{SC/I} = \begin{bmatrix} 0 \\ \dot{\theta} - n \\ 0 \end{bmatrix} \quad (2.91)$$

Consequently, the equation of pitch motion is derived from Euler's equation as follows:

$$\ddot{\theta} = (3 + \xi)\sigma_2 n^2 \sin \theta \cos \theta = 0 \quad (2.92)$$

By integrating this equation, the following equation is obtained:

$$\dot{\theta}^2 + (3 + \xi)\sigma_2 n^2 \sin^2 \theta = C_\theta \quad (2.93)$$

where C_θ is an integration constant. Let τ denote a time unit defined by the equation below.

$$\tau \equiv \frac{1}{n\sqrt{(3 + \xi)\sigma_2}} \quad (2.94)$$

By using this time unit, Eq. (2.93) can be normalized as

$$\left(\frac{d\theta}{d\hat{t}}\right)^2 + \sin^2 \theta = \hat{C}_\theta \quad (2.95)$$

where

$$\hat{t} = \frac{t}{\tau}, \quad \hat{C}_\theta = C_\theta \times \tau^2 \quad (2.96)$$

Once \hat{C}_θ is specified by an initial condition, the pitch angle θ and its change rate $d\theta/d\hat{t}$ are governed by Eq. (2.95). In other words, Eq. (2.95) describes the law of the conservation of energy. Note that it is assumed that Eqs. (2.89) and (2.90) are satisfied, and thus the attitude motion is in a stable state.

Figure 2.34 illustrates the phase plane plot for the pitch motion described by Eq. (2.95). This figure is displayed as a contour map of the \hat{C}_θ value. It is to be noted that the values in this figure are normalized by the time unit. A stable equilibrium appears at $\theta = 0^\circ$, and unstable equilibria appear at $\theta = \pm 90^\circ$. It can also be observed that the pitch motion exhibits two modes of behavior depending on the \hat{C}_θ value: libration motion when $\hat{C}_\theta < 1$ and tumbling motion when $\hat{C}_\theta > 1$. The boundary between these two modes is a separatrix defined as $\hat{C}_\theta = 1$, which is represented as a bold contour line in the figure. This phase plane does not depend on the system parameters because it is scaled by the time unit τ , which is a function of the ξ value. The proposed analytical method enables comprehending the mode of the pitch motion of an E-Glider subject to electrostatic torque.

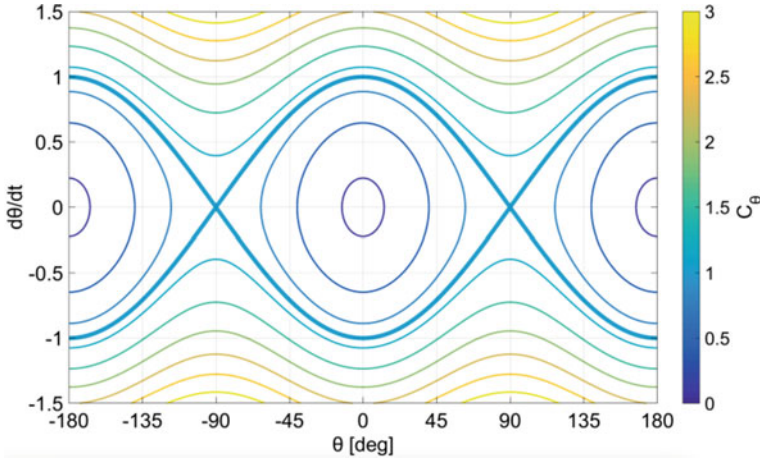


Fig. 2.34 Phase plane for the pitch motion

2.12 Coupled Orbital Attitude Stability

The stability of the attitude motion of a spacecraft subject to electrostatic torque has been analyzed using an analytical approach. Although these analyses enable an understanding of the dynamical structure of the attitude motion, they are performed based on linearization and approximations. This section, therefore, investigates the attitude motion through numerical simulations to verify the validity of the analytical theories established in the previous section.

2.12.1 Coupled Orbit–Attitude Equations of Motion

Let $(\mathbf{e}_{O,x}, \mathbf{e}_{O,y}, \mathbf{e}_{O,z})$ denote a set of unit vectors of the orbital coordinate. From the definition of the orbital coordinate, $\mathbf{e}_{O,x}$, $\mathbf{e}_{O,y}$, and $\mathbf{e}_{O,z}$ are calculated as

$$\mathbf{e}_{O,z} = -\frac{\mathbf{r}}{|\mathbf{r}|}, \quad \mathbf{e}_{O,y} = \frac{\mathbf{e}_{O,z} \times \mathbf{v}}{|\mathbf{e}_{O,z} \times \mathbf{v}|}, \quad \mathbf{e}_{O,x} = \mathbf{e}_{O,y} \times \mathbf{e}_{O,z} \quad (2.97)$$

By using these basis vectors, the rotational transformation matrix from the Hill coordinate system to the orbital coordinate system can be expressed as follows (Hughes 1986):

$${}^O C_H = [{}^H \mathbf{e}_{O,x} \quad {}^H \mathbf{e}_{O,y} \quad {}^H \mathbf{e}_{O,z}]^T \quad (2.98)$$

Considering a 2–1–3 rotation sequence from the orbital coordinate to the spacecraft-fixed coordinate, the attitude of a spacecraft is expressed by Euler angles (ϕ, θ, ψ) . Then the following kinematic equation is derived.

$$\begin{aligned}
\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} &= \begin{bmatrix} 1 & \sin \phi \tan \theta & \cos \phi \tan \theta \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi \sec \theta & \cos \phi \sec \theta \end{bmatrix} \begin{bmatrix} \tilde{\omega}_x \\ \tilde{\omega}_y \\ \tilde{\omega}_z \end{bmatrix} \\
&= \begin{bmatrix} 1 & \sin \phi \tan \theta & \cos \phi \tan \theta \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi \sec \theta & \cos \phi \sec \theta \end{bmatrix} \begin{bmatrix} \omega_x - \Omega_x \\ \omega_y - \Omega_y \\ \omega_z - \Omega_z \end{bmatrix}
\end{aligned} \tag{2.99}$$

where ${}^{SC}\boldsymbol{\omega}_{SC/I} = [\omega_x, \omega_y, \omega_z]^T$, ${}^{SC}\boldsymbol{\omega}_{SC/O} = [\tilde{\omega}_x, \tilde{\omega}_y, \tilde{\omega}_z]^T$, and ${}^{SC}\boldsymbol{\omega}_{O/I} = [\Omega_x, \Omega_y, \Omega_z]^T$. The angular velocity ${}^{SC}\boldsymbol{\omega}_{O/I}$ is given by the equation below.

$${}^{SC}\boldsymbol{\omega}_{O/I} = {}^{SC}\mathbf{C}_O ({}^O\boldsymbol{\omega}_{O/H} + {}^O\mathbf{C}_H^H \boldsymbol{\omega}_{H/I}) \tag{2.100}$$

Here, ${}^H\boldsymbol{\omega}_{H/I} = [0, 0, N]^T$ holds from the definition; ${}^{SC}\mathbf{C}_O$ and ${}^O\mathbf{C}_H$ are obtained from Eqs. (2.73) and (2.98), respectively; and ${}^O\boldsymbol{\omega}_{O/H}$, which is the angular velocity vector of the orbital coordinate with respect to the Hill coordinate, is obtained from the relationship below.

$$[{}^O\boldsymbol{\omega}_{O/H}]^x = {}^O\mathbf{C}_H ({}^O\dot{\mathbf{C}}_H)^T \tag{2.101}$$

${}^O\dot{\mathbf{C}}_H$ can be computed by numerically integrating the position and velocity of a spacecraft. Here, the notation $[\mathbf{u}]^x$ requires the formation of a skew-symmetric matrix from the elements of \mathbf{u} according to the following equation.

$$[\mathbf{u}]^x \equiv \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \tag{2.102}$$

The equation of attitude motion given by the Euler equation can be expressed in the spacecraft-fixed frame as follows:

$$\begin{bmatrix} I_x \dot{\omega}_x \\ I_y \dot{\omega}_y \\ I_z \dot{\omega}_z \end{bmatrix} = \begin{bmatrix} (I_y - I_z) \omega_y \omega_z \\ (I_z - I_x) \omega_z \omega_x \\ (I_x - I_y) \omega_x \omega_y \end{bmatrix} + {}^{SC}\mathbf{T}_G + {}^{SC}\mathbf{T}_E \tag{2.103}$$

The gravity gradient torque \mathbf{T}_G is calculated from Eq. (2.31), and the electrostatic torque \mathbf{T}_E is calculated from Eq. (2.50) based on the detailed model. Both \mathbf{T}_G and \mathbf{T}_E are dependent on the position of a spacecraft in addition to its attitude, and thus the attitude motion is coupled with the orbital motion. Consequently, Eqs. (2.9), (2.99), (2.103), and the relationship between position and velocity provide a total of 12 equations; thus, the following 12 variables can be calculated via numerical integration:

$$(x, y, z, \dot{x}, \dot{y}, \dot{z}, \phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}) \tag{2.104}$$

2.12.2 Stable and Unstable Attitude Motion

The attitude motion of an E-Glider is simulated for four cases that are listed in Table 2.3. Numerical simulations are performed for two different periodic orbits with two different sets of moment of inertia parameters. Figure 2.35 shows the reference electrostatic periodic orbits for the numerical simulations. The orbit illustrated in Fig. 2.35a has a ξ value of -0.18 , and thus the attitude stability mode is categorized as class 2, as shown in Fig. 2.36a. On the other hand, the orbit illustrated in Fig. 2.35b has a ξ value of -3.86 , exhibiting the class 3 stability mode, as presented in Fig. 2.37a. Figures 2.36b and 2.37b show that one set of moment of inertia parameters is a stable case and that the other set is unstable. Initial conditions are given such that the state variables are in the equilibrium state, but small errors are added in the initial ϕ and θ values.

The simulation results are presented in Fig. 2.38. The history of the Euler angles (ϕ, θ, ψ) is provided in this figure. Figure 2.38a and b correspond to the simulations for the orbit illustrated in Fig. 2.35a, c and d illustrate the results for the orbit shown in Fig. 2.35b. Figure 2.38a indicates that each Euler angle oscillates about an equilibrium state without diverging. Note that the broken line in the figure represents the

Table 2.3 Simulation conditions for stable and unstable attitude motions

	Charge	Altitude	ξ	Category	I_x, I_y, I_z	Stability
Case A	$-30 \mu\text{C}$	20 m	-0.18	Class 2	$4, 5, 2 \times 10^{-3} \text{Kg/m}^2$	Stable
Case B	$-30 \mu\text{C}$	20 m	-0.18	Class 2	$2, 5, 4 \times 10^{-3} \text{Kg/m}^2$	Unstable
Case C	$-280 \mu\text{C}$	58 m	-3.86	Class 3	$2, 5, 4 \times 10^{-3} \text{Kg/m}^2$	Stable
Case D	$-280 \mu\text{C}$	58 m	-3.86	Class 3	$4, 5, 2 \times 10^{-3} \text{Kg/m}^2$	Unstable

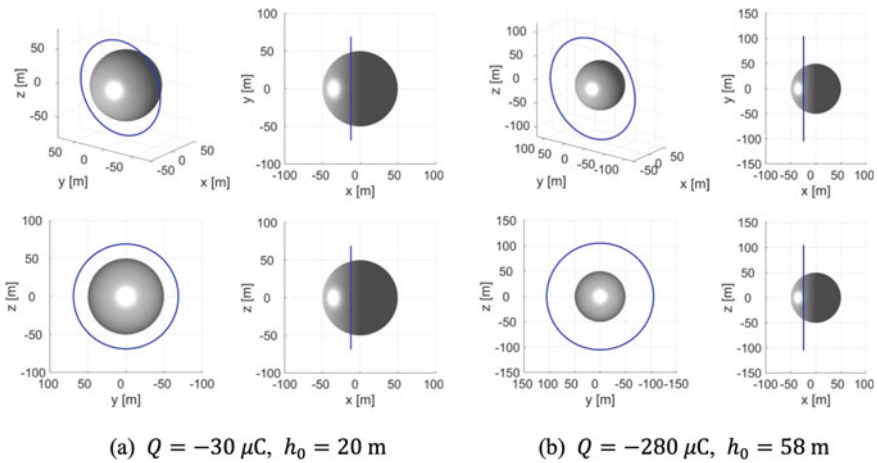


Fig. 2.35 Reference electrostatic periodic orbits for simulations of the attitude motion

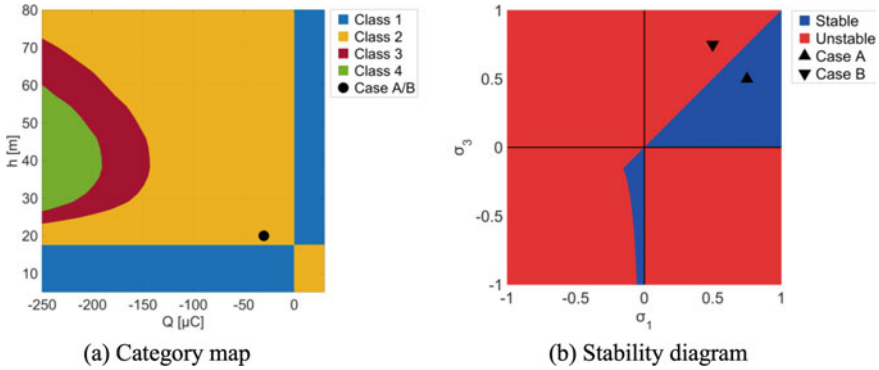


Fig. 2.36 The stability of the attitude motion for cases A and B

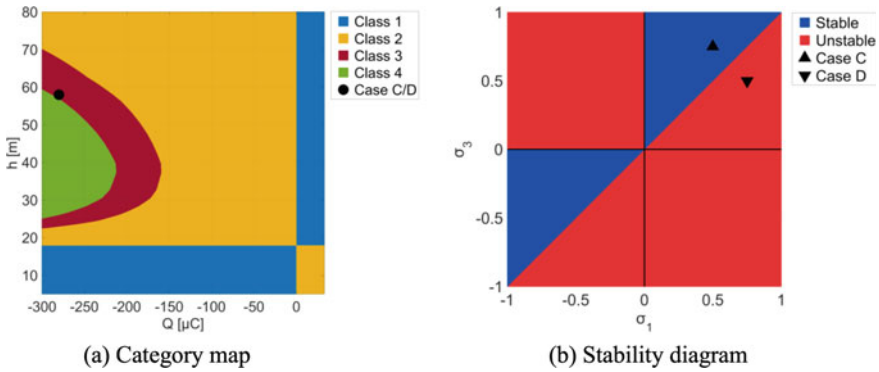


Fig. 2.37 The stability of the attitude motion for cases C and D

equilibrium roll angle that is calculated numerically based on the nonlinear equations of motion. This result demonstrates that an E-Glider can orbit around a small body with a stable attitude motion even under the influence of electrostatic torque. In contrast, Fig. 2.38b illustrates that an unstable set of moment of inertia parameters leads to a large deviation in the Euler angles. Therefore, the shape or orientation of an E-Glider must be designed properly to achieve a stable attitude motion.

Another intriguing result is presented in Fig. 2.38c and d. It can be observed from Fig. 2.38c that although the spacecraft configuration is unstable for the orbit given in Fig. 2.35a, it can achieve a stable attitude motion for the orbit presented in Fig. 2.35b. This result demonstrates that even when the attitude motion of a spacecraft is unstable under gravity gradient torque, it can be stabilized by inducing electrostatic torque. Figure 2.38d shows that a spacecraft that is stable for gravity gradient torque can exhibit an unstable attitude behavior under the electrostatic environment. The simulation results have revealed that the attitude motion of an E-Glider exhibits unique characteristics that cannot be observed in classical attitude dynamics subject

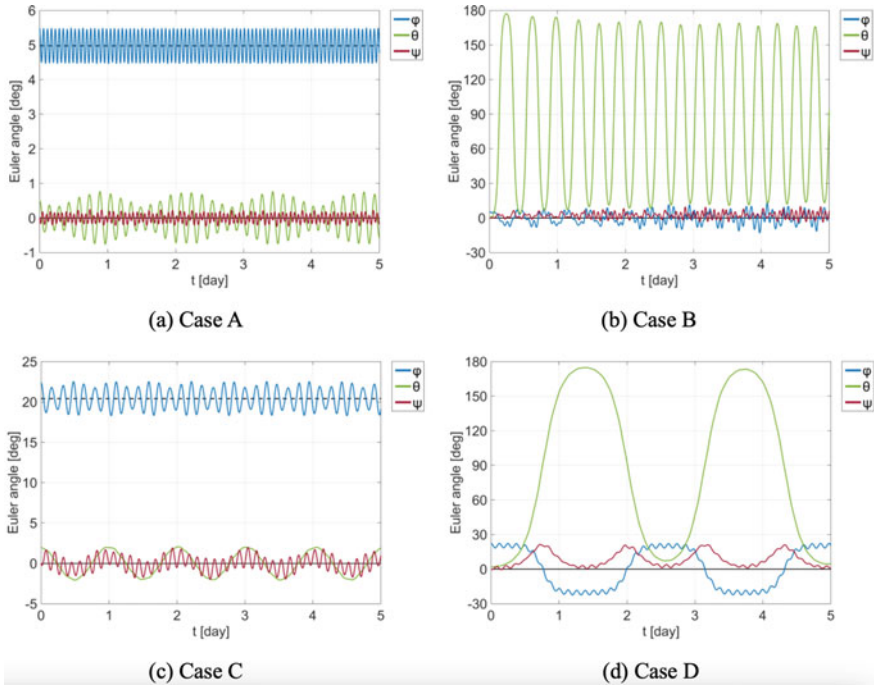


Fig. 2.38 Numerical simulation results of stable and unstable attitude motions

Table 2.4 Simulation conditions for libration and tumbling in the pitch motion

	$\dot{\theta}_0$	\hat{C}_θ	Pitch motion
Figure 2.39a	2×10^{-4} rad/s	0.19	Libration
Figure 2.39b	6×10^{-4} rad/s	1.73	Tumbling

only to gravity gradient torque. Moreover, the simulation results confirm that the numerical simulations agree with the analytical theories discussed in the previous section.

2.12.3 Libration and Tumbling in the Pitch Motion

As described in Sect. 2.11.3, when the roll and yaw motions are in equilibrium states, the pitch motion can be solved independently based on an analytical theory. This section demonstrates the validity of the analytical analysis by performing numerical simulations.

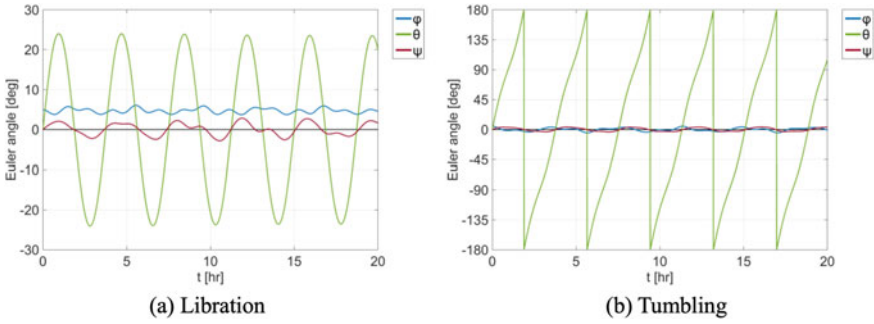


Fig. 2.39 Numerical simulation results of libration and tumbling in the pitch motion

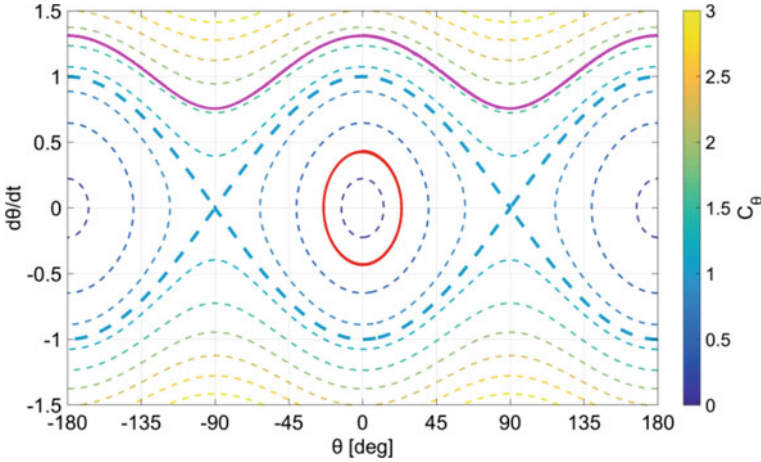


Fig. 2.40 Simulated trajectories in the phase plane for the pitch motion

The simulation is performed for the case A condition presented in Table 2.3, which provides a stable attitude motion. Equation (2.95) indicates that the \hat{C}_θ value is conserved for the pitch motion, and $\hat{C}_\theta < 1$ corresponds to libration, while $\hat{C}_\theta > 1$ corresponds to tumbling motion. The attitude motion of an E-Glider is numerically simulated for two cases with different initial pitch rates $\dot{\theta}_0$ that involve different pitch motion modes. Table 2.4 provides the simulation conditions, and Fig. 2.39 depicts the simulation results. It can be observed from Fig. 2.39a that when the initial pitch rate is relatively small, the pitch motion exhibits stable libration. On the contrary, Fig. 2.39b shows that a larger initial pitch rate results in tumbling about the pitch axis.

Figure 2.40 illustrates the histories of the pitch motion that are displayed in the phase plane. The trajectories represented as the solid red and magenta lines correspond to Fig. 2.39a and b, respectively. The broken lines show a contour map of the analytically calculated \hat{C}_θ value, which is also illustrated in Fig. 2.34. Figure

2.40 demonstrates that the numerically calculated trajectories approximately follow the contour lines, and reasonable agreement is found between the analytical theory and the numerical simulations. It is to be noted that both $d\theta/d\hat{t}$ and \hat{C}_θ described in the figure are nondimensional values that are scaled by the time unit τ given in Eq. (2.94). Because τ is a function of the ξ parameter, the effect of the electrostatic torque implicitly appears in Fig. 2.40. Therefore, these simulation results confirm that the attitude motion of an E-Glider subject to electrostatic torque can be analyzed based on classical approaches by introducing the ξ parameter.

2.13 Hovering and Its Stability

Hovering is a type of active control in which a continuous control thrust cancels out the nominal accelerations acting on the spacecraft (Broschart and Scheeres 2005; Kominato et al. 2006). For the E-Glider, the net charge Q of the spacecraft itself generates thrust or better, a continuous force. The solar radiation pressure has a strong influence on the shape of the orbits near a small body like an asteroid (leading also to the instability in some cases Scheeres and Marzari 2002), due to the weak gravitational attraction. Hovering can be a solution to avoid these problems by eliminating through an active control the accelerations of the spacecraft, thus creating an artificial equilibrium point at a desired location. Until now, fuel restrictions have limited hovering applications, but for the E-Glider, these limitations do not apply, since only the power constraints of the spacecraft will set a limit and not the fuel level. The case of fixed hovering with respect to the Sun has been analyzed in Bechini et al. (2021), Bechini (2020) and discussed here, while the case of hovering over a specific location of an asteroid is not considered (this scenario will be by far more complex due to the alternance of sunlit/shadow phases). Before approaching the hovering problem, we analyzed the zero-velocity curves to better understand the potential field near the asteroid. The equations of motion contain the electrostatic potential term to make this analysis relevant to the E-Glider case. The analysis points out that the Nitter model, used up to now, is not adequate to describe the electrostatic field in close proximity of an asteroid since by using the more refined model provided by the PIC analysis, the identification of new and more equilibrium points with respect to the ones predicted by the Nitter model is possible. This led also to the possibility of identifying stable equilibrium points in subsolar hovering on the sunlit side which is at an altitude of about 10–100 m from the surface of the asteroid. An E-Glider can hover on these points by using a charge level lower than the one predicted by using the Nitter model. The sensitivity analysis shows that the charge over mass ratio needed to hover grows quadratically with the radius of the spacecraft.

Next, we list the assumptions used to define the case study and under which the obtained results are valid. The bodies in the system under investigation are the Sun, a reference asteroid (considered as the main body) in a heliocentric orbit, and a spacecraft orbiting in close proximity to the asteroid. The main body has a circular

Table 2.5 Reference spacecraft characteristics

Parameter	Value
Shape	Sphere
Radius	0.065 m
Mass	1.33 Kg
C_{pa}	0.14
C_{ps}	0.43
C_{pd}	0.43

heliocentric orbit by assumption; thus, the equations of motion can be written by using the Clohessy–Wiltshire approximation (Curtis 2010).

We assumed the spacecraft as a solid sphere. The mass of the spacecraft is constant and equal to 1.33 Kg (as for a 1U CubeSat CubeSat 2019). The radius of the reference spacecraft is equal to 0.065 m (it is computed from the volume of a sphere equivalent to a 1U CubeSat). The computation of the SRP properties of the spacecraft can be performed by assuming a layer of Mylar with a coefficient of absorptance (C_{pa}) equal to 0.14 (Finckenor 1999) that covers the surface. The coefficients of specular (C_{ps}) and diffuse (C_{pd}) reflection have the same value, equal to 0.43, by assumption. The SRP force is modeled by using the backward ray-casting model. The charge Q is considered as fixed and modeled as a point charge concentrated in the center of mass of the spacecraft; thus, the first moment of charge is zero. Table 2.5 summarizes the spacecraft characteristics.

The main body of these simulations is the reference asteroid used also for the PIC (particle-in-cell) analysis (Yu et al. 2016). The reference main body is a spherical asteroid with a radius of 14 m. The gravity model used is the point mass model with $\mu = 0.0017 \text{ m}^3/\text{s}^2$. The circular heliocentric orbit of the asteroid has a radius of 1 AU and a period of 365.25 days at epoch 2451545 JD. The rotational parameters are assumed to be measured at the same epoch of the orbital parameter. The right ascension of the rotational axis is set to 0° , and the declination is set to 90° with a rotational period of $1^\circ/\text{day}$. Table 2.6 summarizes the orbital and the rotational parameters of the asteroid.

2.13.1 Zero Velocity Curves

The Clohessy–Wiltshire equations (see Eqs. (2.11) and (2.12)) can be rewritten in the RIC frame as (Curtis 2010):

Table 2.6 Reference asteroid characteristics

Orbital parameters		Rotational parameters	
Epoch	2451545JD	Epoch	2451545JD
Epoch offset	0	Epoch offset	0
Mean anomaly	0	Rotational axis RA	0
Orbital period	365.25 days	Rotational axis DE	90 deg
Semimajor axis	1 AU	Prime meridian position	0
Eccentricity	0	Rotational period	1 deg/day
Inclination	0	Geometry	
RA of ascending node	0	Shape	Sphere
Argument of periapsis	0	Radius	14 m
Gravity		Plasma	
Gravity model	Point mass	Plasma field	True
Gravitational parameter	0.0017 m ³ /s ²		

$$\begin{aligned}
 \ddot{x} - 2N\dot{y} &= -\frac{\partial U}{\partial x} \\
 \ddot{y} + 2N\dot{x} &= -\frac{\partial U}{\partial y} \\
 \ddot{z} &= -\frac{\partial U}{\partial z}
 \end{aligned} \tag{2.105}$$

In Eq. (2.105), U is the potential as a function of the position of the spacecraft \mathbf{r} . For the case under analysis, the contributions to the total potential are the gravitational effect, the centrifugal effect, and the equivalent potential given by the solar radiation pressure and by the electrostatic potential. Thus,

$$\begin{aligned}
 -\frac{\partial U}{\partial x} &= 3N^2x + a_{g,x} + a_{p,x} + \frac{Q}{M}E_x \\
 -\frac{\partial U}{\partial y} &= a_{g,y} + a_{p,y} + \frac{Q}{M}E_y \\
 -\frac{\partial U}{\partial z} &= -N^2z + a_{g,z} + a_{p,z} + \frac{Q}{M}E_z
 \end{aligned} \tag{2.106}$$

By integrating and by assuming a simplified case in which the potential of the SRP force is $U_p = \mathbf{a}_p \cdot \mathbf{r}$ as in Scheeres (1999), the potential can be written as

$$U(\mathbf{r}) = -U_g(\mathbf{r}) - \frac{N^2}{2}(3x^2 - z^2) + \frac{Q}{M}\phi_e(\mathbf{r}) - \mathbf{a}_p \cdot \mathbf{r} \tag{2.107}$$

where $U_g(\mathbf{r})$ is the gravitational potential (resulting from $\mathbf{a}_g = \nabla U_g$), and $\phi_e(\mathbf{r})$ is the electrostatic potential in \mathbf{r} resulting from $\mathbf{E} = -\nabla\phi_e$. Equation (2.105) can be

written in a more compact form; thus,

$$\frac{\partial^2 \mathbf{r}}{\partial t^2} = -\nabla U \quad (2.108)$$

But

$$\frac{\partial \mathbf{r}}{\partial t} \cdot \frac{\partial^2 \mathbf{r}}{\partial t^2} = -\frac{\partial \mathbf{r}}{\partial t} \cdot \nabla U \quad (2.109)$$

where

$$\frac{\partial \mathbf{r}}{\partial t} \cdot \frac{\partial^2 \mathbf{r}}{\partial t^2} = \mathbf{v} \cdot \dot{\mathbf{v}} = \frac{1}{2} \frac{d^2}{dt} v^2 \quad (2.110)$$

where \mathbf{v} and $\dot{\mathbf{v}}$ are computed in the co-moving frame relative to the co-moving frame. Moreover, we can notice that

$$\frac{\partial \mathbf{r}}{\partial t} \cdot \nabla U = \frac{\partial \mathbf{r}}{\partial t} \nabla U^T = \frac{\partial U}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial U}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial U}{\partial z} \frac{\partial z}{\partial t} = \frac{dU}{dt} \quad (2.111)$$

Thus,

$$\frac{1}{2} \frac{d}{dt} v^2 = -\frac{dU}{dt} \Rightarrow \frac{1}{2} \frac{d}{dt} v^2 + \frac{dU}{dt} = 0 \Rightarrow \frac{d}{dt} (v^2 + 2U) = 0 \quad (2.112)$$

This leads to $v^2 + 2U \doteq C_j$ which is constant and corresponds to an integral of motion (or the Jacobi integral). The total energy can be written as $\frac{1}{2}v^2 + U = E_{TOT}$. Since $v^2 \geq 0$ is always true, $U(\mathbf{r}) \leq C_j$ must hold. This last equation defines a constraint for the allowable regions of the spacecraft. The boundaries of these regions are the zero-velocity curves. The potential $U(\mathbf{r})$ is affected by the spacecraft charge Q (see Eq. (2.107)); thus, an analysis for different levels of Q in proximity to the asteroid allows better understanding of the presence of equilibrium points and the U transitions. The equilibrium points can be obtained by imposing $\frac{d}{dt} = 0$ in Eq. (2.105), resulting in the equilibrium conditions $\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = \frac{\partial U}{\partial z} = 0$. The analysis here presented is reduced to the xy plane in which the negative x is the sunlit side of the asteroid, with the sunlight coming from the $-x$ -direction.

First, we analyzed the case for $Q = 0 \mu\text{C}$. The results obtained and shown in Fig. 2.41 are consistent with the ones already presented in Kikuchi (2017). Only one equilibrium point exists in proximity to the asteroid on the dark side. In the equilibrium point, the effect of the SRP plus the centrifugal force (both acting in the $+x$ -direction) counter the effect of the gravitational acceleration of the asteroid. The potential is negative, with negative peaks in proximity to the planet.

Next, we analyzed the case when the spacecraft has a positive charge. In this case, the influence of the charge itself on the potential field is not strong. On average, if the positive charge is increased, the potential field is “pushed” through more negative values. The main behavior in close proximity to the asteroid is not strongly affected by the charge as the potential keeps decreasing through highly negative values as in

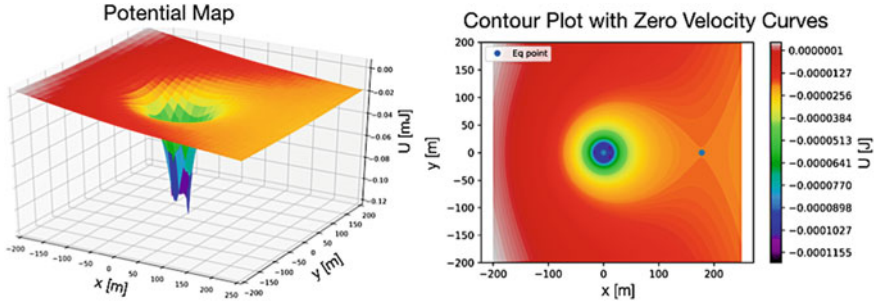


Fig. 2.41 Potential map and zero-velocity curves for $Q = 0 \mu\text{C}$

the case for $Q = 0$. Figure 2.42 shows the zero-velocity curves plots for the cases of $Q = 15 \mu\text{C}$ (top), $Q = 25 \mu\text{C}$ (mid), and $Q = 75 \mu\text{C}$ (bottom).

The blue dots are the equilibrium positions. The plots on the right column of Fig. 2.42 are the details of the plots on the left computed only for the sunlit region ($-55 \leq x \leq 0$ and $0 \leq y \leq 40$) with a more refined mesh. By looking first at the plots on the left column, we can notice that different levels of positive charge do not have a strong impact on the shape of the potential field. Moreover, the effects of the plasma wake are evident (obviously, they are not present if $Q = 0$, see Fig. 2.41). The modifications in the potential field given by the influence of the plasma are different from the results previously obtained by using the Nitter model (Kikuchi 2017), especially in the terminator line (transition between sunlit side and dark side) and in close proximity to the asteroid surface. By neglecting the infeasible equilibrium point located at $x = y = z = 0$, there is at least one collinear ($y = 0$) equilibrium position on the dark side, far from the surface at about 105–110 m. These equilibrium points disappear (as shown in Fig. 2.42 for $Q = 75 \mu\text{C}$) if the level of charge reaches high values, in this case, higher than $50 \mu\text{C}$. Due to the presence of a photoelectron sheath near the surface of the asteroid, the presence of equilibrium points can be hypothesized for a positively charged spacecraft in this region. To detect these points, we needed an extremely refined mesh, for this reason, the plots on the right column of Fig. 2.42 were produced. These analyses indicated the presence of more equilibrium points. These new equilibrium points are non-collinear (they have $y \neq 0$), and there are more than one non-collinear equilibrium points for the same level of charge Q . This result is the consequence of the complex shape of the electric field near the surface of the asteroid. These points were not detectable with the Nitter model, since it uses an oversimplified model for the electric field computation with respect to the PIC results. The mesh used does not allow us to find collinear equilibrium on the sunlit side, but their presence can be predicted by considering again the photoelectron sheath. The mesh must be extremely fine to detect these points, leading to a strong increment in the computational time. The analyses restricted to the case $y = z = 0$ and $x \neq 0$ reported in the following sections confirm the presence of these equilibrium points at an altitude below ≈ 2 m (altitude at which the electrostatic potential has a minimum, thus an inversion of the sign in E_x), but with a strong

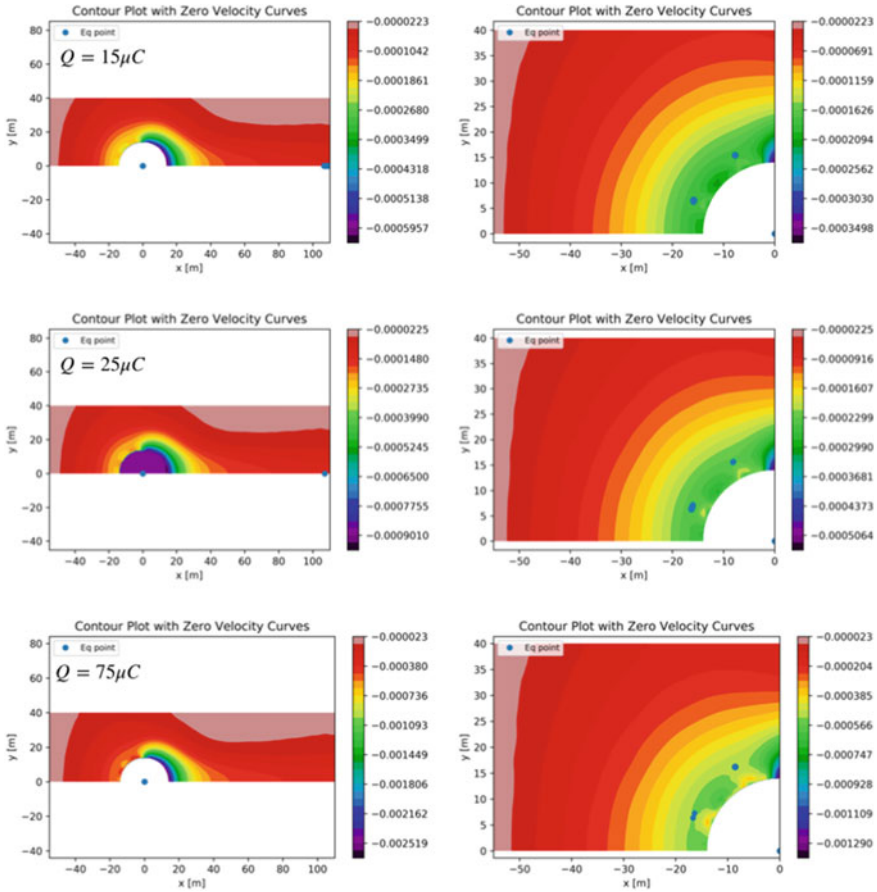


Fig. 2.42 Zero-velocity curves for $Q > 0 \mu C$

gradient of variation of Q for small variation in the position x . The non-collinear points obtained are not very sensitive to variations in the charge levels; furthermore, they are still present even for levels of charge in which the dark side equilibrium point has already disappeared. The coexistence (at least for a low level of positive charge) of equilibrium points both on the sunlit side and on the dark side is ensured, with both collinear and non-collinear points on the sunlit side.

Second, we analyzed the case in which the spacecraft has a negative charge. The charge of the spacecraft affects the potential field in a way such that the potential on the sunlit side falls to extremely negative values, while on the dark side, the potential becomes strongly positive in close proximity to the surface. Figure 2.43 shows the zero-velocity curves for the cases of $Q = -5 \mu C$ (left) and $Q = -10 \mu C$ (right).

The equilibrium conditions are present both on the sunlit side and on the dark side even for a low level of charge; moreover, there is more than one equilibrium

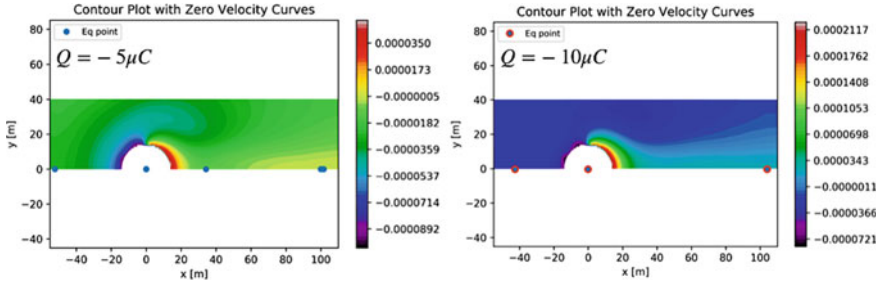
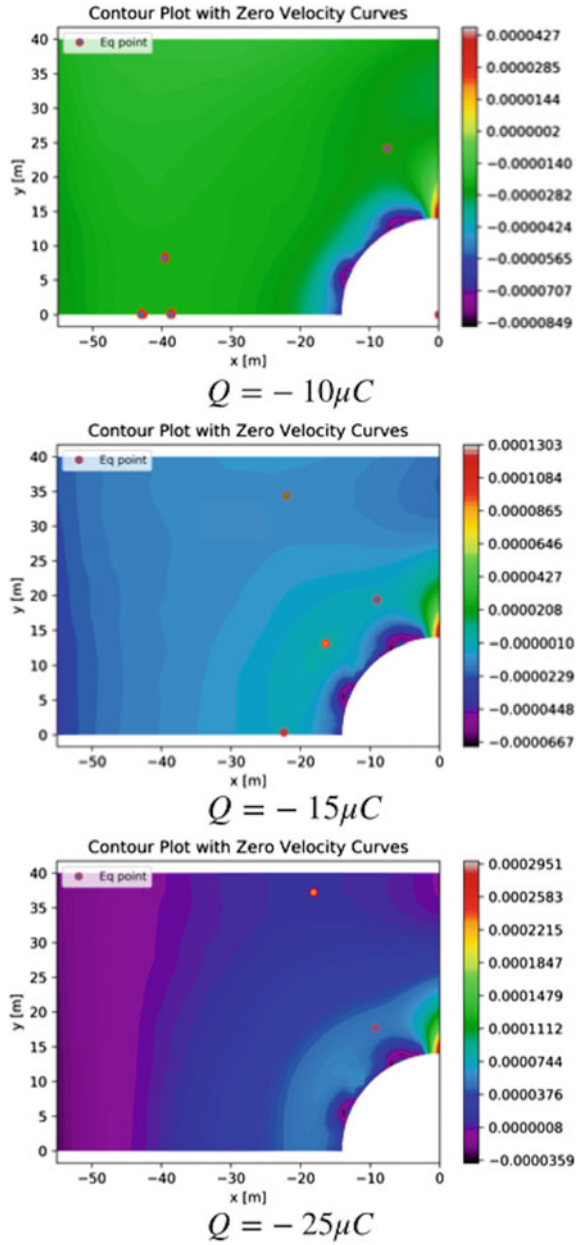


Fig. 2.43 Zero-velocity curves for $Q < 0 \mu C$

condition on the same side for the same level of charge. This result is in contrast with the previous study (Kikuchi 2017) based on the Nitter model. By using the Nitter model, the equilibrium conditions were obtained only on one side per time with a negatively charged spacecraft; moreover, the equilibrium conditions on the sunlit side were obtained only for extremely high levels of charge (Kikuchi 2017). A comparison between the equilibrium points on the sunlit side for the two cases reported indicated that the equilibrium points are closer to the asteroid surface for a charge higher in modulus. Also, this result is in contrast with the Nitter theory, where an augmenting negative charge moves the spacecraft far away from the asteroid towards the direction of the Sun (Kikuchi 2017). The detection of equilibrium positions on the sunlit side is affected by the dimensions of the mesh, thus, in this case, to also augment the number of points in the grid without strongly affecting the computational time, some analyses have been performed for the case in which $-55 \leq x \leq 0$ and $0 \leq y \leq 40$ with a more refined mesh. The results are in Fig. 2.44.

There are more equilibrium positions, for the same level of charge, on the subsolar axis, as deduced before. Moreover, several noncollinear equilibrium points (equilibrium points with both x and y different from zero) are present also for a low level of charge. A comparison between the three cases reported in Fig. 2.44 indicated a drift of the equilibrium points for different levels of charge. By making the charge more negative, the collinear equilibrium points move towards the surface of the asteroid until they disappear (see the case of $Q = -25 \mu C$), while the non-collinear points seem to drift away, and some of them seem to appear and then disappear (a better understanding of this phenomenon can be obtained by using an extremely refined mesh). These results cannot be obtained by using the Nitter model; in particular, the presence of more than one non-collinear equilibrium position has not been assessed before.

Fig. 2.44 Zero-velocity curves for $Q < 0 \mu\text{C}$ on sunlit side



2.13.2 Subsolar Hovering

The analysis of the zero-velocity curves points out that there are several points of equilibrium dependent from the charge. The analysis was restricted to only the subsolar axis (thus only along the x-direction, with $y = 0$ and $z = 0$) in order to better characterize the equilibrium conditions in this region by achieving a higher accuracy without strongly refining the mesh. The equation for the hovering along the x-axis (Eq. (2.113)) can be derived from the general equation of motion written under the Clohessy–Wiltshire assumptions; thus,

$$3N^2x + \frac{f_{g,x}}{M} + \frac{f_{p,x}}{M} + E_x \frac{Q}{M} = 0 \quad (2.113)$$

The solution of Eq. (2.113) for Q/M gives the charge over mass ratio needed to hover at each position along the x-axis. The stability has been evaluated for each equilibrium condition. The stability conditions for the case under exam are

$$\frac{\partial U}{\partial x} = 0 \quad (2.114)$$

$$\left. \frac{\partial^2 U}{\partial x^2} \right|_{eq} > 0 \quad (2.115)$$

where the selected Q/M always verifies Eq. (2.114). The x derivative of Eq. (2.113) leads to

$$\frac{\partial^2 U}{\partial x^2} = -3N^2 - \frac{\partial^2 U_g}{\partial x^2} - \frac{\partial a_p}{\partial x} - \frac{Q}{M} \frac{\partial E_x}{\partial x} \quad (2.116)$$

The second derivative of the gravitational potential is computed by using Pines' algorithm with a modified recursion formula (Gottlieb 1993; Pines 1973; Lundberg 1988; Fantino and Casotto 2009). The derivative of the electric field can be numerically computed as the x-component of the gradient of the electrostatic field. The term $\frac{\partial a_p}{\partial x}$ can be neglected since the E-Glider is supposed to fly in close proximity with respect to the asteroid; thus, $\frac{\partial a_p}{\partial x} \approx 0$ can be assumed. The results of this first analysis are reported in Fig. 2.45.

The central gray band in Fig. 2.45 represents the asteroid, and the blue dotted line represents the limit of the nominal photoelectron sheath (which corresponds to the nominal photoelectron Debye length, equal to 1.38 m). Only a few points of stable hovering over the sunlit face exist. These points are at about 42 m from the center of the asteroid and can be interesting from a "real mission" point of view. A negatively charged spacecraft could achieve almost all the hovering conditions, except some positions on the dark side (at more than 100 m of distance from the center of the asteroid) and for altitude below the photoelectron sheath (which are not reported in the figure for scale issues, since these points are located at values even equal to $Q/M = 0.0024$ C/Kg). This analysis, in agreement with the results obtained in

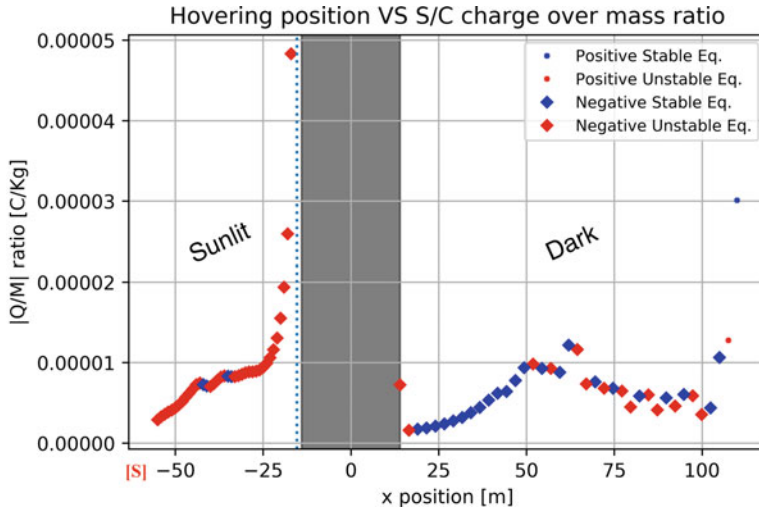


Fig. 2.45 Subsolar hovering conditions

Table 2.7 Radius values for simulations

Reference [m]	Values [m]
0.065	0.1 0.25 0.5 0.75 1.0 2.0 5.0 7.5 10

Corradino (2018), confirmed that in the nominal case, for the assumed spacecraft and main body parameters, the Q/M ratio required for the hovering over the sunlit face is in the order of 10^{-5} C/Kg. A stable hovering at about 10 – 100 m of altitude is possible by assuming that the levels of power and voltage required are achievable. Moreover, the level of charge needed to achieve the hovering condition in subsolar positions is lower than the one predicted by using the Nitter model (Kikuchi 2017).

A sensitivity analysis of the equilibrium conditions has been performed by changing the value of the equivalent radius of the sphere which represents the spacecraft to improve the characterization of the hovering conditions. The values used for the simulations are reported in Table 2.7.

The reference value is equal to the radius used for the previous analysis. The other parameters are kept fixed, such that only the radius changes. An analytical study has been performed before the numerical analysis. The Q/M equation can be written as

$$\frac{Q}{M} = \left[-3N^2x - \frac{f_{g,x}}{M} - \frac{f_{p,x}}{M} \right] \frac{1}{E_x} \tag{2.117}$$

By considering the spacecraft as a “point” concentrated in the center of mass of the equivalent sphere, the only term dependent on the radius of the sphere R_{SC} is the SRP force. By using the simple cannonball model (only for the analytical formulation), the derivative of Eq. (2.117) with respect to R_{SC} is

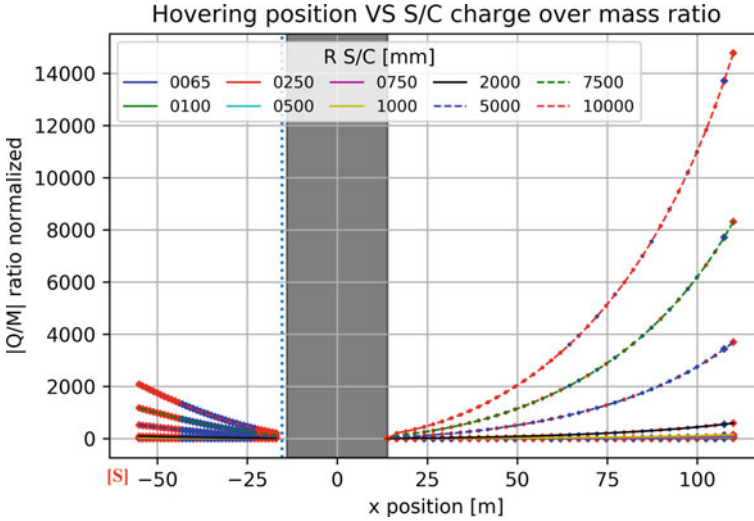


Fig. 2.46 Variation of normalized $|Q/M|$ as a function of R_{SC}

$$\begin{aligned} \frac{\partial Q/M}{\partial R_{SC}} &= \frac{\partial}{\partial R_{SC}} \left(\frac{\Phi}{c} \pi R_{SC}^2 \left(C_{ps} + \frac{13}{9} C_{pd} + C_{pa} \right) \hat{s} \right) \\ &= \frac{\Phi}{c} \pi 2 R_{SC} \left(C_{ps} + \frac{13}{9} C_{pd} + C_{pa} \right) \hat{s} \end{aligned} \quad (2.118)$$

$\frac{\partial Q/M}{\partial R_{SC}}$ is linearly increasing for positive values of R_{SC} ; thus, the Q/M needed for the hovering in proximity to an asteroid increases quadratically with the increment of the equivalent radius of the spacecraft (the minimum is achieved for $R_{SC} = 0$ which is a infeasible solution). The numerical results are reported in Fig. 2.46.

The big red marker stands for unstable equilibrium positions achievable with a negative spacecraft charge, while the big blue marker stands for the stable equilibrium positions obtained with a negative charge. The small markers stand for the equilibrium conditions achieved with positive charges (the color code for stable and unstable equilibria is the same). The simplified analytical approach is compliant with the numerical results. The Q/M needed to achieve the hovering condition increases with a quadratic law with the radius of the spacecraft. The Q/M ratios reported in Fig. 2.46 are the absolute value of the ones computed with a prefixed radius in a defined position normalized by the Q/M ratio obtained at the very same position for the reference radius. The normalized $|Q/M|$ increases both on the sunlit and on the dark side by moving away from the surface. The variation of the spacecraft radius does not affect the stable equilibrium region. Figure 2.47 shows that the behavior on the dark side seems to be quite different from the predicted one.

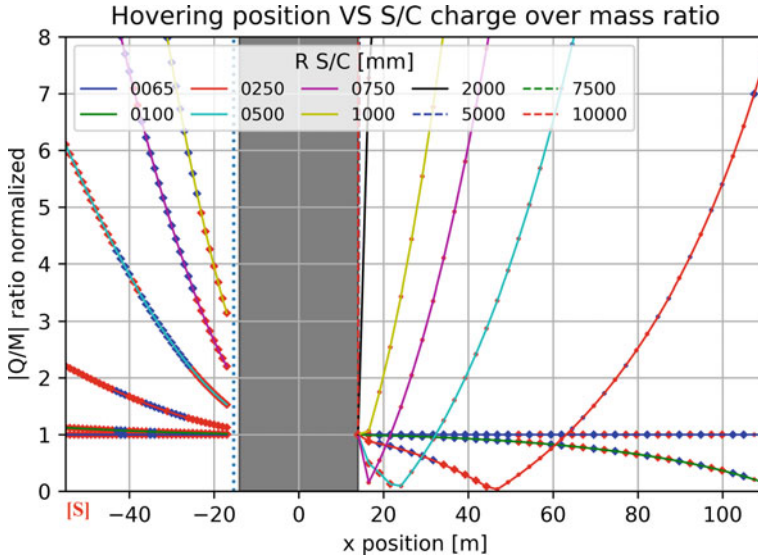


Fig. 2.47 Detail of Fig. 2.46

Table 2.8 Main body positions

Position [AU]	Reason
0.4	Mercury orbit
1.0	Earth orbit
1.5	Mars orbit
2.2	Inner asteroid belt radius
2.75	Mid asteroid belt radius
3.3	Outer asteroid belt radius

The behavior of these curves is due to the fact that the normalized $|Q/M|$ is computed as the ratio between two absolute values. If the radius of the spacecraft increases, the transition from a negative to a positive Q/M ratio (obviously the sign is dictated by the charge Q) is anticipated, as can be seen in Fig. 2.48; For $R_{SC} = 0.065$, the transition happens at ≈ 105 m, while for $R_{SC} = 0.5$ m, the transitions happens at about 25 m from the center of the asteroid. For a radius higher than 1 m, the Q/M needed for the hovering on the dark side is positive for regions close to the asteroid. A second transition region located at about 105 m can be identified in Fig. 2.48. This is due to the electric field x-component that becomes positive.

A sensitivity analysis to evaluate the effects of the position of the main body with respect to the Sun on the $|Q/M|$ ratio has been performed. The five different positions reported in Table 2.8 are considered in this analysis.

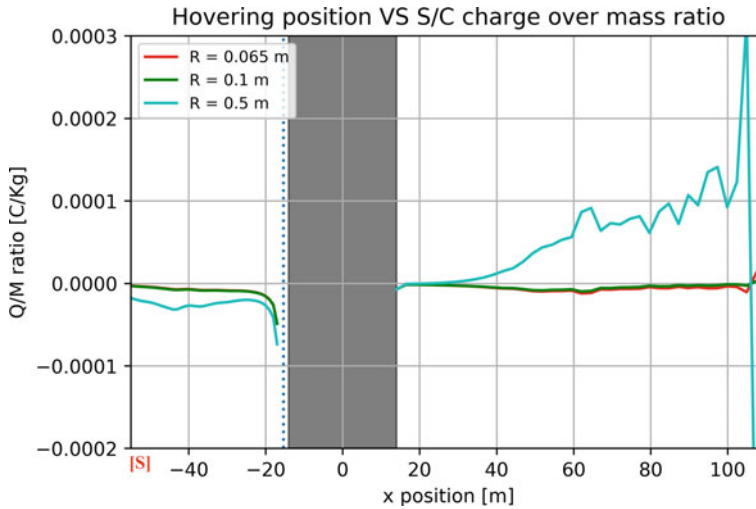


Fig. 2.48 Q/M for the first three cases examined

The positions considered are selected by considering the regions interesting for possible applications of an E-Glider. The case of an asteroid at 1 AU is the reference for the analysis. By changing the distance with respect to the planet, the orbital period (computed in agreement with the selected semimajor axis) and the plasma parameters (the solar wind density and the solar wind ions and electrons temperature) change. The variation of the plasma parameters strongly affects the current collected by the electrodes and the power needed to maintain the charge, but these effects are not considered in this analysis. The results of the numerical analysis are available in Fig. 2.49.

The absolute values of the Q/M obtained for each planet are normalized by the absolute value of the Q/M computed for the reference case. Figure 2.49 indicates that the normalized $|Q/M|$ increases on the sunlit side in the case of an inner asteroid, while it decreases in the case of outer asteroids. Moreover, the increment is higher if compared to the decrements obtained in each case of an outer asteroid. This is due to the strongest effects of the solar wind acting on the sunlit side. By moving away from the asteroid in the direction of the Sun, the difference with the reference case increases by following an exponential law. On the dark side, the trend is the opposite. For an inner asteroid, the Q/M needed switching from negative to positive closer to the surface with respect to both the reference case and the outer asteroids; thus, there is a region (just before the transition) in which the Q/M required is effectively reduced in modulus (see Fig. 2.50). Also, in this case, there is a second transition position located at about 105 m (see Fig. 2.50) from the center of the asteroid, on the dark side region due to the inversion of the sign in E_x . Figure 2.50 confirms that hovering on the sunlit regions requires a higher level of charge for inner asteroids, while on the dark side, the charge required for these asteroids is lower. The solar wind

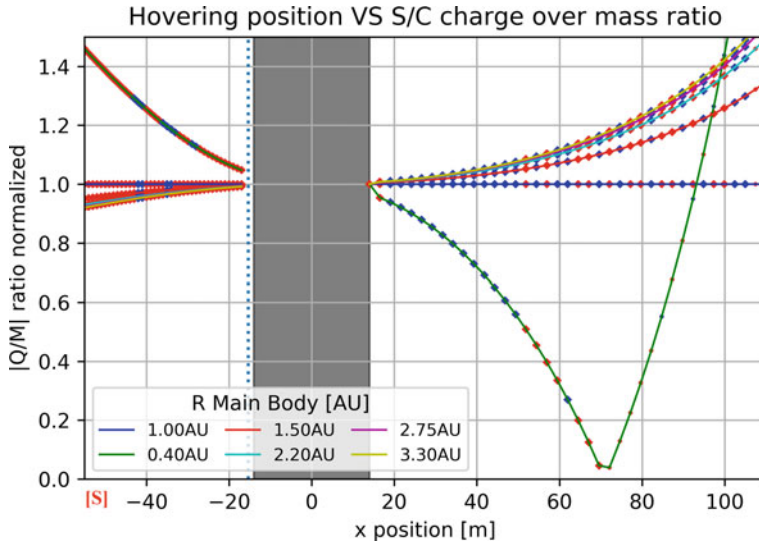


Fig. 2.49 Variation of normalized $|Q/M|$ as function of the asteroid position

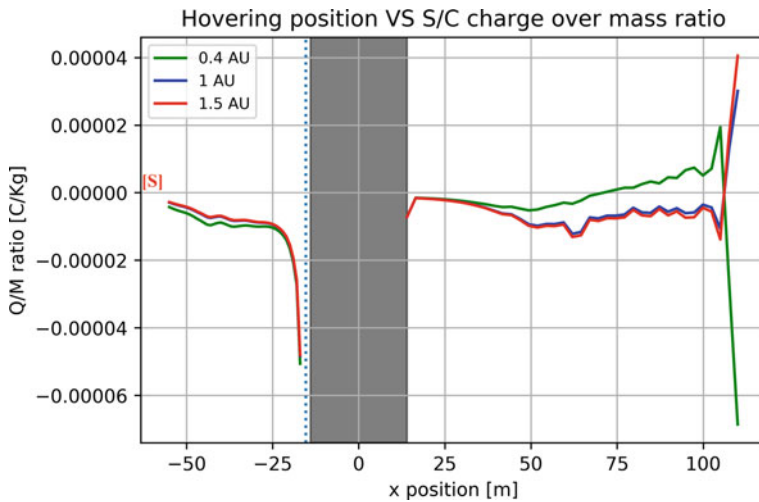


Fig. 2.50 Variation of Q/M as function of the asteroid position

effects on the dark side hovering are quite low for an outer asteroid, since the $|Q/M|$ increment switching from 1.5 AU to 2.2 AU is much stronger than the one computed from 2.2 AU to 3.3 AU (see Fig. 2.49). The positions of the stable equilibria are almost not changed with respect to the reference case (the legend for the equilibrium point is the same of the previous cases).

2.13.3 *Electrostatic Periodic Orbit*

2.13.3.1 Introduction

Previous works assessed the presence of a class of periodic orbits, called *electrostatic periodic orbits* Quadrelli et al. (2017b), Kikuchi (2017), under the assumption of an electrostatic field modeled with the Nitter model. The previous work that used the PIC results (Corradino 2018) only hypothesized the possibility of achieving the electrostatic periodic orbits without defining them. The electrostatic periodic orbits computed by using the Nitter model are found to have different shapes as a function of the charge. Moreover, assessed was the possibility to have more than one periodic orbit for the same level of charge starting from different initial positions (Kikuchi 2017). These orbits are found to be displaced through the direction of the Sun, allowing the E-Glider to orbit the central asteroid on the sunlit side, thus in a more favorable position with respect to the periodic neutral orbit, which is found to be displaced through the dark side of the asteroid (this can have drawbacks from the mission viewpoint). These orbits, as said, were computed by assuming the Nitter model for the electrostatic field close to the main body. As shown above, the electrostatic field described by the PIC results is much more complex than the one obtained by using the Nitter model, especially for a negatively charged spacecraft on the sunlit side (which is the case for which the electrostatic periodic orbits have been obtained); thus, it is necessary to evaluate if these orbits are still present even if the PIC results are used and/or if they have some modifications from the results obtained by using the Nitter model.

2.13.4 *Neutral Periodic Orbit*

The natural periodic orbits determination has been used as a benchmark case for testing the algorithm. The natural periodic orbits are periodic orbits in the RIC reference frame characterized by a neutral total charge $Q = 0 \mu\text{C}$, referred to in the literature as terminator orbits (Quadrelli et al. 2017b; Kikuchi 2017). These orbits are known to be displaced through the dark side of the asteroid. They are Sun-synchronous orbits with the characteristics of being perpendicular to the subsolar axis. Since the neutral orbits are characterized by a charge equal to zero, they are influenced only by the gravitational effects and by the Solar radiation pressure. In Fig. 2.51, shown is an example of a neutral terminator orbit found with the previously described algorithm computed in the RIC reference frame for about $(2.1, 0, 39.94)^T$ as initial position (in meters) and $(0, 0.00656, 0)^T$ as initial velocity (in m/s) propagated for a period of time equal to 10 Earth days using the full-model equation of motion described in Eq. (2.9). From the plot, it can be noticed that the drifting of these orbits is low indeed, and the displacements from the reference initial orbit (the orbit obtained for the first orbital period) are small. The period of the orbit is about 5.48 h. To test

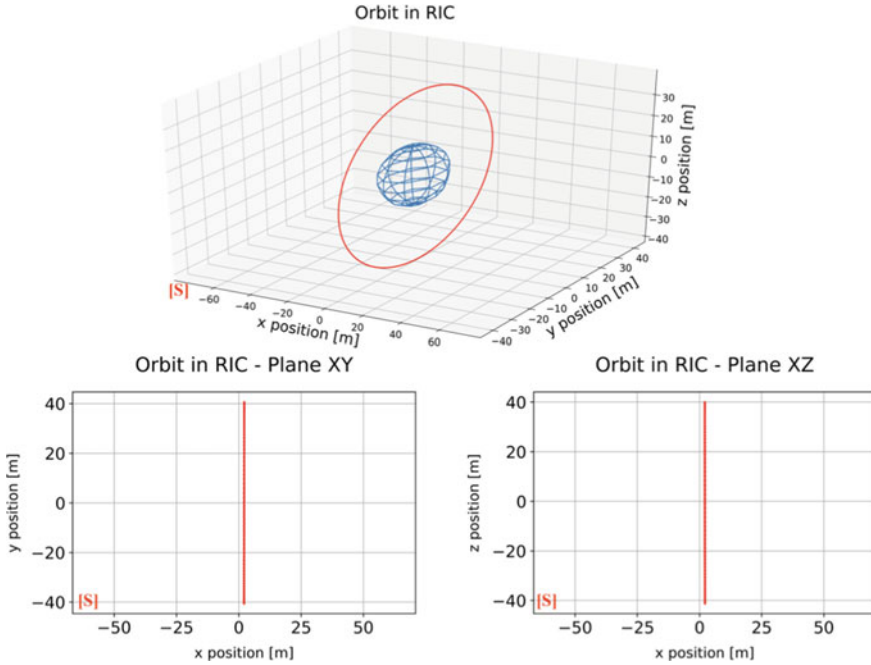


Fig. 2.51 Example of a neutral terminator orbit

the iterative process, the initial conditions that allow us to obtain a neutral periodic orbit have been evaluated. Since the shape and the position of the neutral orbit are influenced only by the gravitational effects and by the SRP, it is expected that a drift of the initial position is needed to obtain a periodic neutral orbit towards the dark side (positive x-direction in the RIC reference frame) if the distance from the center of the asteroid increases. This expectation is fully confirmed by the initial positions computed and reported in Fig. 2.52. The blue line in Fig. 2.52 represents the surface of the asteroid. The maximum displacement computed in the x-direction is about 2.2 m, obtained for $r_0 = 40$ m. These positions are computed by using $\varphi_0 < 90^\circ$ as an initial guess.

2.13.5 Connection with the Electrostatic Periodic Orbit

Once the algorithm was proven to work properly for a neutral orbit, it was used to find the initial condition that allows us to obtain an electrostatic periodic orbit, thus connecting with the previous section where the periodic orbits are described analytically. It has been noticed that the computational cost of this algorithm is extremely high; thus, we reduced the zone of interest in which to select the initial guess con-

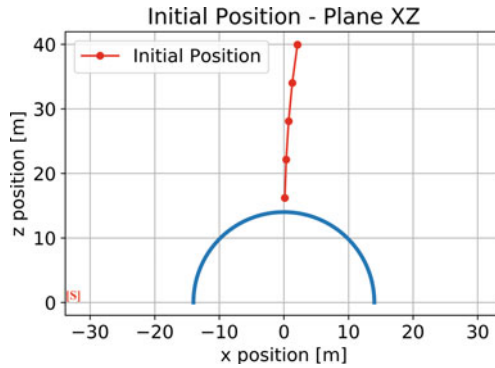


Fig. 2.52 Neutral terminator orbit’s initial positions

ditions. The periodicity of an electrostatic orbit can be achieved by exploiting the spacecraft charge level to cancel out the force component along the x-direction in the RIC reference frame. To identify the level of charge over mass ratio needed to cancel out the x-component of the force, it is possible to use Eq. (2.117), here reported for completeness.

$$\frac{Q}{M} = \left[-3N^2x - \frac{f_{gx}}{M} - \frac{f_{SRPx}}{M} \right] \frac{1}{E_x} \tag{2.119}$$

By evaluating Eq. 2.119 on a domain such that both the x- and the z-directions in the RIC reference frame are different from zero, the searched level of Q/M is obtained. The results of this analysis are reported in Fig. 2.53, while Fig. 2.54

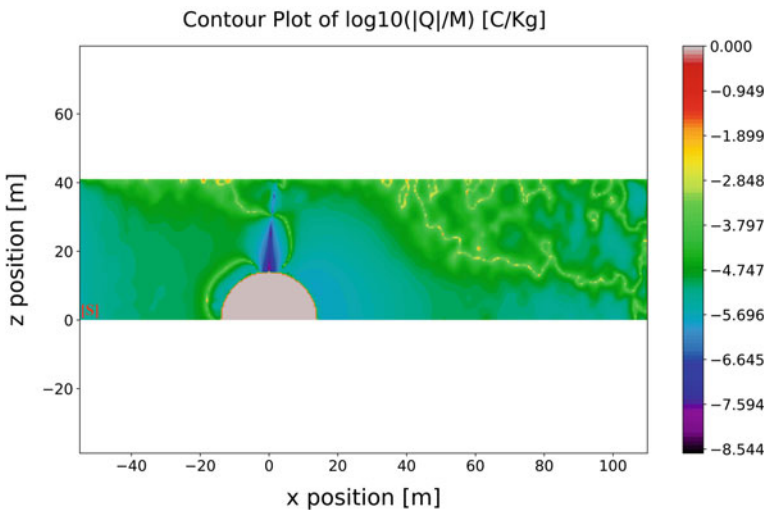


Fig. 2.53 Charge over mass ratio required for orbiting

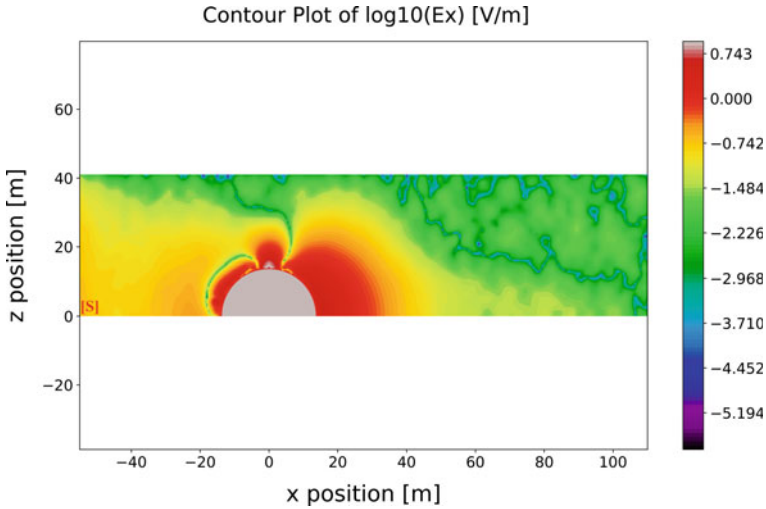


Fig. 2.54 Axial electrostatic field in the xz -plane

presents the level of the axial electric field (along the x -direction) on each point of the evaluation domain of the Q/M ratio.

Figure 2.53 allows us to identify those regions in which the charge over mass ratio required for the orbiting is less than the one required for the subsolar hovering, resulting in a more efficient strategy that can also enable the achievement of more interesting observation points. The sunlit regions that allow for a smaller charge over mass ratio with respect to the hovering case are few and located near the terminator region. This result is expected since the highest strength of the axial electrostatic field is computed above this region (as demonstrated in Fig. 2.54). In Fig. 2.53, the region in which it is possible to establish a periodic neutral orbit can be also found (colored in violet in the figure). It can be noticed that the neutral orbit region obtained in Fig. 2.53 is coincident with the one computed with the previously discussed algorithm and reported in Fig. 2.52. By exploiting these “low-charge” regions, it could be possible to “tug” the orbit in the sunlit direction. It is possible to identify also some regions of low required charge near the subsolar position, in close proximity to the equilibrium points identified when the charge is negative. Anyway, orbiting in these regions can be extremely risky since the favorable regions are surrounded by extremely adverse regions, in which the electrostatic field is weak, thus requiring a high level of charge. Moreover, these regions are close to the surface of the asteroid; thus, in a real case, the perturbations given by the irregularity of the surface itself can be dangerous for the mission. In conclusion, it can be stated that the electrostatic orbiting is advantageous (from the charge over mass ratio point of view) and mainly feasible near the terminator region. Therefore, the selection of the guess initial point can be limited to this region. Furthermore, the charge can also be limited to a level of Q/M lower than the one required for the hovering. This reduction of the zone of interest allows speeding up

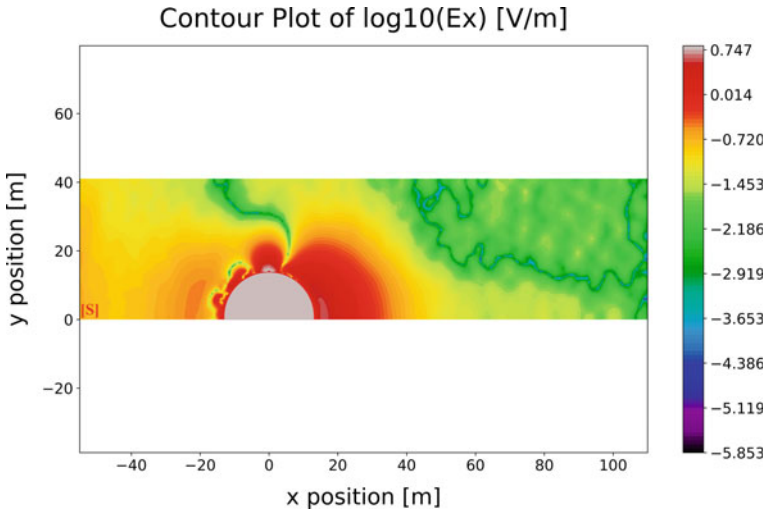


Fig. 2.55 Axial electrostatic field in the xy -plane

the entire process, but the drawback is that the possibility of achieving an electrostatic periodic orbit is not explored nor for regions further from the asteroid surface nor for a higher level of charge. This means that the solution space is not completely explored; thus, a better characterization (with a more powerful algorithm) could be needed to fully explore the solution space. It is better to stress here a main difference between the Nitter model used for past analysis and the model resulting from the PIC method used here. The electrostatic field around a spherical object in a 3D domain can be defined using the Nitter model by computing the field itself on a reference plane (the xy -plane in the RIC reference frame for example) and then performing a rotation of 360° of the obtained results around the x -axis. The electrostatic field obtained from the PIC results is computed by interpolating the data that outcome from the numerical analysis on the 3D sector identified by the x -axis and the positive semi-axis of y and z directions in the RIC reference frame. The 3D complete field can then be obtained by mirroring the 3D sector with respect to the xy -plane and then with respect to the xz -plane. By evaluating the x -component of the electrostatic field on the xy -plane (reported in Fig. 2.55) and comparing the results with the ones reported in Fig. 2.54, some variations between the plots are evident. Especially on the sunlit region, the differences between the x -component of the electrostatic field on the xy -plane and the one in the xz -plane are quite strong in magnitude and shape of the field itself. Thus, the PIC-based model has a higher degree of complexity with respect to the Nitter model since it does not show the rotational symmetry previously discussed for the Nitter model. This loss of symmetry in the numerical results of the PIC analysis can be compared with the Nitter model computed for the case of a slightly ellipsoidal main body in Kikuchi (2017); thus, in this case, an electrostatic field can be obtained which does not have a rotational symmetry about the x -axis

in RIC reference frame. Hence, this “asymmetry” can affect the periodicity of the orbits, also leading to open orbits as reported in Kikuchi (2017).

By running the simulation, it is possible to compute the initial conditions that allow us to obtain those orbits that cross the xz -plane perpendicularly. These vectors of initial conditions are stored in an external file compatible with MatLab. This file is post-processed in order to propagate the obtained initial conditions for an entire orbital period and check the distance between the starting point and the final point. This process of evaluation of the obtained orbits is a fast way to verify if they are closed, thus periodic, and if not, how big the distance between the starting and the final condition is. The results of the post-process analysis show that by using a PIC-based electrostatic field model and under the restriction previously reported, in the region of low required charge defined above, there are no periodic electrostatic orbits. This means that it is not possible to obtain a closed periodic orbit displaced through the sunlit direction by imposing a constant negative charge to the spacecraft. The analysis of the obtained distances shows that the smallest deviations (less than 1 m) are obtained for a radius of the starting orbit below 17 m and level of charge of the order of 10^{-7} C. For a higher radius of the initial position, by imposing a charge level lower than 10^{-5} C, deviations lower than 1 m after an orbital period have not been computed. The minimum distance computed is 17 cm obtained for $R_0 = 16$ m, $V_{y0} = 0.0105$ m/s and $Q_0 = -1.9 \cdot 10^{-7}$ C. In the following, the orbit obtained by imposing a charge equal to $Q = -1.6 \mu\text{C}$ and a starting distance from the center of the asteroid equal to 21 m is taken as an example. The algorithm gives as initial position the vector $\mathbf{r} = (-3.245, 0, 20.748)$ in meters, and as initial velocity the vector $\mathbf{v} = (0, 6.562, 0)$ in millimeters per seconds. Both \mathbf{r} and \mathbf{v} are expressed in the RIC reference frame. The condition of perpendicularity at the crossing of the xz -plane is satisfied since by propagating the orbit for the resulting half-period (which is about 1.94 h) it is obtained in which the x-component of the velocity vector is of the order of magnitude of 10^{-13} , while the z-component is of the order of magnitude of 10^{-14} . The y-component of the velocity at the crossing position is -1.09 cms per second. Thus, the velocity vector results to be perpendicular to the xz -plane as expected. The orbit obtained for a propagation of one orbital period is shown in Fig. 2.56.

It is evident that even if the trajectory crosses the xz -plane perpendicularly, the orbit results in being an open orbit, thus not periodic. The displacement of the final position with respect to the initial condition is both through the positive x-direction and through the positive z-direction in the RIC reference frame. This result is in agreement with that stated in Kikuchi (2017) for the case of an ellipsoidal main body; thus, the result confirms that the loss of a degree of symmetry given by the PIC results affects the possibility of defining electrostatic periodic orbits, as previously hypothesized. The main cause of the asymmetry is thus in the electrostatic field in which the spacecraft moves along its orbit. The electrostatic field components along the orbit under analysis for one orbital period are reported in Fig. 2.57.

From Fig. 2.57, it can be noticed that in correspondence with the half-period, the electrostatic field components have a strong modification (with a high gradient) with respect to the instant of time just before and just after the crossing of the xz -plane.

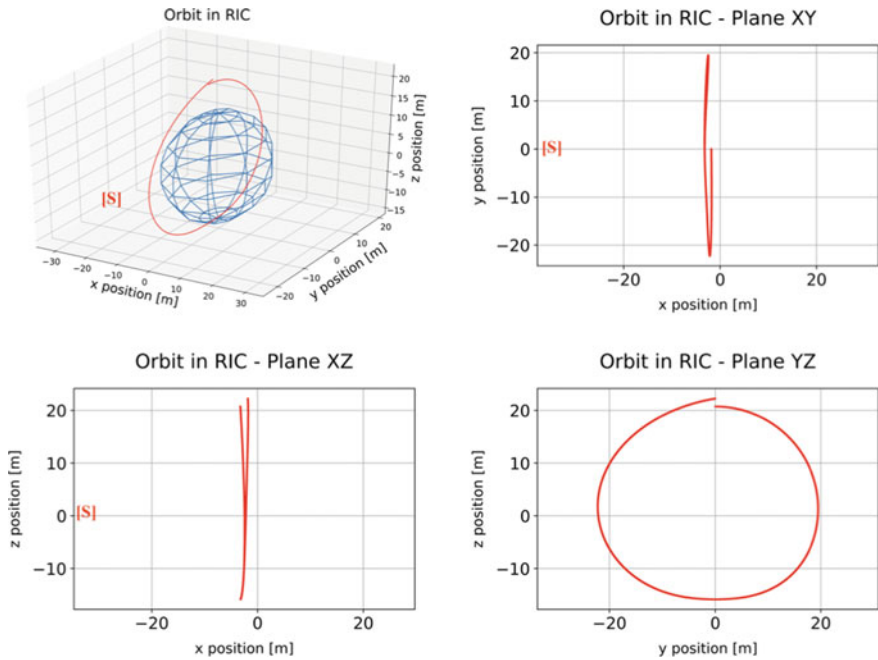
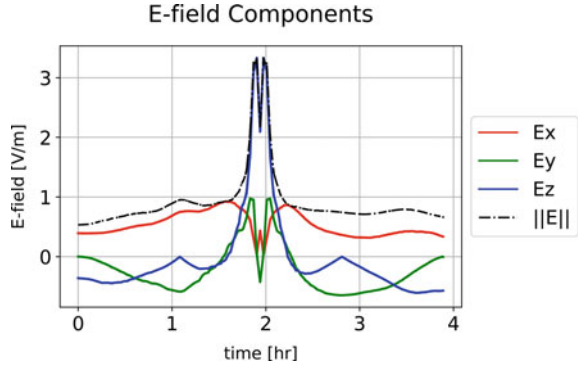


Fig. 2.56 Electrostatic orbit

Fig. 2.57 Electrostatic field components along the orbit



This modification has an effect on the electrostatic force acting on the spacecraft. Like in the motion of a pendulum, to have periodicity in the trajectory, it is required that the resultant of the forces acting in the out-of-plane direction is zero when the pendulum is in the vertical position in order to avoid perturbations in the trajectory and thus rotations of the plane of motion and loss of periodicity. The same must be verified for the orbital motion here considered, but due to the shape of the electrostatic

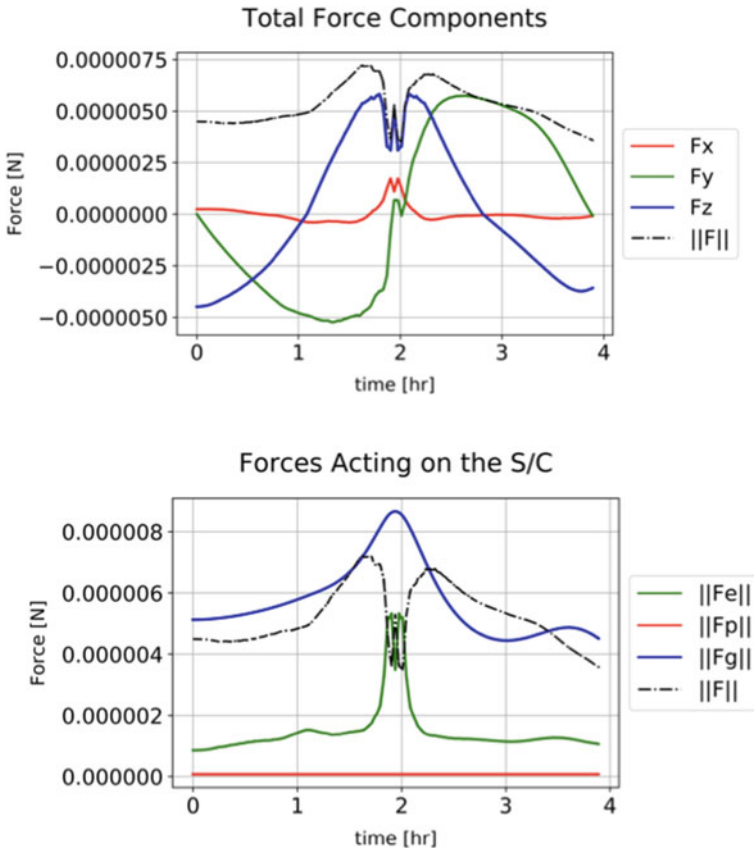


Fig. 2.58 Total force components along the orbit (top) and norm of the forces acting on the spacecraft (bottom)

field, the balance between the forces acting in the x-direction (but also along the z-direction) in the RIC reference frame is not satisfied, as can be noticed in Fig. 2.58 in the upper plot. The non-zeroing of the force component along the x-axis in the RIC reference frame and the unbalancing of the forces along the z-direction cause the displacement of the final point with respect to the initial position and the asymmetry of the orbit and thus of the resultant of the forces on the spacecraft (as can be noticed in the plot on the bottom in Fig. 2.58).

2.14 Control

2.14.1 Control Approach

Key characteristics of small-body targets are lower gravity and lack of atmosphere. The low gravity allows for (1) longer timelines for surveillance and characterization of the target site, (2) gradual descent to the target, (3) multiple landings or contacts and ascent, and (4) aborting and restarting during critical activities. Low-gravity maneuvering differs fundamentally from high gravity in the timescales, requirement for high thrust, and the need for closely monitoring the trajectory and attitude-control loops. An important characteristic of these missions is the lack of a priori information about the body. *JPL's* AutoNav (Quadrelli and Bhaskaran 2019) is ideally suited for E-Glider operations and is capable of achieving position control to within 3 m and horizontal velocity control better than 2 cm/s. Landmark-based autonomous navigation with terrain relative navigation (TRN) and hazard detection and avoidance (HDA) will be necessary for the E-Glider to reach critical landing sites of high scientific interest that are surrounded by terrain hazards. TRN is an image processing method that extracts kinematic position (and optionally attitude) information from onboard sensor data (e.g., camera images, LIDAR range image/map, etc.) for subsequent use in an estimation filter. HDA is a landing function that uses data collected onboard to identify safe landing sites in real time as the vehicle descends. The NEAR and Hayabusa asteroid landings demonstrated that such missions are feasible using ground-in-the-loop navigation at tens of meters of accuracy. Future proximity operations and landings on small bodies may need to achieve accuracies of less than 5 m. A typical timeline for the E-Glider, in the context of a small-body mission, is discussed next. Once released, the vehicle extends its wings and hovers. Through an array of Langmuir probes that measure the spatial distribution of the charges surrounding the vehicle, a map of the local electrostatic field is generated. This map is the result of the differentials between the model and the measurements which are continuously updated in flight. Once the electric potential has been mapped, the E-Glider is able to use this electrostatic topographic map for path planning and navigation. Further articulation of the electrodes would generate a component of lift depending on the articulation angle. This selective maneuvering capability would lead to electrodynamic (rather than aerodynamic) flight. In this context, a potential field approach to path planning for navigation (Quadrelli et al. 2004) is a likely candidate. For navigation, the important determination is which low-altitude ranging sensors (i.e., altimetry) would be needed closer to the ground, if it would be more advantageous for the E-Glider to descend/ascend cyclically in response to solar illumination condition, or what is needed for stable station keeping. Another concern is how to differentially bias the charge on the surfaces relative to the body being orbited since solar wind can cause charge neutralization within a fraction of a second on exposed spacecraft surfaces. To provide continually varying charge emission to control the spacecraft potential relative to the space environment and asteroid, proper orbital design will significantly mitigate this concern by leveraging the natural charging, first hovering

in the dark side (where both the E-Glider and the surface are charged negatively) and then approaching the positively charged surface in the sunlit side at much lower altitude when both the E-Glider and the surface are charged positively. E-Glider navigation requires a local measurement of the direct current (DC) electric field near the spacecraft (and a feedback loop for control). Measuring DC electric fields requires double-probe sensors on long deployable booms (typically 30 m or more in 1-AU solar wind). These measurements are a function of the spacecraft's electrostatic environment including photo- and secondary emission, current bias setting, etc.

In the rest of this section, we consider the E-Glider as an extended body, particularly a dipole. We examine the possibility for a dumbbell E-Glider to achieve a hovering position with a desired attitude on the sunlit side in the RIC reference frame starting from a given initial position. Since the equations of translational motion and the ones of the attitude motion are highly coupled for a dumbbell spacecraft, the assessment of the capabilities of an E-Glider to perform both attitude and orbital control together by exploiting the electrostatic force and torque is fundamental. The coupled control is here developed for the planar case. An initial attempt to control a point mass E-Glider by acting on both the net and the differential charges has been performed in Corradino (2018) but under strong assumptions (e.g., the linearity of the electrostatic field). In this work, the case in which the spacecraft is composed by separated electrostatically active masses linked by a rigid tether controlled by adopting a control strategy similar to the one presented in Corradino (2018) is investigated. The problem is reduced to the planar case by considering only the xy -plane in the RIC reference frame.

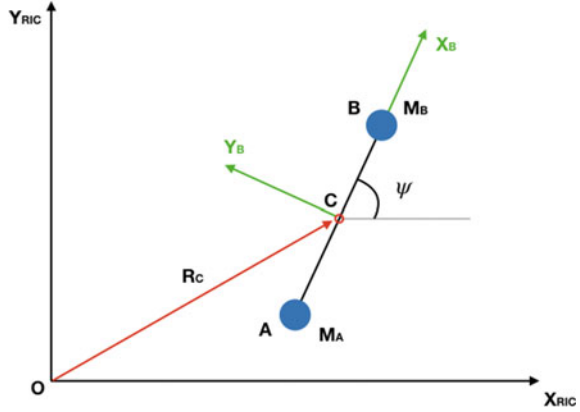
2.14.2 Assumptions of the Hovering Dipole Model

The system can be represented in the planar case as in Fig. 2.59. The position of the center of mass \mathbf{R}_c and the attitude angle ψ , defined as the angle between the x -direction in the RIC reference frame and the x -direction in the BF reference frame (see Fig. 2.59), describe the system under investigation.

The objective is the achievement of a desired hovering configuration defined by \mathbf{R}_{cd} and ψ_d starting from a given initial position \mathbf{R}_{c0} and ψ_0 . The hovering conditions have been previously found under the assumption of the Clohessy–Wiltshire equations (see Sect. 2.13.2). Here, the RIC reference frame is used to derive an ideal control law, but the terms related to the heliocentric motion of the spacecraft have been neglected; since the order of magnitude of the apparent forces given by the non-inertial RIC reference frame is as low as their effects result in a periodic motion with a characteristic time of about one Earth year, they can be safely neglected in first approximation.

The equations of motion (and so all the terms related) are reduced to the 2D case (only x - and y -components). The electrostatic effects can be modulated by changing the charge of the two point masses A and B. To achieve the desired hovering condition, the total resultant charge must be at least equal to the one obtained in

Fig. 2.59 Reference frame for the control analysis



the previous analysis for the case of a point mass spacecraft (see Sect. 2.13.2). Hence, $Q_d = Q_{Hover}$, where Q_{Hover} can be computed by evaluating Eq. (2.117) at the desired final position and configuration. The level of charge that allows the E-Glider to electrostatically levitate at each position over the asteroid on the sunlit side must be evaluated at each instant of time. The level of charge needed to hover at the current position and the one needed to achieve the final hovering can be decoupled by writing $Q(t) = Q_d + dQ(t)$, where $dQ(t)$ is the net charge control variable.

Dumbbell spacecraft can be considered as a single dipole immersed in the electrostatic field given by the PIC analysis. The basic physics of an electric dipole indicates that the first moment of charge can be obtained by differentially charging the end-points of the dipole itself, hence, by differentially charging A and B. By assuming, for example, a charge $-q$ on mass A and $+q$ on mass B, a dipole moment defined as $S_q = qL$ is generated. The electric dipole moment acts on the direction from the negative to the positive charge (along the x-direction in BF reference frame); thus, it can be expressed in RIC frame components as

$$\begin{aligned} S_{q,x} &= qL \cos \psi \\ S_{q,y} &= qL \sin \psi \end{aligned} \quad (2.120)$$

If the differential charge on A is the positive one, the dipole moment acts in the $-x$ direction in a BF frame. The differential charging affects at the same time the dynamics of the position of the center of mass and the attitude dynamics, since the dipole moment generated tends to align the spacecraft with the local electric field (generating the rotational effects given by $\mathbf{S}_q \times \mathbf{E}(\mathbf{R}_c)$), but it also gives a translational component since the E-field is a local property. By adopting the formulation derived in Sect. 2.9.1, the electrostatic acceleration can be expressed as

$$\frac{\mathbf{f}_e}{M} = \frac{\mathbf{Q}}{M} \mathbf{E}(\mathbf{R}_c) + \mathbf{G}_e(\mathbf{R}_c) \frac{\mathbf{S}_q}{M} \quad (2.121)$$

where $M = M_A + M_B$, $Q = Q_A + Q_B$, $\mathbf{E}(\mathbf{R}_c)$ is the 2D electric field evaluated at \mathbf{R}_c , and $\mathbf{G}_e(\mathbf{R}_c)$ is the 2D electrostatic gradient tensor evaluated at \mathbf{R}_c . Equation (2.121) shows how both the net and the differential charge affect the translational motion of the center of mass.

Since the reference frame used is the RIC reference frame and since the analysis is reduced to the planar case, the solar radiation pressure affects only the motion in the x -direction in the RIC frame by assumption. Due to the rigid dumbbell spacecraft model adopted for the E-Glider, both the total gravitational force acting on the equivalent mass M located on the center of mass and the effects of the decentralized masses M_A and M_B given by the gravity gradient term \mathbf{G}_g must be considered. This last contribution can be expressed as (Beletsky and Lavin 1993)

$$\mathbf{G}_g = \frac{3\mu}{R_c^4} \left[\frac{1}{2}(J_{yy} + J_{zz}) - J_{xx}; J_{xy}; J_{xz} \right]^T \quad (2.122)$$

where

$$\begin{aligned} J_{ii} &= M_A i_A^2 + M_B i_B^2 \text{ with } i = x, y, z \\ J_{ij} &= M_A i_A j_A + M_B i_B j_B \text{ with } i = x, y, z \text{ and } i \neq j \end{aligned}$$

with x, y, z components of the vector $\mathbf{r} = \mathbf{R} - \mathbf{R}_c$ in the local vertical/local horizontal (LVLH) reference frame with the \mathbf{R} position vector in the ACI reference frame of the part considered. Hence, $\mathbf{f}_g = -\frac{\mu M}{R_c^3} \mathbf{R}_c + \mathbf{G}_g$. Also in this case, the vector and the tensor must be reduced to the 2D case by taking only the components related to the x - and y -axis.

The Euler equation previously discussed can be used for the rotational motion. By assuming that the angular rate of the spacecraft with respect to the inertial reference frame expressed in the body reference frame is $\boldsymbol{\omega} = (0, 0, \dot{\psi})^T$ and by assuming that the BF reference frame is aligned with the principal axis of inertia, we obtain that $\mathbf{J}\boldsymbol{\omega} \times \boldsymbol{\omega} = 0$. Thus, the cardinal equation for the rotation around the z -axis in the BF frame can be written as $J_{zz}\ddot{\psi} = T_z$. Solar radiation pressure torque is assumed to be negligible in this analysis, while the electrostatic torque can be computed as $T_{e,z} = \mathbf{S}_q \times \mathbf{E}(\mathbf{R}_c) + T_{eG,z}$ (see Sect. 2.9.1). $\mathbf{S}_q \times \mathbf{E}(\mathbf{R}_c)$ gives the effect of the dipole moment which tends to align the dipole itself with the local electrostatic field, while $T_{eG,z}$ is the third element of the vector \mathbf{T}_{eG} . The vector \mathbf{T}_{eG} is defined by the second moment of charge \mathbf{I}_q ; thus, it depends on the level of charge of the spacecraft. From geometric considerations, we can easily verify that \mathbf{I}_q is symmetric for a dumbbell spacecraft. Moreover, $I_{q,yy}$ and $I_{q,zz}$ are always equal for the spatial symmetry of both the masses and the charges.

If the problem is reduced to the planar case, the vector \mathbf{T}_{eG} has only the component $T_{eG,z}$ different from zero. In order to develop an electrostatic control system based on the charge level of the points A and B, $T_{eG,z}$ must be formulated as an explicit function of the charge Q (since the terms related to the differential charge are auto-balanced). By starting from the computation of \mathbf{I}_q and after some mathematical

steps, we can write $T_{eG,z} = G_{e,xy}(\mathbf{R}_c) \frac{L^2}{4} Q$. This equation indicates that the higher the distance between the charges, the higher the torque given by the electrostatic gradient. By considering the symmetry relations and by remembering that we are analyzing the planar case, we can write the z-component of the gravitational torque as $T_{g,z} = G_{g,xy}(\mathbf{R}_c)(J_{yy} - J_{xx})$. The coupled equations of motion can be written as

$$\begin{aligned} {}^r \ddot{\mathbf{R}}_c &= -\frac{\mu}{R_c^3} {}^r \mathbf{R}_c + \frac{{}^r \mathbf{G}_g}{M} + \frac{{}^r \mathbf{f}_p}{M} + \frac{{}^r \mathbf{E}({}^r \mathbf{R}_c)}{M} Q + \frac{{}^r \mathbf{G}_e({}^r \mathbf{R}_c)}{M} {}^r \mathbf{S}_q \\ \ddot{\psi} &= {}^b G_{g,xy}({}^r \mathbf{R}_c) \left(\frac{J_{yy} - J_{xx}}{J_{zz}} \right) + {}^b G_{e,xy}({}^r \mathbf{R}_c) \frac{L^2}{4J_{zz}} Q + ({}^b \mathbf{S}_q \times {}^b \mathbf{E}({}^r \mathbf{R}_c)) \frac{1}{J_{zz}} \end{aligned} \quad (2.123)$$

In Eq. (2.123), the equations for the translation are in the RIC reference frame, while the equation for the rotation is in the BF frame. The electrostatic effects depend on both Q and \mathbf{S}_q , which can be assumed to be the inputs of the system (even if the real control variables are the net charge Q and the differential charge q); thus, the redefinition of the equation of motion as an explicit function of Q and \mathbf{S}_q is needed in order to properly formulate a control law. This implies that \mathbf{S}_q must be written in a common reference frame, or the RIC or the BF frame for all the equations. For the planar 2D case, the RIC reference frame and the BF reference frame have the z-axis in common, such that the rotation matrix that allows switching from the inertial to the body reference frame ${}^b \mathbf{R}_r$ is the elementary rotation matrix about the z-axis of an angle ψ . By defining ${}^b \mathbf{T}_e = {}^b \mathbf{S}_q \times {}^b \mathbf{E}({}^r \mathbf{R}_c)$, we can write ${}^b \mathbf{T}_e = {}^b \mathbf{R}_r {}^r \mathbf{T}_e$ with ${}^r \mathbf{T}_e = {}^r \mathbf{S}_q \times {}^r \mathbf{E}({}^r \mathbf{R}_c)$. In the RIC reference frame, for the case under analysis, we have ${}^r \mathbf{S}_q = ({}^r S_{q,x}, {}^r S_{q,y}, 0)^T$, thus ${}^r \mathbf{T}_e = (0, 0, {}^r E_y {}^r S_{q,x} - {}^r S_{q,y} {}^r E_x)^T$. By applying the rotation matrix ${}^b \mathbf{R}_r$ to ${}^r \mathbf{T}_e$, we can verify that since the ${}^r \mathbf{T}_e$ has only the z-component different from zero and since the RIC and the BF reference frames have the z-axis in common, ${}^b \mathbf{T}_e = (0, 0, {}^r E_y {}^r S_{q,x} - {}^r S_{q,y} {}^r E_x)^T = {}^r \mathbf{T}_e$; thus, Eq. (2.123) can be rewritten by switching \mathbf{S}_q from the BF frame to the RIC reference frame (to obtain ${}^r \mathbf{S}_q$ in all the equations) as

$$\begin{aligned} {}^r \ddot{\mathbf{R}}_c &= -\frac{\mu}{R_c^3} {}^r \mathbf{R}_c + \frac{{}^r \mathbf{G}_g}{M} + \frac{{}^r \mathbf{f}_p}{M} + \frac{{}^r \mathbf{E}({}^r \mathbf{R}_c)}{M} Q + \frac{{}^r \mathbf{G}_e({}^r \mathbf{R}_c)}{M} {}^r \mathbf{S}_q \\ \ddot{\psi} &= {}^b G_{g,xy}({}^r \mathbf{R}_c) \left(\frac{J_{yy} - J_{xx}}{J_{zz}} \right) + {}^b G_{e,xy}({}^r \mathbf{R}_c) \frac{L^2}{4J_{zz}} Q + ({}^r \mathbf{S}_q \times {}^r \mathbf{E}({}^r \mathbf{R}_c)) \frac{1}{J_{zz}} \end{aligned} \quad (2.124)$$

2.14.2.1 Proportional-Derivative (PD) Active Control

In an initial analysis, a proportional-derivative (PD) controller has been selected to command the required accelerations that must be given to reach the desired configuration. The PD receives as input the state vector of the errors, composed of the error itself (given by the difference between the position at a given instant of time and the reference position) and its time derivative and gives as output the required

accelerations by scaling by a factor K_P the error and by a factor K_D the error time derivative. A PD controller has been used for each variable in the analysis proposed, so the system can be assumed to be composed by three parallel PDs, each one with its own K_P and K_D parameters to be tuned. The parameters of each PD controller have been selected by starting from reference values (Corradino 2018) and then adjusted for the case under analysis by using the trial-and-error method. The outputs of the PDs are the inputs of an ideal actuator which processes these demanded accelerations in order to obtain the net charge and the differential charge needed. The control system is thought to shift the spacecraft from an equilibrium condition (artificially generated by tuning the value of Q) to the next one, hence starting from \mathbf{R}_{c0} to \mathbf{R}_{cd} . The total charge is considered as $Q = Q_d + dQ$ such that $\ddot{\mathbf{R}}_c(\mathbf{R}_c, Q, q) = \ddot{\mathbf{R}}_{cd}(\mathbf{R}_{cd}, Q_d, 0) = 0$. The control equations are derived by assuming that \mathbf{R}_{c0} is close to \mathbf{R}_{cd} (small displacements) such that once that the error state vector is processed by the PD block, the following control equations can be written:

$$\begin{Bmatrix} \Delta \ddot{x}_c \\ \Delta \ddot{y}_c \\ \Delta \ddot{\psi} \end{Bmatrix} = \begin{bmatrix} \frac{r E_x}{r M_y} & \frac{r G_{e,xx}}{r M_y} & \frac{r G_{e,xy}}{r M_y} \\ \frac{r E_y}{M} & \frac{r G_{e,yy}}{r M_y} & \frac{r G_{e,xy}}{r M_y} \\ b G_{e,xy} \frac{L^2}{4J_{zz}} & \frac{r E_y}{J_{zz}} & -\frac{r E_x}{J_{zz}} \end{bmatrix} \begin{Bmatrix} dQ \\ S_{q,x} \\ S_{q,y} \end{Bmatrix} \quad (2.125)$$

where $\Delta \ddot{\mathbf{R}}_c = (\Delta \ddot{x}_c, \Delta \ddot{y}_c, \Delta \ddot{z}_c)$ and $\Delta \ddot{\psi}$ are the demanded accelerations given by the PDs. From simple geometrical relations, we can derive the equation in matrix form that relates the first moment of charge in the RIC reference frame to the differential charge q . Thus,

$$\begin{Bmatrix} dQ \\ S_{q,x} \\ S_{q,y} \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & L \cos \psi \\ 0 & L \sin \psi \end{bmatrix} \begin{Bmatrix} dQ \\ q \end{Bmatrix} \quad (2.126)$$

By defining as \mathbf{A} the matrix that relates the output of the PDs to the vector $(dQ, S_{q,x}, S_{q,y})^T$ and naming \mathbf{B} the matrix that links the vector $(dQ, S_{q,x}, S_{q,y})^T$ with the net charge and the differential charge, we can compute the ideal actuator equations by computing the pseudoinverse of the 3×2 $[\mathbf{A} \cdot \mathbf{B}]$ matrix.

2.14.3 Ideal Hovering Control for a Single-Dipole Spacecraft

The simulations are run by considering a spacecraft composed by identical spheres (which are the “point masses”) of 0.065 m of radius linked by a tether such that the distance between the center of mass (and charge) of sphere A from the center of mass (and charge) of sphere B is 10 m. The gravity model used for the asteroid is the point mass model. The ideal control developed relies on the fact that the inertia matrix of the spacecraft is perfectly known. The exact knowledge of the inertia matrix can be an issue for a long-term mission or for objects which stay in space for a long period of time. The measurements of the state of the system (thus position, velocity,

and acceleration of both the center of mass and the angular coordinate) are exact by assumption. Moreover, the electrostatic field and its gradient at each instant of time and for each position are assumed to be exactly known. This last assumption can be extremely difficult to be satisfied in practical cases since even if there is the capability of measuring the electrostatic field at each position through some Langmuir probes (Chen 2003), the time needed to know the electrostatic field must be infinitesimal, and the electrostatic field may also change as a function of time. Lastly, the control system is assumed to be capable of determining the needed net and differential charges instantaneously and the actuator to be capable of providing them to the system itself in a short period of time. The masses are supposed to have the capability of assuming a certain level of charge as it is transmitted as input, without time delays. To make the ideal actuator model more realistic, a certain threshold for the maximum and the minimum charge that the actuator can deliver has been considered in the simulations. If the net or the differential charges reach the saturation level, the actuator model automatically scales the other charge to maintain the alignment of the resultant force equal to the one that can be obtained without the saturation limit.

Due to the strong coupling already noticed and since both the net control charge and the differential control charge affect both the translational and the rotational dynamics together, we analyzed first the capability of achieving a desired position disregarding the attitude control; then the attitude control was added to explore the possibility of performing both the translation and the attitude control simultaneously.

This control strategy aims to verify the possibility of achieving a desired final position given an initial state vector by imposing a control action only on the translation of the spacecraft. The attitude dynamics is not directly controlled here, but it evolves under the electrostatic effects since the equations of motions are coupled. In particular, due to the electrostatic torque given by the total charge (term \mathbf{T}_{eG}) and due to the dipole moment effects, the spacecraft is expected to start spinning about the z-axis in the BF frame. A certain level of charge (Q_0) must be always maintained in order to guarantee the hovering condition, such that a certain level of torque is always acting on the spacecraft. Each change in the charges causes a torque acting on the spacecraft which cannot be balanced by any other torque (since the spacecraft is composed of a single dipole and does not have any momentum exchange device), and which induces a rotation of the spacecraft itself. The control of such a spacecraft is also made extremely difficult by the fact that the forces along the x- and y-axis in the RIC reference frame depend on the attitude of the dipole, making the control action and the translational motion itself strictly bounded to the attitude motion and constrained by that. The initial scenario of the simulation is given by the spacecraft hovering at an arbitrarily chosen initial position and attitude over the sunlit face of the main body, as reported in Table 2.9.

Since the spacecraft at the initial condition is hovering, the initial velocities are all equal to zero. The reference final position has been arbitrarily chosen as reported in Table 2.10.

The maximum charge level is set to be equal to $\pm 750 \mu\text{C}$. All the considerations and issues related to charging a spacecraft to such a high level have not been taken into account in this analysis. The simulation time is set to 20 h (as done in Corradino

Table 2.9 Initial position and attitude in hovering

x [m]	y [m]	ψ [rad]
-25	2	1

Table 2.10 Reference position in hovering

x [m]	y [m]
-23	1

Fig. 2.60 Time evolution of the position errors for the hovering control without attitude control

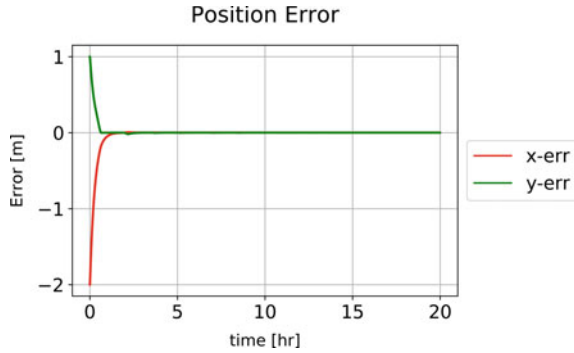


Fig. 2.61 Forces acting on the spacecraft for hovering control without attitude Control



2018). The results of the simulation are reported in Fig. 2.60 by the means of the error between the position at the current instant of time and the reference position (e.g., $e_x(t) = x(t) - x_{ref}$).

Both the position errors are dropped to zero in a short amount of time by using the ideal control law previously discussed. From Fig. 2.60, the possibility to maintain the hovering condition is evident even if the desired location is not a stable equilibrium point, as in the case here presented (see Sect. 2.13.2). The norm of the forces acting on the spacecraft is in Fig. 2.61.

As expected, the electrostatic force has a peak during the first instants of time due to the control action demanded by the PD controller. When the spacecraft reaches

Fig. 2.62 Electrostatic force components for the hovering control without attitude Control

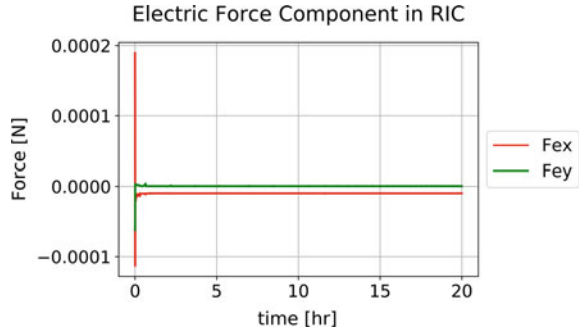
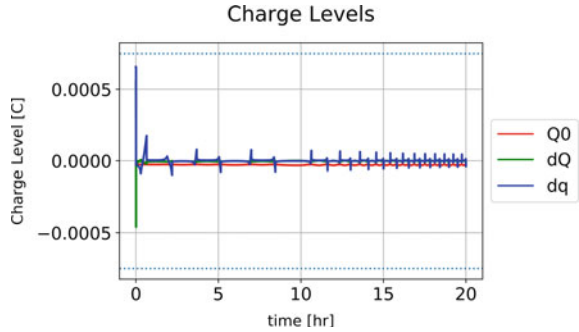


Fig. 2.63 Control charges for hovering control without attitude control



the desired final position, the electrostatic force is maintained almost constant in order to keep the position. The delivered electrostatic force is always higher than the gravitational force acting on the spacecraft, which is due to the fact that to maintain the hovering position, the electrostatic force must counter the summation of both the gravitational and the SRP forces. Hence, the hovering is confirmed to be an energetically inefficient strategy. The electrostatic force is expected to be delivered mostly along the x -direction in the RIC frame in order to maintain the hovering position. This is confirmed in Fig. 2.62.

The levels of charge (both net and differential) needed are reported in Fig. 2.63 together with the level of charge Q_0 that must be maintained to achieve the hovering at the reference position. Once the final position has been achieved, the net charge dQ required to maintain the final position is almost zero, while a continuous control on the differential charge is needed to avoid drifting of the spacecraft. In Fig. 2.63, the blue dotted lines show the saturation levels imposed for both the total and the differential charge.

To reduce the control effort and the pulsations in the differential charge, a dead band in which the spacecraft is free to drift should be considered. The dead band can be defined in terms of both x and y positions. If the spacecraft is inside this region, the only charge applied is Q_0 , while if the spacecraft exits the predefined region, the

Fig. 2.64 Electrostatic torque for hovering control without attitude control

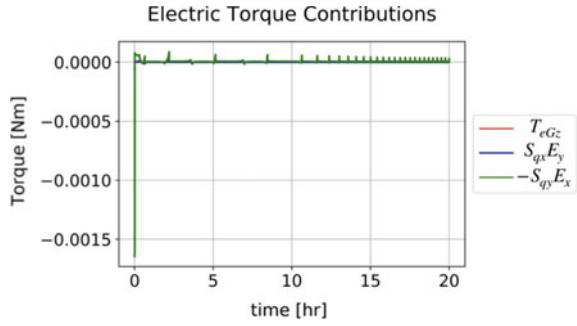
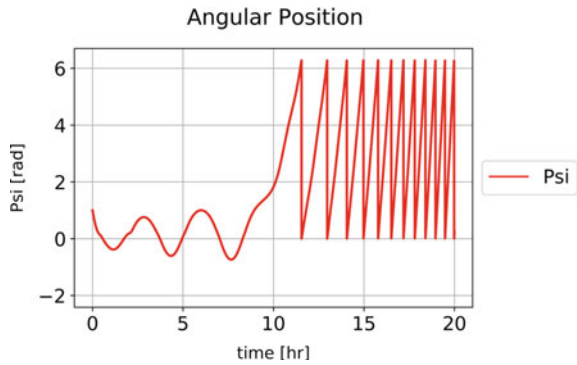


Fig. 2.65 Angular position of the spacecraft during hovering control without attitude Control



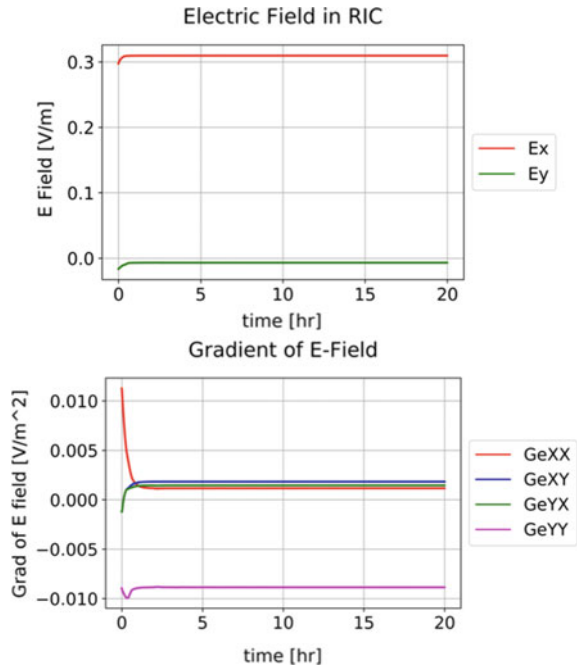
complete control law previously described is applied again. The differential control charge also generates a torque that affects the attitude motion of the spacecraft, modifying the spinning rate. The electrostatic torque contributions are reported in Fig. 2.64.

Figure 2.64 points out that the torques introduced during the position keeping are all positive, meaning that the spacecraft rotates in a counterclockwise direction in the RIC reference frame around the z-axis with an increasing spin rate. This result is confirmed by Fig. 2.65.

The highest contribution to the total electrostatic torque, in this case, is given by $-S_{q,y}E_x$ because the electric field is mostly directed along the x-direction in the RIC reference frame (see Fig. 2.66).

As expected, after an initial phase in which the spacecraft angular position oscillates under the torques given by the control charges required to translate the spacecraft itself from the initial position to the reference one, the angular position becomes positive during the position-keeping phase, and the spin rate increases. To limit the increment in the spin rate, the dead-band control law previously discussed can be used (Fig. 2.66).

Fig. 2.66 Electrostatic field (upper) and gradient of the electrostatic field (lower) components in 2D in the RIC frame



2.14.4 Position and Attitude Control for Single-Dipole Spacecraft

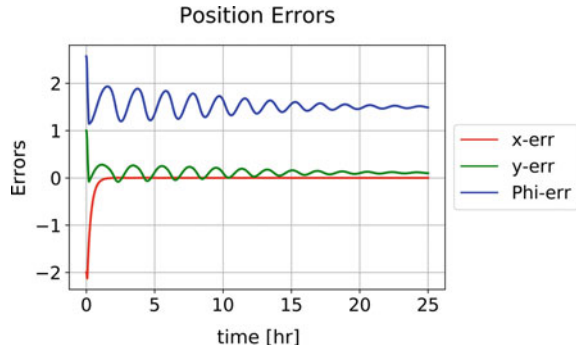
Once the capabilities of achieving a desired position in hovering starting from an initial hovering condition with a single-dipole spacecraft have been explored, the possibility of achieving a desired final position with a desired final attitude starting from a given initial hovering condition is analyzed. The initial conditions are maintained equal to the ones reported in Table 2.9, while the final reference conditions are set as reported in Table 2.11.

The spacecraft has a line connecting the two masses perpendicular to the subsolar axis (the $-x$ direction in the RIC reference frame) at the final desired attitude. The scenario described in this case is far more complex than the previous one because here a single electrostatic dipole that acts as an ideal actuator (that has coupled effects on the rotational and on the translational dynamics) should control a spacecraft with the rotational and the translational dynamics highly coupled. Furthermore, if the electric

Table 2.11 Reference position and attitude in hovering

x [m]	y [m]	ψ [rad]
-23	1	$-\pi/2$

Fig. 2.67 Position and attitude errors for hovering and attitude control



dipole is aligned with the electric field described in each point by the PIC results, the torque component given by the first dipole moment is null, such that the modulation of the net charge gives the only control term on the attitude motion.

Figure 2.67 shows the results of the simulation which has been selected as the best one since it leads to a “stable” final configuration.

By considering the results in Fig. 2.66, we expected that the highest torque will be delivered by the term $-E_x S_{q,y}$. Moreover, the term $G_{e,yy}$, which is the highest in modulus, causes a strong coupling between the y-position and the attitude of the spacecraft (see also Eq. (2.124)), making the control extremely difficult. On the contrary, the x-position is not strongly affected by the attitude dynamics, since the predominant component of the control action is the net charge term. Figure 2.67 confirms this expectation. The reference x-position is achieved in a relatively low amount of time (about 2 h), while after 25 h of simulation, both the y-position and the angular position dynamics show low-amplitude and low-frequency oscillations around a reference error value which is different from zero. In particular, the residual error for the y-position is in the order of centimeters far from the reference value reported in Table 2.11, while the angular position error is of about 1.5 radians.

Hence, a PD controller can drive single-dipole spacecraft to a certain fixed final configuration, performing both the translational and the attitude control simultaneously. The main problem is that the PD controller is not capable of making the error equal to zero in steady-state conditions. This drawback is due to the proportional part of the PD controller; thus, a simple solution could be to augment the gain K_P associated with the proportional error, putting a bigger effort in the control action and making the steady-state error closer to zero or to substitute the PD controller with a proportional-integral-derivative (PID) one. The integrative part of this new controller will ensure the zeroing of the steady-state error.

The biggest issue for the case here analyzed is the coupling of the equations of motions and, as a consequence, also the ones of the actuator. In particular, the net charge is always different from zero, and it gives a constant torque which always acts on the attitude motion of the spacecraft. By changing the structure of the spacecraft, the torque term related to $T_{eG,z}$ can be canceled out for symmetry, making it possible to partially decouple the equation of motions and the equation of the ideal actuator

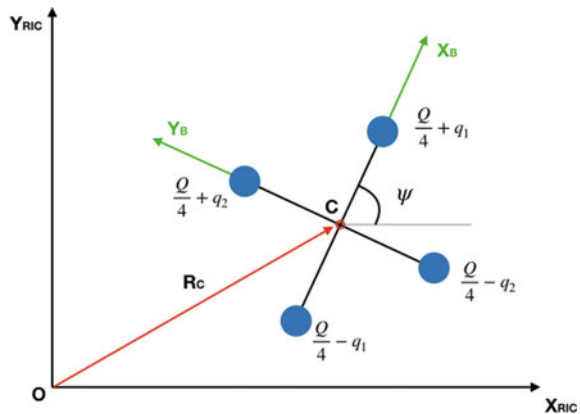
(indeed, the adoption of a new geometry can be seen as a change of the ideal actuator). The new geometry adopted and the new hovering control results are described in the following section.

2.14.5 Ideal Hovering Control for a Double-Dipole Spacecraft

The spacecraft with the new geometry adopted is composed by four spheres connected by two rigid tethers in a cross shape. The length of the tether is maintained equal to the one of the single-dipole spacecraft previously analyzed (10 m). The radius of each sphere is equal to $R_s = 0.046$ m such that each sphere of the new spacecraft has half cross-sectional area with respect to the single-dipole spacecraft. A mass equal to 0.665 Kg per sphere has been fixed to maintain the total mass of the spacecraft constant. Figure 2.68 shows the new geometry adopted and the reference frames used.

The total charge Q is now split in the four spheres, resulting in a net charge equal to $Q/4$ per each sphere. The differential charge associated with the spheres aligned with the x-axis in the BF reference frame is named q_1 (or first differential charge), while q_2 (or second differential charge) is the differential charge of the spheres on the y-axis in the BF frame. If the charges are assigned as in Fig. 2.68, the vector S_q in the RIC reference frame can be redefined as the summation of the effects given by the two dipoles formed by the masses along the x-axis and along the y-axis in the BF frame; thus, the matrix that links the vector $(dQ, S_{q,x}, S_{q,y})^T$ with the net and the differential charges can be rewritten as

Fig. 2.68 Reference frame for the control analysis: four-sphere spacecraft



$$\begin{Bmatrix} dQ \\ S_{q,x} \\ S_{q,y} \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & +L\cos\psi & -L\sin\psi \\ 0 & +L\sin\psi & +L\cos\psi \end{bmatrix} \begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} \quad (2.127)$$

This new configuration is double symmetric with respect to the x-axis and the y-axis in the BF frame. Thus, the term given by the gravity gradient can be canceled out by the equation of motion for the rotation ψ . The symmetry of the spacecraft leads also to $I_{q,yy} = I_{q,xx}$; thus, $T_{eG,z} = G_{e,xy}(I_{q,yy} - I_{q,xx}) = 0$ in the Euler equation for the attitude motion. Thus, the equations of motion for the four-sphere spacecraft can be written as

$$\begin{aligned} {}^r\ddot{\mathbf{R}}_c &= -\frac{\mu}{R_c^3} {}^r\mathbf{R}_c + \frac{{}^r\mathbf{G}_g}{M} + \frac{{}^r\mathbf{f}_p}{M} + \frac{{}^r\mathbf{E}({}^r\mathbf{R}_c)}{M} Q + \frac{{}^r\mathbf{G}_e({}^r\mathbf{R}_c)}{M} {}^r\mathbf{S}_q \\ \ddot{\psi} &= ({}^r\mathbf{S}_q \times {}^r\mathbf{E}({}^r\mathbf{R}_c)) \frac{1}{J_{zz}} \end{aligned} \quad (2.128)$$

In Eq. 2.128, the effects of the total charge are present only in the translational equations of motion. Moreover, in the ideal case here analyzed, the rotational motion is affected only by the electrostatic effects induced by the dipoles, which are directly delivered by the actuator. Thus, a finer control of the rotational motion can be achieved.

By following the very same procedure described for the case of a single-dipole spacecraft, the accelerations commanded by the controller can be linked to the vector $(dQ, S_{q,x}, S_{q,y})^T$ through the matrix \mathbf{A} , here redefined as

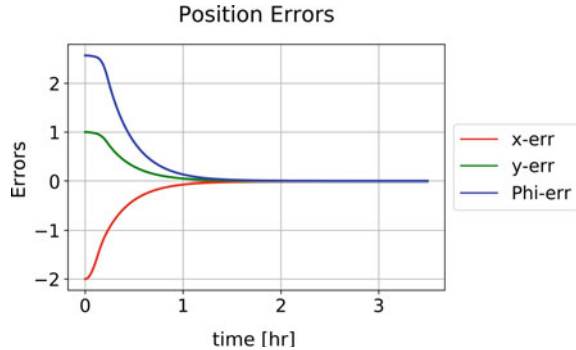
$$\begin{Bmatrix} \Delta\ddot{x}_c \\ \Delta\ddot{y}_c \\ \Delta\ddot{\psi} \end{Bmatrix} = \begin{bmatrix} \frac{{}^rE_x}{rM} & \frac{{}^rG_{e,xx}}{rM} & \frac{{}^rG_{e,xy}}{rM} \\ \frac{{}^rE_y}{rM} & \frac{{}^rG_{e,yx}}{rM} & \frac{{}^rG_{e,yy}}{rM} \\ 0 & \frac{J_{zz}}{J_{zz}} & -\frac{{}^rE_x}{J_{zz}} \end{bmatrix} \begin{Bmatrix} dQ \\ S_{qx} \\ S_{qy} \end{Bmatrix} = \mathbf{A} \begin{Bmatrix} dQ \\ S_{qx} \\ S_{qy} \end{Bmatrix} \quad (2.129)$$

The problem of the loss of controllability due to the alignment of the electric dipole with the electrostatic field is no more an issue since even if one of the dipoles is aligned with the electric field lines, the other dipole is still capable of delivering a certain level of torque due to the perpendicularity of the two dipoles. Due to the partial decoupling of the equations of motions and of the actuator achieved with the geometry here presented, the coupled orbital and attitude control of the E-Glider can be achieved more easily with respect to the single-dipole case.

2.14.6 Position and Attitude Control for a Double-Dipole Spacecraft

Here, the results for the coupled control of both attitude and orbital motion to achieve a desired hovering condition with a desired angle of the x-axis in the BF reference

Fig. 2.69 Position and attitude errors: double dipole



frame with respect to the x -axis in the RIC reference frame are discussed. The parameters used are the same presented for the single-dipole case (see Tables 2.9 and 2.11). The saturation level for this simulation is set equal to $\pm 750 \mu\text{C}$ for both the net and the differential control charges. In this case, the control law adopted is capable of driving to zero the errors both in the translational motion and in the attitude motion without any overshoot and in a relatively low amount of time, as Fig. 2.69 points out.

The higher level of decoupling of the equations reached by adding the second dipole makes possible finer control of both the attitude and the orbital motion in the ideal case. By comparing the error rejection here presented with the results obtained for the coupled control in the single-dipole case, the improvement offered by adopting the second dipole is clear, as all the errors are driven to zero almost simultaneously without any oscillations or coupling both in the y error and in the angular error.

In the upper part of Fig. 2.70, the electrostatic force components $F_{e,x}$ and $F_{e,y}$ are reported, while in the lower portion, there are the two terms $S_{q,x}E_y$ and $-S_{q,y}E_x$, the sum of which gives the electrostatic torque on the spacecraft. The biggest effort is related to the x -component of the electrostatic force also in this case. Regarding the electrostatic torque terms, after an initial phase of almost constant low torque (in which the control action is mostly devoted to reducing the error in the x -position), there is a strong peak due to an abrupt variation of the component $-S_{q,y}E_x$. After this strong peak, both the terms of the electrostatic torque have a positive value which gradually decreases to almost zero once the ψ error becomes null.

In the upper portion of Fig. 2.71, the commanded charge levels to obtain the previously discussed electrostatic forces and torques are reported, while in the lower part, the actuator electrostatic effects are presented. The control charges needed are well under the saturation limit imposed, also during the initial transient phase in which the charges reach the highest level due to a strongest control action needed. The strong variations and the changes in the polarity of both q_1 and q_2 during the initial phases determine the previously noticed peaks in the electrostatic force and torque. In particular, by considering the initial attitude of the spacecraft, even if the magnitudes of the differential charges are comparable, values of $S_{q,x}$ notably higher than those of $S_{q,y}$ are expected during the initial phase. This is confirmed by the plot

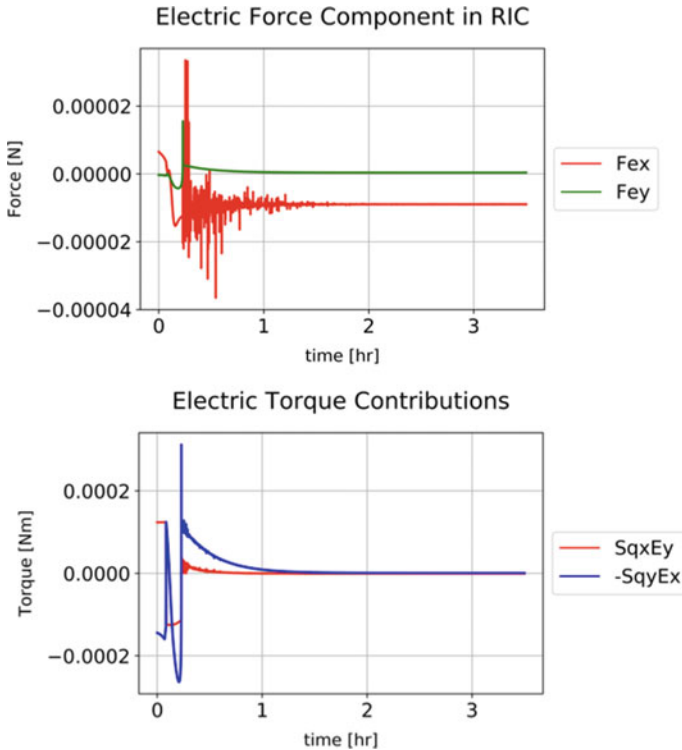


Fig. 2.70 Electrostatic force (upper) and torque (lower) components: double dipole

of the actuator electrostatic effects. The high value of $S_{q,x}$ in this phase, coupled with the also relatively high levels of the net control charge dQ , gives the peaks previously noticed in the $F_{e,x}$ components. The strong peak in the electrostatic torque is given by the $S_{q,y}$ component of the first momentum dipole even if $S_{q,x}$ is higher, due to the components of the electric field in the region considered. During the final phase in which the main objective is to maintain the hovering condition previously achieved, the charge levels are very low such that the resulting $S_{q,x}$ component is positive and slightly higher in modulus than $S_{q,y}$ to ensure the zero torque condition (given by $S_{q,x}E_y - S_{q,y}E_x = 0$) once the system reaches steady-state conditions. Since the component E_x of the electric field is always positive and greater than E_y (which is also negative) in the region of interest, the general rule $|S_{q,x}| > |S_{q,y}|$ can be derived for the case here analyzed. Moreover, $S_{q,x}$ and $S_{q,y}$ must have opposite signs in order to balance the torque contribution.

As said, the saturation level of $\pm 750 \mu\text{C}$ was arbitrarily imposed on the actuator, and it could be reduced to a more realistic value. A high limit of saturation for the levels of charge implies a wider range of feasibility for the maneuvers tested, and it makes it possible to increase the proportional gains K_P of the PD controller here adopted, reducing the error at steady state. This means that the proper definition

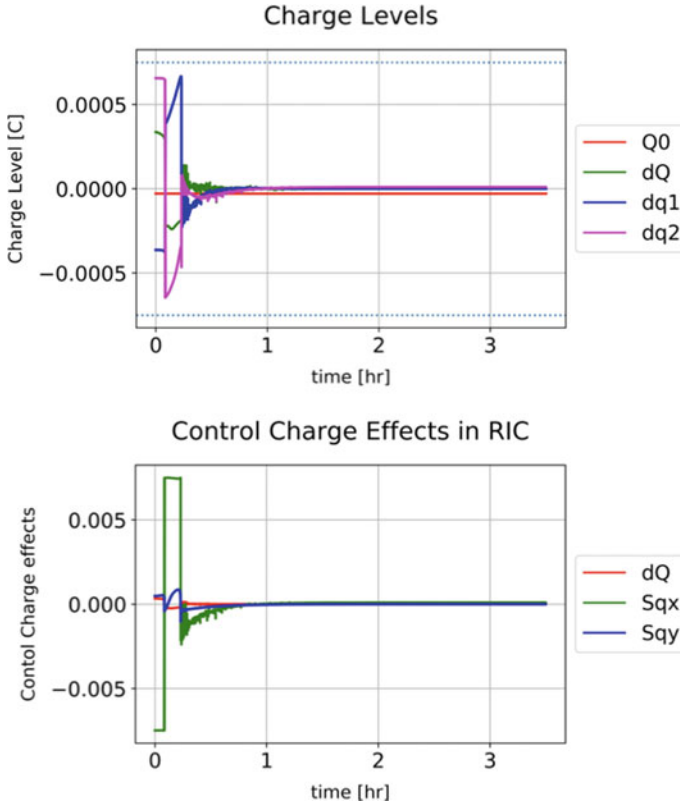


Fig. 2.71 Control charges (upper) and actuator electrostatic effects (lower): double dipole

of a saturation level affects the choice of the controller and the error in steady-state condition for a PD controller. The selection of a saturation charge below the predicted levels of Q/M defined in Sect. 2.13.2 is not possible; otherwise, the actuator will saturate before reaching the charge level needed to levitate over the asteroid surface, making the hovering not feasible. By considering Fig. 2.45, we can conclude that by reducing the saturation level, a sort of “no-fly” zone is created starting from the asteroid surface (the height of this zone increases as the saturation level decreases). This zone must be avoided to prevent the spacecraft from irreversibly collapsing on the asteroid.

2.14.7 Tether Length Sensitivity

In this section, an analysis of the effects of the length of the tether that links the four masses of the spacecraft is presented. The length of the tether can be seen as

Table 2.12 Tether length values for simulations

Value 1 [m]	Value 2 [m]	Value 3 [m]	Value 4 [m]	Value 5 [m]
0.25	1.00	2.00	5.00	10.00

a parameter of the actuator since it affects the outputs $S_{q,x}$ and $S_{q,y}$ delivered by the actuator in the RIC reference frame once the net control charge dQ and the differential charge q_1 and q_2 are given as input. During this analysis, five different spacecraft are considered. All of them are equal, except for the length of the tethers. The tether lengths used for these simulations are reported in Table 2.12.

As for the previous analysis, the tether length is considered as the separation distance between the center of mass of two spheres with opposite position with respect to the center of mass of the spacecraft. Since the charge for this analysis is assumed to be concentrated in the center of mass of each sphere, the length of the tether is the separation distance between the two charges that constitutes one of the two dipoles of the spacecraft. It is assumed that the two tethers of a single spacecraft are equal in length in order to maintain the symmetry of both the masses and the charges, making it possible to use the very same control law developed for the four-sphere spacecraft. The parameters used here are the very same as in previous simulations. The proportional and derivative gains K_p and K_d of the PD controllers have been kept constant and equal to the ones selected for the simulation discussed previously. Also, the initial and final conditions are kept constant with respect to the previous analysis in order to detect only the variations due to the different lengths of the tether selected. The saturation level is $\pm 750 \mu\text{C}$, which is high enough to ensure the convergence of all the error dynamics to about zero for all the cases here analyzed. Figure 2.72 presents the results obtained from the simulations of the cases reported in Table 2.12.

Only the level of charges and the rejection of the errors for each case are reported in the figure since it is possible to identify the effects given by changing the tether length by analysis of these two plots. Generally speaking, it is possible to identify two regions in both the plot of the charge level and the plot of the errors. The first region is referred to the transient phase and the second one to the hovering conditions maintained. The transient phase in the plot of the charge level is characterized by the abrupt variation of the differential charges and the high level of the net charge dQ . In the plots of the errors, the transient phase is identified as the region in which the errors move from the initial value to a band close to the condition of zero error and do not escape from this region. It is immediately clear that the length of the tether affects the duration of the transient phases. By increasing the length of the tether, the time duration of the transient phase is reduced. The reduction is more evident in the plot of the charges. It is better to remark that the levels of charge are influenced by the error through the accelerations imposed by the PD controllers and that the errors are affected by the charges since they define the level of force and torque applied to the system in a closed-loop system; thus, a reduction in the transient phase of the

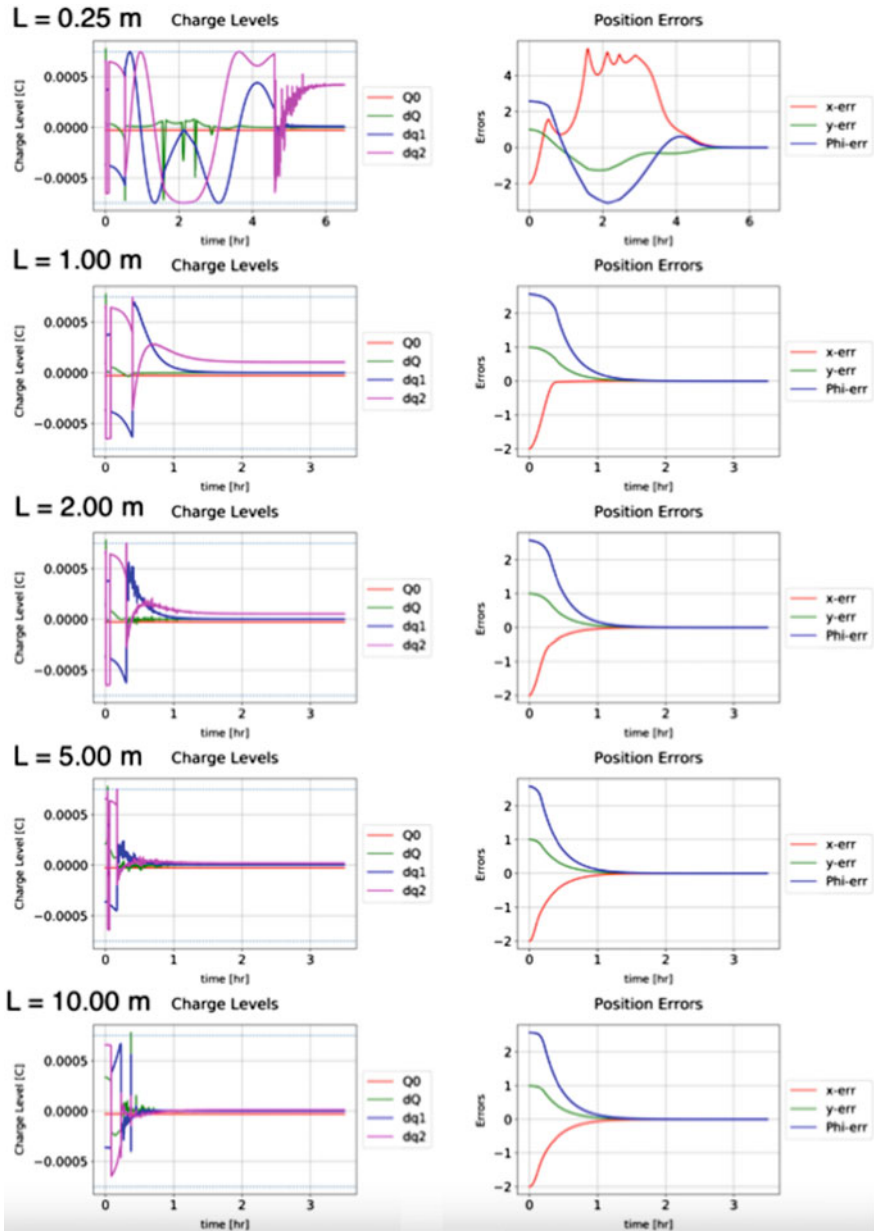


Fig. 2.72 Control charges (left) and error dynamics (right) for different tether lengths

charges implies a reduction in the transient phase of the errors. The reduction of the transient phase of the errors is evident by doing a comparison between the case of

tether length equal to 0.25 m and the one which measures 1 m. In the former case, the error is driven to zero in almost 6 h, while in the latter case, it takes only about 2 h. By doing the same comparison for the other case, it can be seen that the settling time (defined as the time at which the errors enter in the band close to zero and do not exit again) is reduced by just a few minutes. On the contrary, on the plot of the charges, it is evident that the initial transient phase is strongly reduced. The transient phase of the charges can be divided into two sub-phases; in the first one, the control charges are varying in a continuous way, and in the second one, they behave in an impulsive way. The increment of the length of the tether mostly reduces the time of the first phase, while the second one is not largely affected by the length of the tether. It can be noticed that the amplitude of the impulses in the control charges highlighted by the simulations during the second phase of the transient is extremely reduced if the tether is 1 m long. Moreover, during the transient phase in all the cases, except the last one reported, the charges reach the saturation level. Now the second part of the plots, the hovering conditions maintained, is taken into account. The increment of the tether length affects the level of the differential charges q_1 and q_2 needed to maintain the position achieved and to obtain the condition of zero torque previously identified as necessary to maintain the desired attitude. In particular, the reduction of the level of q_2 is evident. This is due to the fact that the level of q_2 at $\psi = -\pi/2$ affects only $S_{q,x}$, while q_1 affects only $S_{q,y}$. From this, and by recalling that the electric field in the region swept by the spacecraft during this maneuver has a positive component along the x-axis and a negative one along the y-axis lower in modulus than the positive one and also that the electrostatic torque can be computed as $T_e = S_{q,x}E_y - S_{q,y}E_x$, the reason for which a value for q_2 always notably higher than q_1 is obtained, it is rather clear. In general, it can be concluded that a higher tether length implies a lower level of charge needed to accomplish the same maneuver in a lower amount of time. For the cases analyzed here, it is also evident that for a tether length equal to 0.25 m, achievement of the previously defined hovering condition is extremely difficult, and the feasibility of the maneuver cannot be ensured since there is a phase during the transient in which the center of mass of the spacecraft is extremely close to the asteroid surface, reaching a minimum distance from the center of the asteroid of about 18 m; thus, it should be verified that the attitude conditions during this phase are such that the electrodes do not impact the surface of the asteroid. It can be noticed that the four peaks of maximum error in the x-position of the spacecraft with a tether of 0.25 m correspond to the peaks in the net charge $Q = dQ + Q_0$ equal to the saturation level. This analysis highlights the fact that adopting a short separation of the charges can make the mission unfeasible from the energetic point of view. This is due to the fact that the level of charge needed by each electrode affects the total power level required onboard; in particular, a higher charge means a higher electric potential (in modulus) of the spacecraft and thus a higher power needed to maintain the level of charge.

2.15 Including the Effect of Charging Electrodes

2.15.1 Spacecraft Charging

The power supply voltage for charging a spacecraft, which is regarded as the electrostatic potential of the spacecraft relative to the ambient plasma potential, is given by the equation below.

$$V_{sc} = \frac{Q}{C} \quad (2.130)$$

A spacecraft in the plasma environment around an asteroid collects charged particles. Besides, the spacecraft itself is also exposed to the solar radiation and emits photoelectrons. As a result, the charge of the spacecraft varies due to the current flux from/to the ambient plasma, as expressed by the equation below (Nitter et al. 1998; King et al. 2002).

$$\frac{dQ}{dt} = I_i - I_e - I_p + I_{p,sc} \quad (2.131)$$

where I_i , I_e , and I_p are the currents from the solar wind ions, the solar wind electrons, and the photoelectrons emitted from the asteroid surface, respectively; and $I_{p,sc}$ is the photoelectron current from the spacecraft. These currents can be expressed as follows (Nitter et al. 1998; Havnes et al. 1987; Hirata and Miyamoto 2012):

$$\begin{aligned} I_i &= \begin{cases} N_{sp} \cdot \pi r_{sp}^2 e n_i v_i \left(1 - \frac{2eV_{sc}}{m_i v_i^2}\right) & (V_{sc} < \frac{m_i v_i^2}{2e}) \\ 0 & (V_{sc} \geq \frac{m_i v_i^2}{2e}) \end{cases} \\ I_e &= \begin{cases} N_{sp} \cdot \pi r_{sp}^2 e n_e \sqrt{\frac{8k_B T_e}{\pi m_e}} \exp\left(\frac{eV_{sc}}{k_B T_e}\right) & (V_{sc} < 0) \\ N_{sp} \cdot \pi r_{sp}^2 e n_e \sqrt{\frac{8k_B T_e}{\pi m_e}} \left(1 + \frac{eV_{sc}}{k_B T_e}\right) & (V_{sc} \geq 0) \end{cases} \\ I_p &= \begin{cases} N_{sp} \cdot \pi r_{sp}^2 e n_p \sqrt{\frac{8k_B T_p}{\pi m_e}} \exp\left(\frac{eV_{sc}}{k_B T_p}\right) & (V_{sc} < 0) \\ N_{sp} \cdot \pi r_{sp}^2 e n_p \sqrt{\frac{8k_B T_p}{\pi m_e}} \left(1 + \frac{eV_{sc}}{k_B T_p}\right) & (V_{sc} \geq 0) \end{cases} \\ I_{p,sc} &= \begin{cases} N_{sp}^* \cdot \pi r_{sp}^2 \frac{\varepsilon_{sc} J_0}{d^2} & (V_{sc} < 0) \\ N_{sp}^* \cdot \pi r_{sp}^2 \frac{\varepsilon_{sc} J_0}{d^2} \exp\left(-\frac{eV_{sc}}{k_B T_p}\right) & (V_{sc} \geq 0) \end{cases} \end{aligned} \quad (2.132)$$

where k_B is the Boltzmann constant; $v_i \equiv (v_D^2 - 2e\phi/m_i)^{1/2}$ is the velocity of the solar wind ions; T_e and T_p are the temperatures of the solar wind electrons and photoelectrons, respectively; m_i and m_e are the masses of an ion and an electron,

respectively; ϵ_{sc} is the photoemissivity of a spacecraft; $J_0 \equiv 4.5 \times 10^{-6}$ A/m² is the photoemission current density at 1 AU (Nitter et al. 1998); and d is the distance of an asteroid from the Sun expressed in AU. These equations indicate that a positively charged spacecraft attracts more electrons, while a negatively charged spacecraft attracts more ions. The last equation shows that the photoelectric effect is less likely to occur for a positively charged spacecraft. Note that Eq. (2.132) is derived based on the assumption that $v_{T,i} \ll v_D \ll v_{T,e}$.

Since the charge of the spacecraft is affected by the ambient plasma, the spacecraft must emit current constantly in order to maintain the charge level (King et al. 2002). The emitted current I_{sc} is given by the following equation:

$$\begin{aligned} \frac{dQ}{dt} &= I_i - I_e - I_p + I_{p,sc} - I_{sc} = 0 \\ \therefore I_{sc} &= I_i - I_e - I_p + I_{p,sc} \end{aligned} \tag{2.133}$$

Although possible methods for emitting current from a spacecraft are not discussed in this paper, several solutions have been proposed in previous research (Quadrelli et al. 2017a; Schaub et al. 2004; King et al. 2002). Consequently, the required power for electrostatic levitation with a constant charge Q is calculated from Eqs. (2.130)–(2.133) as follows:

$$P = |V_{sc} I_{sc}| \tag{2.134}$$

The flow chart to calculate the required power is described in Fig. 2.73.

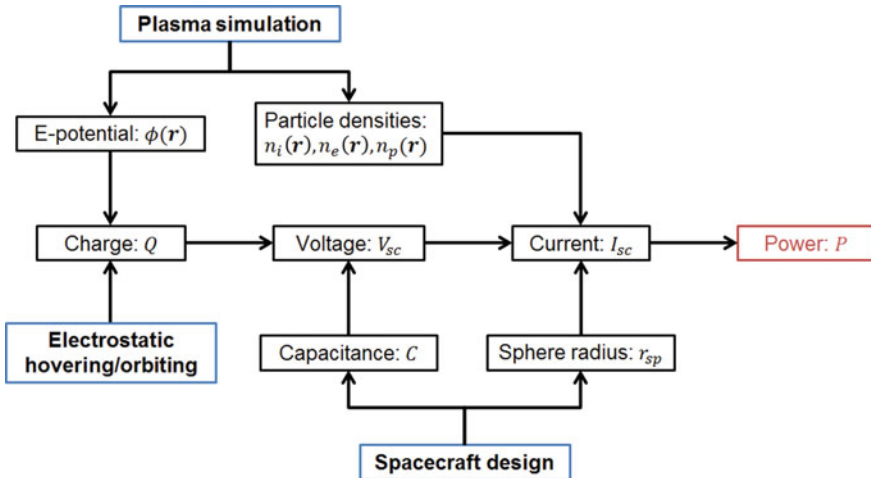


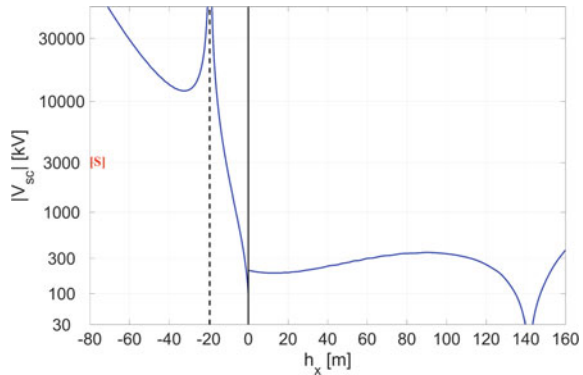
Fig. 2.73 Flow chart of power calculation

2.15.2 Power Required for Electrostatic Hovering

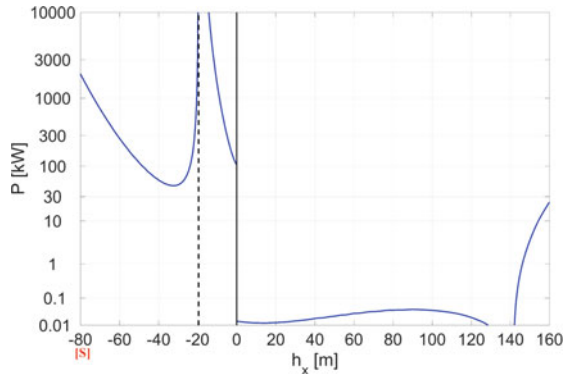
Electrostatic hovering above an asteroid can potentially be achieved by creating artificial equilibrium points with electrostatic force. This section investigates the feasibility of electrostatic hovering from the perspective of power requirement. Analyses are performed for the collinear equilibrium point solutions obtained in the previous subsection.

Figure 2.74a shows the magnitude of voltage required for electrostatic hovering at the corresponding altitude. The power supply voltage is calculated from the required charge based on Eq. (2.130). This result indicates that dayside hovering requires at least 100 kV levels of charge. Although such high-voltage charging itself might not cause any risk to a spacecraft, it can cause electrostatic discharge, which is harmful to spacecraft subsystems. Therefore, the spacecraft must be designed to prevent electrostatic discharge itself or instrument damage due to it, as discussed in previous research (Schaub et al. 2004; King et al. 2002).

Fig. 2.74 Required voltage/power for electrostatic hovering at a collinear equilibrium point



a) Voltage



b) Power

Figure 2.74b illustrates the power required to hover the spacecraft at an equilibrium point. It is indicated that hovering on the dayside near the surface requires the power of as much as 100 kW, while hovering on the nightside requires only about 15 W, despite almost the same charge levels. This difference primarily stems from two reasons. First, the mass of an electron is much smaller than that of an ion, and thus, electrons are much more mobile in a plasma. This results in a large negative current flux, requiring much power to maintain a positive charge. Second, a dense photoelectron layer near the surface is present around the subsolar region, as shown in Fig. 2.11. This environment also involves a large negative current for a positively charged spacecraft. These results imply that electrostatic hovering with a negative charge is more feasible than that with a positive charge from the perspective of power requirement.

2.15.3 Power Required for Electrostatic Orbiting

The power requirement for electrostatic orbiting is analyzed in this subsection. Figure 2.75 shows the power history of the orbit provided in Fig. 2.21a during one orbital period. Even though the charge Q is given as a constant, the required power varies because the current flux from an ambient plasma depends on the position of a spacecraft with respect to an asteroid. The broken line in the figure represents the average power defined by the following equation:

$$P_{ave} = \frac{1}{T} \int_0^T P(t) dt \quad (2.135)$$

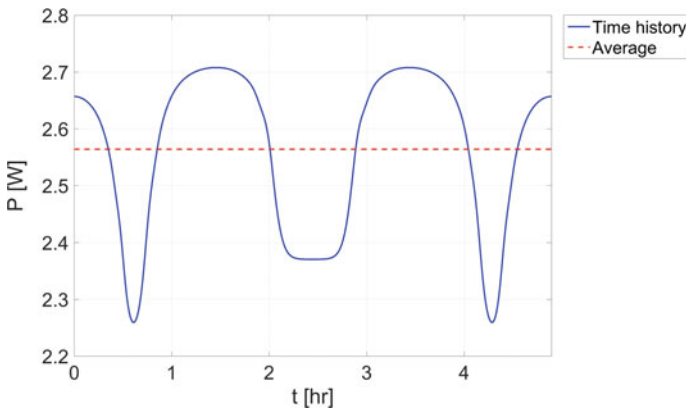
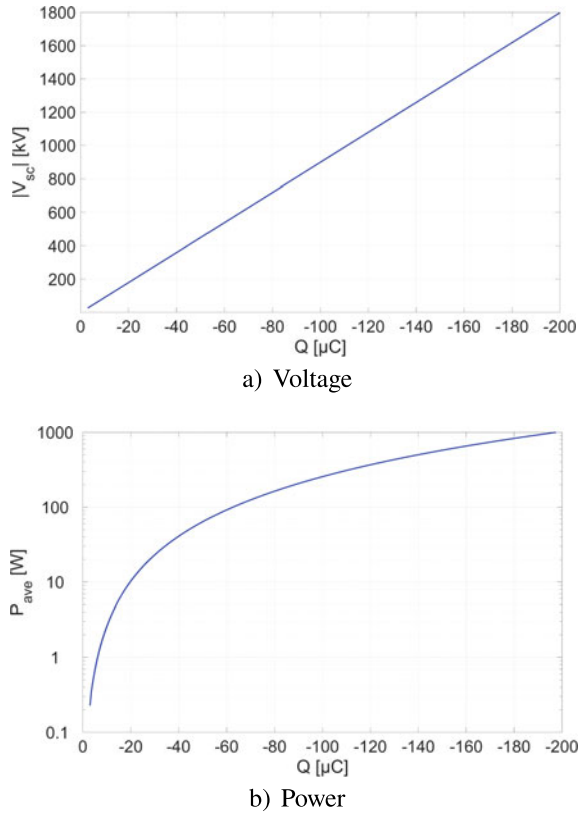


Fig. 2.75 Power consumption during one orbital period

Fig. 2.76 Required voltage/power for electrostatic orbiting



The average power required for this electrostatic periodic orbit is calculated as 2.56 W, and it appears to be feasible for missions.

Figure 2.76 shows the voltage and power required to achieve electrostatic periodic orbit solutions provided in Fig. 2.23. Figure 2.76a is simply obtained from the relationship between the charge and the voltage, and Fig. 2.76b plots the average power of each single periodic orbit. Interestingly, even though the solution space structure of electrostatic periodic orbits is complex and involves bifurcation, the power diagram exhibits a simple profile as shown in Fig. 2.76b. According to this analysis, example values of the required voltage and power are $|V_{sc}| = 89.9$ kV and $P_{ave} = 2.56$ W for the orbit with $Q = -10$ μC ; and $|V_{sc}| = 449$ kV and $P_{ave} = 63.8$ W for the orbit with $Q = -50$ μC .

Comparing Figs. 2.74b and 2.76b, electrostatic orbiting requires considerably lower energy than electrostatic hovering on the dayside. Although an E-Glider must be inserted into an orbit either by itself or a mother spacecraft, after the insertion, it can orbit around an asteroid without requiring any fuel. The required voltage and power largely depend on the design of an E-Glider, and thus, further investigations must be carried out to optimize the entire system design of the E-Glider. In addition,

as mentioned in Sect. 2.15.3, the spacecraft must be designed to have the capability of handling high voltage.

2.15.4 Current Collection for Spherical Electrodes

As already indicated above, the electrodes are assumed to be totally insulated from the spacecraft structure and also from the other electrodes such that the mutual capacitance is equal to zero. This gives a diagonal capacitance matrix for the overall spacecraft given only by the self-capacitance of each electrode.

To correctly evaluate the total power needed to maintain a certain level of charge, computation of the current collected by each electrode of the spacecraft is required. The current collection in a plasma can be analyzed by considering two different regimes: the sheath-area-limited (SAL) regime and the orbit-motion-limited (OML) regime (Bhattarai and Mishra 2017).

The SAL regime can be adopted when the radius of the electrode R_{el} is comparable or higher than the sheath dimension given by the local value of the Debye length λ_D . Hence, if $R_{el}/\lambda_D \geq 1$, the thin-sheath approximation of the SAL model can be used. In the SAL regime, all the particles that enter the Debye sheath are assumed to be captured by the electrode; thus, the current density is mostly limited to the thermal current on the spacecraft surface (Mott-Smith and Langmuir 1926).

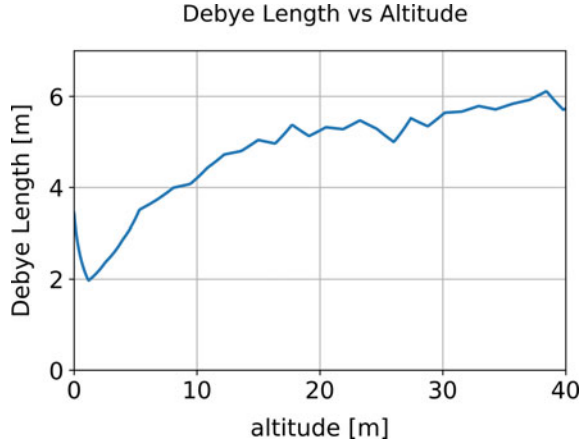
Instead, the OML regime can be assumed when the radius of the electrode is lower than the sheath dimension, such that $R_{el}/\lambda_D \ll 1$. In this case, under the assumption of the OML regime, only a small percentage of the particles that enter the Debye sheath are captured by the electrode (only the particles with a trajectory that approaches the electrode with a minimum distance below a certain threshold), while most of the particles are deflected following a curved trajectory that does not bring them close enough to the electrode to be captured. Thus, in the OML regime, it can be assumed that the current collected is mainly limited by the electrode dimension.

First, the current collected by a spherical electrode in a plasma environment as the one described by the PIC analysis is evaluated. The density of each species at each position is known from the outputs of the PIC analysis. By knowing these densities and the temperature of each species, the total Debye length can be computed at each position as

$$\lambda_D(\mathbf{R}) = \sqrt{\frac{\varepsilon_0 k_B / e^2}{\frac{N_p(\mathbf{R})}{T_p} + \frac{N_e(\mathbf{R})}{T_e} + \frac{N_i(\mathbf{R})}{T_i}}} \quad (2.136)$$

where \mathbf{R} is the position vector in the RIC reference frame, ε_0 is the vacuum permittivity, k_B is the Boltzmann constant, e is the elementary charge, $N_j(\mathbf{R})$ is the density of the species j function of the position of evaluation, and T_j is the temperature of the species j . By assuming a plasma field as the one previously described, the total Debye length can be evaluated by using the PIC results as explained. In Fig. 2.77,

Fig. 2.77 Total Debye length as a function of the altitude



the Debye length as a function of the altitude on the subsolar axis (x-axis in the RIC frame) is reported.

The total Debye length is almost always higher than 2 m. Thus, for the case of a spherical electrode with radius $R_s = 0.046$ m, the OML regime is the most suitable. Under this assumption, the collected ion current I_i for each spherical electrode can be computed as (Quadrelli et al. 2017b; Nitter et al. 1998)

$$I_i = \frac{1}{4} A_{sph} e N_i v_i \left(1 - \frac{2eV_{el}}{m_i v_i^2} \right) \quad \text{if } V_{el} < \frac{m_i v_i^2}{2e} \quad (2.137)$$

$$I_i = 0 \quad \text{if } V_{el} > \frac{m_i v_i^2}{2e}$$

where A_{sph} is the external surface of the sphere considered as an electrode, e is the elementary charge, $v_i = (v_D^2 - 2e\phi/m_i)^{0.5}$ is the solar wind ion velocity computed by knowing the potential ϕ of the plasma and the drift velocity, V_{el} is the bias electrostatic potential of the spherical electrode, and m_i is the proton mass.

If the electrode has a negative potential ($V_{el} < 0$), the collected electron and photoelectron currents are computed as (Bhattarai and Mishra 2017)

$$I_e = \frac{1}{4} A_{sph} e N_e \sqrt{\frac{8k_B T_e}{\pi m_e}} \exp\left(\frac{eV_{el}}{k_B T_e}\right) \quad (2.138)$$

$$I_p = \frac{1}{4} A_{sph} e N_p \sqrt{\frac{8k_B T_p}{\pi m_e}} \exp\left(\frac{eV_{el}}{k_B T_p}\right)$$

Both the electron and the photoelectron can be expressed by using the very same formulation. Note that in Eq. 2.138, the electrode potential V_{el} is negative; hence, the entire exponential term is negative, leading to an exponential decrement of the

collected current if the potential is decreased. This can be expected since the more negative the potential, the better the approximation of an ion-saturated sheath (Bhat-tarai and Mishra 2017).

If the electrode has a positive potential ($V_{el} > 0$), the electron and photoelectron current collection is enhanced and can be computed as (Laframboise and Parker 1973)

$$\begin{aligned} I_e &= \frac{1}{4} A_{sph} e N_e \sqrt{\frac{8k_B T_e}{\pi m_e}} \left(1 + \frac{eV_{el}}{k_B T_e} \right) \\ I_p &= \frac{1}{4} A_{sph} e N_p \sqrt{\frac{8k_B T_p}{\pi m_e}} \left(1 + \frac{eV_{el}}{k_B T_p} \right) \end{aligned} \quad (2.139)$$

By using these equations, the power needed by the spacecraft here considered can be evaluated to assess the feasibility of hovering from the energetic point of view. The total power needed is computed by evaluating both the potential V_{el} and the current collected I_{el} by each electrode of the spacecraft. Once that these values are available, the total power needed is computed as the sum of the power required by each electrode at each instant of time. Knowledge of the exact position in the RIC frame of each electrode leads to a more accurate evaluation of the collected current since the exact density of each species on that location is known from the PIC results. The scenario used for the simulation here presented is exactly the one previously analyzed in Sect. 2.14.6; thus, the length of the rigid tethers is 10 m. The results in terms of potential for each electrode (upper plot) and total power needed onboard (lower plot) for the repositioning and pointing maneuver analyzed in depth in Sect. 2.14.5 are shown in Fig. 2.78.

Note that the legend of this figure says which color is associated with the electrode located on a certain semi-axis in the BF reference frame. The symmetry of the potential of the electrodes located on two opposite sides with respect to the center of mass of the spacecraft is given by the fact that they are all equal from the geometric point of view and because the two electrodes that form a dipole have the same level of net charge Q and opposite differential charge dq . The potential bias of the electrodes is extremely high, reaching values even higher than 10000 KV in modulus during the initial transient phase. During the final position-keeping phase, the potentials drop to lower values in modulus, being about -1400 kW for the electrodes along the x-axis in BF and 600 KV and about -3400 KV for the electrodes on the +y and -y semi-axes, respectively. These levels are clearly unfeasible, especially for a small CubeSat like the one simulated here. Such a high potential is given by an extremely low capacitance due to the extremely reduced size of the spheres. By increasing the size of the spheres, the enhancement of the self-capacitance could be possible, but the hypothesis of the OML theory could be not valid for the needed increment in the size of the sphere; moreover, a higher surface area implies a higher current collection that can affect the total power needed. The total power required during the initial transient phase shows peaks of about 800 W, while during the final position-keeping phase, the power required is about 75 W. These values are extremely high for the

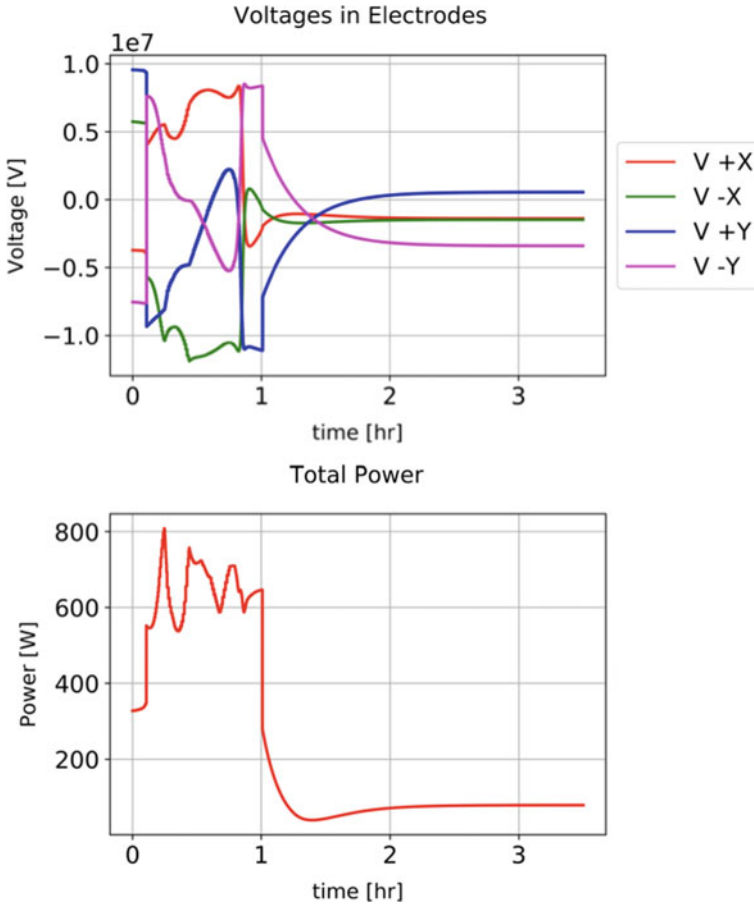


Fig. 2.78 Electrode potentials (upper) and total power needed (lower) for a four-sphere spacecraft model

state-of-the-art technologies available for the power production in a small CubeSat (NASA 2020). By considering both the total power level and the potential reached by the electrodes, we can state that neither the maneuvering phase nor the final hovering phase are feasible with the spherical electrodes assumed here.

In Corradino (2018), an analysis that investigates the behavior of different electrode shapes in plasma is reported, and from this analysis the wire electrodes, in particular, if they are shaped like a loop, appear to be capable of reducing both the power consumption and the potential with respect to the spherical electrodes here presented. A detailed analysis of the simulations run by using the wire electrodes is presented in Sect. 2.15.5.

2.15.5 Current Collection for Wire Electrodes

From the analysis of several electrode shapes reported in Corradino (2018), good improvements from the power consumption point of view can be obtained by using a wire electrode. In particular, the high efficiency of the hoops in achieving an extremely low power consumption with a lower increment in the voltage with respect to a straight wire has been highlighted. This effect is given by the fact that hoop electrodes can be modeled exactly as wire electrodes but with a total length equal to the loop length. For this reason, in this section, the benefits that have been theoretically evaluated for a simpler static case in Corradino (2018) are simulated for the previously analyzed scenario in order to assess the feasibility of the maneuver itself by using a different electrode model (the hoops in this case).

The assumption of OML regime is valid also for type of electrodes, in particular for the case in which the radius of the wire is small (as in this case in which a radius of 1 mm has been assumed). The total current collection for a wire electrode is here modeled by using the empirical formulas derived in Choiniere et al. (2003) and Fuhrhop (2007) used also in Corradino (2018). For the case of negative hoop potential we can write that

$$I_{el} = \frac{1}{4} A_l e N_i \sqrt{\frac{8k_B T_i}{\pi m_i}} \frac{2}{\sqrt{\pi}} \sqrt{T_i + F - V_l} \quad (2.140)$$

where A_l is the surface area of the loop, F is the energy of the ion beam computed as $F = \frac{1}{2} m_i v_D^2$, and V_l is the potential of the wire. In the case of positive potential, the collected current can be computed as

$$I_{el} = \frac{1}{4} A_l \left(e N_e \sqrt{\frac{8k_B T_e}{\pi m_e}} + e N_p \sqrt{\frac{8k_B T_p}{\pi m_e}} \right) \frac{2}{\sqrt{\pi}} \sqrt{1 + X} \quad (2.141)$$

where X is the nondimensional potential defined in Choiniere et al. (2003) as $X = V/T_e$.

To perform simulations also for this case, a new spacecraft model is needed. To verify the possibility of using a wire-hoop electrode, a 1U CubeSat (CubeSat 2019) has been selected. Hence, the spacecraft's central body is a cube of $10 \times 10 \times 11$ cm of mass 1.33 Kg with optical properties equal to the previously analyzed cases. The hoops are assumed to be tangent to the rectangular side faces of the CubeSat. Each hoop is modeled as a circumference with the center of mass at 2.5 m from the center of mass of the central body, such that each hoop in this case has a radius of 2.445 m (see Fig. 2.79). As previously declared, the thickness of the wire that composes the hoop is assumed to be 1 mm. By using this new configuration, the center of mass of two opposite hoops is located at exactly 5 m of distance.

By assuming that the total charge of each hoop is concentrated in the center of mass of the hoops, the same model developed in Sect. 2.14.5 can be used to simulate

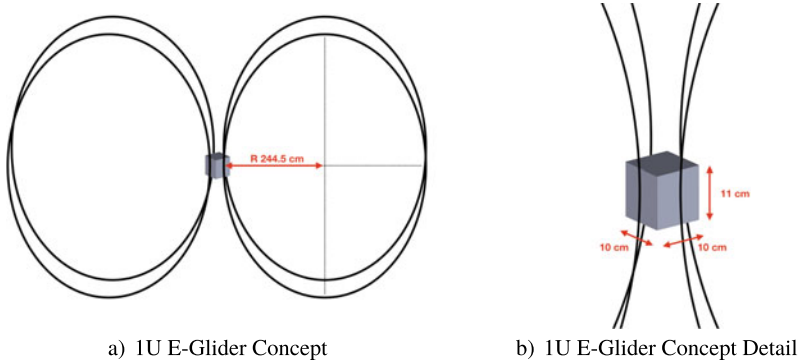


Fig. 2.79 1U E-Glider concept with wire electrodes (not to scale)

the spacecraft dynamics in 2D. For the case here presented, the distance between the two charges of a dipole is equal to 5 m.

Since the total mass of the CubeSat is halved with respect to the previous cases, the saturation level for both the net and differential charge can be reduced to $\pm 50 \mu\text{C}$. For the maneuver here analyzed, the standard distance of 10 m between the charges cannot be used because the further portion of the “most inner hoop” may collide with the asteroid surface.

The results of the potential and power analysis are reported in Fig. 2.80. The upper plot shows the potential level of each electrode, while the second plot shows the total power needed to perform the maneuver and to keep the final condition.

Also in this case, the symmetry of the potential level assumed by two opposite electrodes can be noticed. During the initial transient phase, the potential of the electrodes is extremely high for all of them, reaching levels of about 500 KV and also showing strong oscillations from strongly positive to strongly negative values as a function of the polarity of the charges. During the position-keeping phase the levels of charge needed are lower, and as a result, the potential of the electrodes settles down to about 50 KV in modulus for the electrodes on the x-axis in the body frame, -150 KV for the electrode on the -y-axis in the BF frame, and about 100 KV for the electrode on the +y-axis in the BF frame.

The values obtained for the position-keeping phase are in agreement with the analysis of this type of electrodes developed in Corradino (2018), where the analysis was carried out using the convergence method previously introduced that led to the computation of the “true” capacitance C and the “true” potential V_{el} of an electrode in plasma. By using the results of the analysis in Corradino (2018), for a wire electrode with a radius of the wire of the order of 10^{-3} m and length of the order of 10^1 m (as the case here analyzed), a potential of the order of about $10^5 - 10^6$ volts can be expected for a spacecraft of comparable mass. Thus, by computing the potential of the electrodes as Q/C , a slightly overestimated value for the actual potential is obtained but inside the range provided in Corradino (2018). This makes the analysis carried

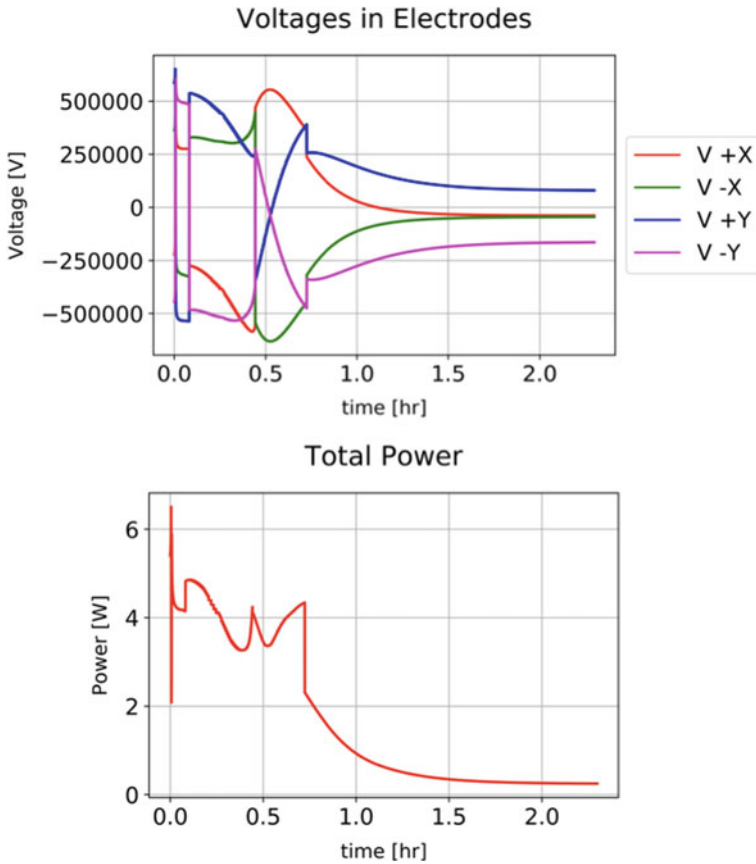


Fig. 2.80 Electrode potentials (upper) and total power needed (lower) for a four-Hoop 1U CubeSat

out in this work suitable at least for the preliminary evaluation of the feasibility of hovering on the sunlit side of an airless body.

The extremely high level of the potential is mainly due by the fact that the capacitance of a wire electrode is extremely low, of the order of 10^{-11} Farad. By considering the total power needed, we can immediately notice that the amount of power to be provided is low even during the initial transient phase. During the initial phase of the simulation, the power needed has a peak of about 6 W due to a peak on the potentials of all the electrodes. During the position keeping, the total power needed is extremely low, settling down to about 0.3 W. Also in this case, the result of the simulation is in agreement with the results of the parametric analysis on the total power required by a wire electrode reported in Corradino (2018) since a total power of the order of 10^{-1} W was expected.

In this case, by looking at the power level required, the maneuver and, in general, the hovering can be stated to be feasible with a state-of-the-art solution for onboard

power generation. The limiting factor here is confirmed to be the extremely high level of the potential, as already highlighted by previous analysis (Quadrelli et al. 2017b; Corradino 2018). The potential can be reduced by substantially increasing the length of the electrode. The increment in the total length can be achieved by using multiple hoops, but this can make the control law extremely complex due to the presence of a higher number of dipoles and due to the fact that the interferences among the electrodes (that are neglected here) can become relevant due to the short distance among two consecutive electrodes.

To assess the effects of a higher mass on the required power and on the electrode potential, we analyzed the case of a 12U CubeSat with four hoop electrodes.

2.15.6 Example of a 12U CubeSat with Four Hoop Electrodes

In this section, the results obtained from the analysis of a 12U CubeSat are reported. The central box-shaped body used for this simulation has the same dimensions of the 1HOPSat satellite (1HOPSat 2020) ($22.6 \times 22.6 \times 34$ cm) and a total mass of 22 Kg. Once again, the hoops are located on the side rectangular faces with the center of mass at a distance of 2.5 m from the center of mass of the central 12U structure (thus a radius of the hoop equal to 2.387 m; see Fig. 2.81). Also in this case, the thickness of the wire is set to be equal to 1 mm. The initial and the final conditions for the maneuver analyzed are the same as the previous case with a 1U CubeSat, and the optical properties of the external faces are unchanged. Due to the higher mass, the saturation level used in this case is $\pm 750 \mu\text{C}$. The increment of the mass with respect to the 1U CubeSat previously analyzed affects only the net charge, resulting in a higher Q_0 and in a higher dQ needed to achieve the maneuver. The higher level of $Q = Q_0 + dQ$ makes the maneuver unfeasible for the low saturation limit since $|Q_0| > 50 \mu\text{C}$.

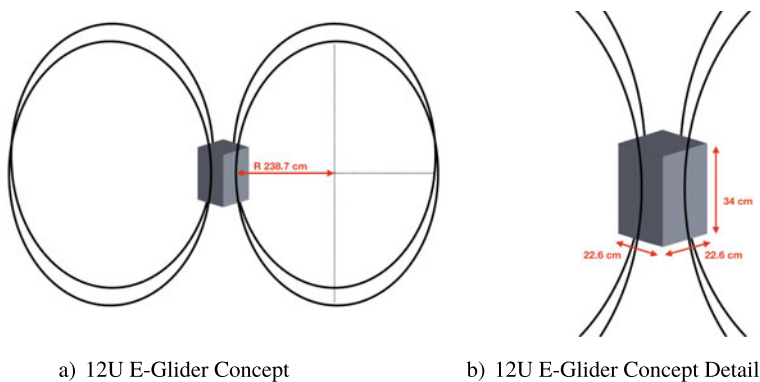


Fig. 2.81 12U E-Glider concept with wire electrodes (not to scale)

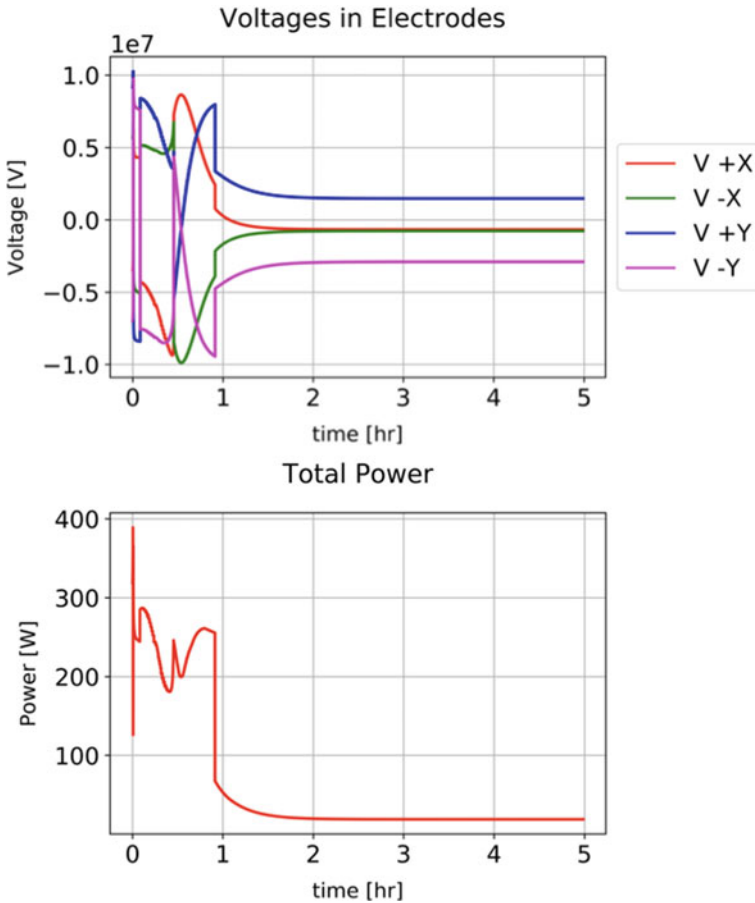


Fig. 2.82 Electrode potentials (upper) and total power needed (lower) for a four-Hoop 12U CubeSat

The selection of a new saturation level of $\pm 750 \mu\text{C}$ leads to an increment in the saturation level itself with respect to the previous case which is almost proportional to the increment in the spacecraft mass. This makes the dynamics of the two simulations almost equal (the low differences detected are due to the different surface areas). This makes it possible to evaluate the changes in the total power required and in the electrode potentials due to a higher mass, and thus a higher level of charge needed. From the analysis already developed for a single static electrode in Corradino (2018), a nonlinear increment in the total power needed with respect to the increment in the mass of the spacecraft can be expected.

The resulting potential of the electrodes (upper) and the resulting total power (lower) are reported in Fig. 2.82. The results are proportional to the ones reported in Fig. 2.80.

During the transient phase, the potentials are extremely high, reaching peaks of about 10000 KV, which are practically unfeasible with the technology available nowadays for space applications. Also during the final position keeping, the potentials are high, reaching a level higher than 650 KV in modulus for each electrode (-650 KV is the potential of the electrodes aligned with the x-axis that results to be the ones with lower potential in modulus). The potential reaches scales linearly with the increment of mass (the small variations here are due to a different size of the hoops).

On the other side, the total power required is below 400 W during the entire transient phase. The total power settles down to a level of about 20 W and has an increment of about 60 times the power computed for the 1U CubeSat previously analyzed. This value is in agreement with the sensitivity analysis reported in Corradino (2018) in which a superlinear increment (power of $\sqrt{2}$) with the spacecraft mass has been identified. The total power levels obtained from this analysis are feasible for a 12U CubeSat spacecraft (e.g., the 6U solar panels produced by *Innovative Solutions In Space* can be used as a power source onboard).

The maneuver here analyzed is clearly feasible for a 12U E-Glider from a dynamical and power point of view, but it results to be unfeasible for the extremely high level of potential on each electrode. This analysis confirmed that the limiting factor for the development of an E-Glider mission is the potential level of the electrodes and not the power required to enable the electrostatic flight. Moreover, the total power needed and also the potential reached during the transient phase are much higher than the values computed for the final position keeping. This means that the feasibility of a given maneuver must also be carefully evaluated from the energetic viewpoint. The enhancement of the power needed during the maneuvering phase is due to the fact that the charge levels of the electrodes are substantially increased during this phase with respect to the ones needed during the position keeping. Also in this case, the increment obtained is not linear, in agreement with the analysis of the power consumption sensitivity with respect to the Q/M ratio developed in Corradino (2018), in which, again, a superlinear increment (power of $\sqrt{2}$) is identified.

The potential levels reached here for all the cases analyzed make mandatory a new evaluation of the effects of a spacecraft with such a high potential on the environment in which it operates. For example, if the plasma sheath of the spacecraft (enhanced by the extremely high potential) is wide enough to reach the surface of the asteroid, it may affect the photoelectron sheath of the main body itself, which can cause change of the flow of the particles and a modification in the solar wind effects, leading to strong variations in the previously computed conditions of hovering. Corradino (2018) reported a numerical simulation (provided by William Yu and Dr. Wang from USC) run by using the PIC method in which it is proven that a spacecraft with a potential of 1 KV, for the plasma here analyzed, has a sheath of a few Debye lengths, thus not high enough to strongly affect the results previously obtained. The same type of evaluation should also be carried out for higher potentials in order to try to define a limit potential that should not be exceeded. The extremely high potential may also prevent this type of technology to be applied to a planetary environment (thus in presence of an atmosphere) due to the atmospheric breakdown threshold that

should not be exceeded in order to avoid strong discharges (glow discharges), peak in the power needed for the levitation, and also possible damages to the electrode itself.

From the material viewpoint, the composition of all the electrodes here analyzed is a challenge; in particular, the sputtering and the erosion of the electrodes that can strongly limit the lifetime of an E-Glider mission must be evaluated. Finally, we showed that the current collection can be enhanced by a flowing plasma (Choiniere et al. 2003); thus, more exact numerical simulations with the PIC method should be useful to evaluate this increment for the cases here presented and eventually find a correction factor for the basic OML formulation assumed here.

2.16 Conclusions

This paper has proposed a novel flight mechanism around airless bodies in the solar system, utilizing the electrostatic field around them. The two distinct types of operations have been presented, namely, electrostatic hovering and electrostatic orbiting. We showed that both of these methods allow dayside operation without requiring any fuel. Therefore, the electrostatic flight around an asteroid offers significant advantages against conventional methods based on natural dynamics, from the perspective of mass budget, optical observation, solar power generation, and thermal design. By inducing the electrostatic force, several different artificial equilibrium points around an asteroid can be created, including the dayside equilibrium that cannot be observed in the natural dynamics. Electrostatic hovering can be achieved by placing a spacecraft at these artificial equilibrium points, and it could potentially be an option for effective proximity operation around asteroids. However, the power analyses showed that electrostatic hovering on the dayside requires high levels of power, based on our current model. For this reason, the electrostatic orbiting method was also discussed as an alternative strategy for an E-Glider operation. A new class of periodic orbits, called electrostatic periodic orbits, was successfully designed. These orbits exist on the dayside of an asteroid, requiring only a few watts of power for some cases. Moreover, the specific impulse of the E-Glider was found to be significantly higher than conventional propulsion systems. From these observations, the electrostatic orbiting strategy appears to be promising for asteroid missions.

Finally, the effects of an irregularly shaped asteroid, including the irregular gravitational field and the irregular electrostatic field, have been formulated and evaluated. As a result, an electrostatic periodic orbit solution can serve as a good approximation for an asteroid with a relatively small oblateness. Although an asteroid with a highly irregular shape perturbs the orbital motion significantly, possible approaches to this problem were also proposed, including feedback control of the spacecraft charge.

While much work remains to be carried out, including gaining additional insight into the engineering behavior, developing approaches for path planning and navigation, and conceiving plans to build and test a prototype, we conclude that the

electrostatic flight method using an E-Glider is useful for airless body missions and exhibits intriguing and valuable dynamic characteristics.

Acknowledgements ©2020. All rights reserved. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, under the NASA Innovative Advanced Concepts Program. The authors are very grateful to Dr. Henry Garrett of JPL and many students who have worked at JPL with Dr. Quadrelli on the E-Glider, including Mr. Filippo Corradino of the Politecnico of Torino, Italy.

Nomenclature

d	Main body to Sun distance, m
h	Angular momentum, m^2/s
a	Semimajor axis, m
e	Eccentricity, -
P	Semilatus rectum, m
θ	True anomaly, deg
μ	Gravitational parameter, m^3/s^2
M	Mass, Kg
f	Force, N
N	Orbital angular velocity in a circular orbit, rad/s
ω	Angular velocity, rad/s
\mathbf{J}	Mass inertia moment matrix, $kg\ m^2$
\mathbf{T}	Torque vector, Nm
\mathbf{a}	Acceleration vector, m/s^2
\mathbf{v}	Velocity vector, m/s
\mathbf{r}	Position vector, m
\mathbf{G}	Gradient tensor of a vector field
C	Solar radiation pressure coefficient
\mathbf{E}	Electric field vector, V
q	Single charge, C
Q	Total net charge, C
\mathbf{S}_q	First moment of the charge vector, Cm
\mathbf{I}_q	Second moment of the charge tensor, Cm^2
<i>Subscript</i>	
a	Asteroid
s	Sun
SC	Spacecraft
g	Gravitational
p	Solar radiation pressure
pa	Absorption
ps	Specular reflection
pd	Diffuse reflection

- e* Electrostatic
- x, y, z* General directions of a component in a vector/tensor
- Superscript* (for rotation matrices, the super- and subscripts refer to this list)
- r* Radial/in-track/cross-track (RIC) frame
- a* Asteroid-centered inertial (ACI) frame
- b* Body-fixed (BF) frame

References

- IHOPSat Formal Orbital Debris Assessment Report (ODAR) and End of Mission Plan (EOMP)*. <https://apps.fcc.gov/els/GetAtt.html?id=218010&x=>. Accessed: 07-01-2020.
- Aplin, K.L., Bowles, N., Urbak, E., and Sawyer, E.C. 2011. Asteroid electrostatic instrumentation and modelling. *Journal of Physics Conference Series*.
- Aslanov, Vladimir, and Hanspeter Schaub. 2019. Detumbling attitude control analysis considering an electrostatic pusher configuration. *Journal of Guidance, Control, and Dynamics* 42 (4): 900–909.
- Bechini, M. 2020. *E-glider. Modeling and simulation of an electrically actuated spacecraft in a PIC-described plasma field*, Politecnico di Milano, Tesi di Laurea Magistrale. <https://www.politesi.polimi.it/handle/10589/153193>.
- Bechini, Michele, Quadrelli, Marco B., Lavagna, Michèle, and Wang, Joseph J. 2021. Hovering of an electrically actuated spacecraft in a small-body plasma field. *Journal of Spacecraft and Rockets* 58 (5): 1461–1476. <https://doi.org/10.2514/1.A34954>.
- Beletsky, V.V., and Lavin, E.M. 1993. *Dynamics of space tether systems*. American Astronautical Society Publication. (Advances in the Astronautical Sciences). ISBN 0877033706.
- Bellerose, Julie, and Yano, Hajime. 2010. Dynamics of asteroid 1999 JU3: target of the Hayabusa follow-on mission. *Transactions of JSASS Aerospace Technology Japan* 8 (ists27): Tk_23–Tk_28.
- Besse, Arthur L., and Allen G. Rubin. 1980. A simple analysis of spacecraft charging involving blocked photoelectron currents. *Journal of Geophysical Research: Space Physics* 85 (A5): 2324–2328.
- Bhattarai, Shankar, and Lekha Mishra. 2017. Theoretical study of spherical Langmuir Probe in Maxwellian plasma. *International Journal of Physics* 5 (08): 73–81.
- Birdsall, C.K., and Langdon, A.B.: *Plasma physics via computer simulations*. Institute of Physics Publishing. (The Adam Hilger Series on Plasma Physics). – ISBN 9780750301176.
- Broschart, S.B., and D.J. Scheeres. 2005. Control of hovering spacecraft near small bodies: application to asteroid 25143 Itokawa. *Journal of Guidance, Control, and Dynamics* 28 (2): 343–354.
- Broschart, Stephen B., Daniel J. Scheeres, and Benjamin F. Villac. 2009. New families of multi-revolution terminator orbits near small bodies. *Advances in the Astronautical Sciences* 135 (3): 1685–1702.
- Chen, F.C. 2003. *Langmuir probes diagnostics*. Electrical Engineering Department, University of California, Los Angeles, 2003 (Minicourse on Plasma Diagnostic).
- Choiniere, Eric, Gilchrist, Brian, Bilén, Sven, and Fuhrhop, Keith. 2003. *Measurement of cross-section geometry effects on electron collection to long probes in mesosonic flowing plasmas*. In *39th AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit*. <https://arc.aiaa.org/doi/abs/10.2514/6.2003-4950>.
- Clark, Beth E., Bruce Hapke, Carlé Pieters, and Daniel Britt. 2002. Asteroid space weathering and regolith evolution. *Asteroids III* 585: 90086–2.
- Colwell, Joshua E., Amanda A. Gulbis, Mihály Horányi, and Scott Robertson. 2005. Dust transport in photoelectron layers and the formation of dust ponds on Eros. *Icarus* 175 (1): 159–169.

- Corradino, F. 2018. *Modeling of orbital and attitude dynamics of a satellite controlled via active electrostatic charging*. Politecnico di Torino, Tesi di Laurea Magistrale. <https://webthesis.biblio.polito.it/6853/1/tesi.pdf>.
- CubeSats Overview. https://www.nasa.gov/mission_pages/cubesats/overview. Accessed 03-10-2019.
- Council, National R. 2011. *Vision and voyages for planetary science in the decade 2013–2022*. Washington, DC : The National Academies Press. <https://www.nap.edu/catalog/13117/vision-and-voyages-for-planetary-science-in-the-decade-2013-2022>. ISBN 978-0-309-22464-2.
- Cui, C, and Wang, J. 2019. Numerical simulations of plasma-spacecraft interactions near irregularly shaped small asteroids. In *Applied Space Environments Conference* Los Angeles, CA.
- Curtis, H.D. 2010. *Orbital mechanics for engineering students*, 3rd edn. Elsevier. – ISBN 9780080977478.
- Czeplia, Steven A., Hugh McManus, and Daniel Hastings. 2000. Charging of composites in the space environment. *Journal of Spacecraft and Rockets* 37 (5): 556–560.
- Eckman, R.A., Brown, A.J., and Adamo, D.R. 2011. Normalization of gravitational acceleration models. In *NASA scientific and technical information*. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110023121.pdf>.
- Evlanov, E.N., M.A. Zavjalov, and P.M. Tyuryukanov. 2013. Electron guns for spacecraft. *Cosmic Research* 51 (5): 388–395.
- Fantino, E., and S. Casotto. 2009. Methods of harmonic synthesis for global geopotential models and their first-, second- and third-order gradients. *Journal of Geodesy* 83 (7): 595–619.
- Finckenor, M.M. 1999. Multilayer insulation material guidelines. In *National aeronautics and space administration*. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19990047691.pdf>.
- Fuhrhop, Keith. 2007. *Theory and experimental evaluation of electrodynamic tether systems and related technologies*, The University of Michigan, Dissertation. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/57663/kfuhrhop_1.pdf?sequence=2&isAllowed=y.
- Fujiwara, Akira, J. Kawaguchi, D.K. Yeomans, M. Abe, T. Mukai, T. Okada, J. Saito, H. Yano, M. Yoshikawa, D.J. Scheeres, et al. 2006. The rubble-pile asteroid Itokawa as observed by Hayabusa. *Science* 312 (5778): 1330–1334.
- Gorham, P. 2013. Ballooning spiders: the case for electrostatic flight. *eprint arXiv:1309473*.
- Gottlieb, R.G. 1993. Fast gravity, gravity partials, normalized gravity, gravity gradient torque and magnetic field: derivation, code and data. *Nasa Contractor Report 188243*. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940025085.pdf>.
- Han, D. 2015. *Particle-in-cell simulations of plasma interaction with asteroidal and lunar surfaces*, University of South California, Dissertation. <http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll3/id/639389>.
- Han, D., J. Wang, and X. He. 2016a. A nonhomogeneous immersed-finite-element particle-in-cell method for modeling dielectric surface charging in plasmas. *IEEE Transactions on Plasma Science* 44 (8): 1326–1332.
- Han, Daoru, Wang, Pu, He, Xiaoming, Lin, Tao, and Wang, J. 2016b. A 3D immersed finite element method with non-homogeneous interface flux jump for applications in particle-in-cell simulations of plasma-lunar surface interactions. *Journal of Computational Physics* 321 : 965 – 980. ISSN 0021-9991.
- Han, Daoru, and Joseph Wang. 2019. 3-D fully kinetic particle-in-cell simulations of small asteroid charging in the solar wind. *IEEE Transactions on Plasma Science* 47 (8): 3682–3688.
- Han, Daoru, Joseph J. Wang, and Xiaoming He. 2018. Immersed finite element particle-in-cell simulations of plasma charging at the lunar terminator. *Journal of Spacecraft and Rockets* 55 (6): 1490–1497.
- Hartzell, C.M. 2012. *The dynamics of near-surface dust on airless bodies*, University of Colorado Boulder, Dissertation, 2012. https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/6w924c073.

- Havnes, O., C.K. Goertz, G.E. Morfill, E. Grün, and W. Ip. 1987. Dust charges, cloud potential, and instabilities in a dust cloud embedded in a plasma. *Journal of Geophysical Research: Space Physics* 92 (A3): 2281–2287.
- Hénon, Michel. 1969. Numerical exploration of the restricted problem, V. Hill's case: periodic orbits and their stability. *Astronomy and Astrophysics* 1: 223–238.
- Hirata, Naoyuki, and Hideaki Miyamoto. 2012. Dust levitation as a major resurfacing process on the surface of a saturnian icy satellite. *Atlas. Icarus* 220 (1): 106–113.
- Hughes, P.C. 1986. *Spacecraft attitude dynamics*, 0471818429. ISBN: Wiley Inc.
- Ip, W.H. 1986. Electrostatic charging and dust transport at Mercury surface. *Geophysical Research Letters* 13: 1133–1136.
- Iwata, Minoru, Arifur R. Khan, Hideyuki Igawa, Kazuhiro Toyoda, Mengu Cho, and Tatsuhiro Fujita. 2012. Development of electron-emitting film for spacecraft charging mitigation. *Journal of Spacecraft and Rockets* 49 (3): 546–552.
- Jeong, H. 2008. *Kinetic simulations of spacecraft charging and plasma interactions in the solar wind*, Virginia Polytechnic Institute and State University, Dissertation. https://vtechworks.lib.vt.edu/bitstream/handle/10919/30237/Dissertation_Jeong.pdf?sequence=1&isAllowed=y.
- Kafafy, R., T. Lin, Y. Lin, and J. Wang. 2005. Three-dimensional immersed finite element methods for electric field simulation in composite materials. *International Journal for Numerical Methods in Engineering* 64 (7): 940–972.
- Kafafy, R., and J. Wang. 2006. A hybrid grid immersed finite element particle-in-cell algorithm for modeling spacecraft-plasma interactions. *IEEE Transactions on Plasma Science* 34 (5): 2114–2124.
- Khan, Arifur R., Minoru Iwata, Kazuhiro Toyoda, Mengu Cho, Setuo Tomonari, and Yuta Takaki. 2013. In-orbit demonstration of newly developed passive electron-emitting film for spacecraft-charging mitigation. *Journal of Spacecraft and Rockets* 50 (4): 853–859.
- Kikuchi, S. 2017. *E-glider: active electrostatic flight for airless body exploration*. Research Report: The University of Tokyo.
- King, B.L., Parker, G.G., Deshmukh, S., and Chong, J.H. 2002. NIAC phase i final report - spacecraft formation-flying using inter-vehicle coulomb forces. In *NASA Innovative Advanced Concepts (NIAC)* (2002). – http://www.niac.usra.edu/files/studies/final_report/601King.pdf.
- Kobrick, Ryan, Hoffman, Jeffrey, Street, Kenneth, and Rickman, Douglas. 2014. Overview of instruments for investigating dust interactions on small solar system bodies by landers and rovers, 09.
- Kominato, Takashi, Masatoshi Matsuoka, Masashi Uo, Tatsuaki Hashimoto, and Jun'ichiro Kawaguchi. 2006. Optical hybrid navigation and station keeping around Itokawa. In *AIAA/AAS Astrodynamics Specialist Conference and Exhibit 2006*: 6535.
- Kryszczyńska, A., La. Spina, Paolicchi Alessandra, Harris Paolo, and AW, Breiter, S, and Pravec, P. 2007. New findings on asteroid spin-vector distributions. *Icarus* 192 (1): 223–237.
- Laframboise, J.G., and L.W. Parker. 1973. Probe design for orbit? limited current collection. *The Physics of Fluids* 16 (5): 629–636.
- Lai, S.T. 1989. An overview of electron and ion beam effects in charging and discharging to spacecraft. *IEEE Transactions on Nuclear Science* 36 (6): 2027–2032.
- Lee, P. 1996. Dust levitation on asteroids. *Icarus* 124 (1): 181–194.
- Lundberg, B.J. 1988. Recursion formulas of legendre functions for use with nonsingular geopotential models. *Journal of Guidance, Control, and Dynamics* 11 (1): 31–38.
- Lyon, H.R. 2004. *Geosynchronous orbit determination using space surveillance network observations and improved radiative force modeling*, Massachusetts Institute of Technology, Master Degree Thesis. <https://dspace.mit.edu/handle/1721.1/17779>.
- Masek, T.D., and H.A. Cohen. 1978. Satellite positive-ion-beam system. *Journal of Spacecraft and Rockets* 15 (1): 27–33.
- Mendis, D.A., Jay R. Hill, Harry L. Houpis, and E.C. Whipple. 1981. On the electrostatic charging of the cometary nucleus. *The Astrophysical Journal* 249: 787–797.

- Miyamoto, H., et al. 2007. Regolith migration and sorting on asteroid Itokawa. *Science* 316 (5827): 1011–1014.
- Mizera, P.F. 1983. A summary of spacecraft charging results. *Journal of Spacecraft and Rockets* 20 (5): 438–443.
- Morley, Erica L., and Daniel Robert. 2018. Electric fields elicit ballooning in spiders. *Current Biology* 28 (14): 2324–2330.
- Mott-Smith, H.M., and Irving Langmuir. 1926. The theory of collectors in gaseous discharges. *Physical Review* 28: 727–763.
- NASA: *State of the art of small spacecraft technologies - NASA*. <https://sst-soa.arc.nasa.gov/03-power>. Accessed: 07-01-2020.
- Nesnas, Issa A., Matthews, Jaret B., Abad-Manterola, Pablo, Burdick, Joel W., Edlund, Jeffrey A., Morrison, Jack C., Peters, Robert D., Tanner, Melissa M., Miyake, Robert N., Solish, Benjamin S., and Anderson, Robert C. 2012. Axel and DuAxel rovers for the sustainable exploration of extreme terrains. *Journal of Field Robotics* 29 (4): 663–685. <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21407>.
- Nitter, T., O. Havnes, and F. Melandsø. 1998. Levitation and dynamics of charged dust in the photoelectron sheath above surfaces in space. *Journal of Geophysical Research: Space Physics* 103 (A4): 6605–6620.
- Nolan, Michael C., Christopher Magri, Ellen S. Howell, Lance A. Benner, Jon D. Giorgini, Carl W. Hergenrother, R.S. Hudson, Dante S. Lauretta, Jean-Luc. Margot, Steven J. Ostro, et al. 2013. Shape model and surface properties of the OSIRIS-REX target Asteroid (101955) Bennu from radar and lightcurve observations. *Icarus* 226 (1): 629–640.
- Olhoef, G.R., and D.W. Strangway. 1975. Dielectric properties of the first 100 meters of the Moon. *Earth and Planetary Science Letters* 24 (3): 394–404.
- Peck, Mason. 2005. Prospects and challenges for lorentz-augmented orbits. In *Collection of technical papers - AIAA guidance, navigation, and control conference* 3: 08.
- Pines, S. 1973. Uniform representation of the gravitational potential and its derivatives. *AIAA Journal* 11 (10): 1508–1511.
- Plis, Elena A., Daniel P. Engelhart, Justin Likar, Ryan C. Hoffmann, Russell Cooper, and Dale Ferguson. 2018. Electrical behavior of carbon-loaded kapton for spacecraft applications. *Journal of Spacecraft and Rockets* 55 (3): 775–777.
- Poppe, Andrew R. 2011. *Modeling, theoretical and observational studies of the lunar photoelectron sheath*, University of Colorado at Boulder, Dissertation.
- Páscoa, José, Teixeira, Odélma, and Ribeiro, Gustavo. 2018. A review of propulsion systems for CubeSats.
- Quadrelli, B.M., and Bhaskaran, S. 2019. *Navigation in low gravity*, Ang M., Khatib O., Siciliano B. ed., Encyclopedia of robotics. Berlin: Springer.
- Quadrelli, B.M., Kowalchuck, S., and Chang, J. 2004. Dynamics and control of a herd of sondes guided by a Blimp on Titan. In *14th AAS/AIAA space flight mechanics meeting*, 2004, 5310.
- Quadrelli, B.M., Wood, L.J., Riedel, J.E., McHenry, M.C., Aung, M., Cangahuala, L.A., Volpe, R., Beauchamp, P.M., and Cutts, J. 2015. Guidance, navigation, and control technology assessment for future planetary science missions, journal of guidance, control, and dynamics. *Journal of Guidance, Control, and Dynamics* 1165–1186.
- Quadrelli, M. B., Garrett, H., Castillo, J., Stoica, A., and Ono, H. 2017a. Schaub: active electrostatic flight for airless body. In *2017 IEEE aerospace conference* 1–16.
- Quadrelli, M.B., Garrett, H., Castillo, J., Stoica, A., and Ono, H. 2017b. Schaub: NIAC phase I final report - E-Glider: active electrostatic flight for airless body exploration. *NASA innovative advanced concepts (NIAC)*. https://www.nasa.gov/sites/default/files/atoms/files/niac_2016_phasei_quadrelli_eglider_tagged.pdf.
- Quadrelli, Marco B., Mazhar, Hammad, and Negrut, Dan. 2012. Modeling and simulation of anchoring processes for small body exploration. In *AIAA SPACE 2012 conference and exposition*, 5310.
- Renno, Kok. 2008. Electrical activity and dust lifting on earth, mars, and beyond. *Space Science Reviews* 137 (11): 419–434.

- Schaub, Hanspeter, Gordon G. Parker, and Lyon B. King. 2004. Challenges and prospects of coulomb spacecraft formation control. *Journal of Astronautical Sciences* 52 (1): 169–193.
- Scheeres, D. 1994. Satellite dynamics about asteroids. In *AAS/AIAA spaceflight mechanics meeting*.
- Scheeres, D. 2007. Orbit mechanics about small asteroids. *NASA Archive* 01. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20080012725.pdf>.
- Scheeres, D. 2012. *Orbital motion in strongly perturbed environments*. Bd. 1. Springer.
- Scheeres, D.J. 1999. Satellite dynamics about small bodies: averaged solar radiation pressure effects. *Journal of the Astronautical Sciences* 47 (1): 25–46.
- Scheeres, D.J., Gaskell, R., Abe, S., Barnouin-Jha, O, and Hashimoto, T. 2006. The actual dynamical environment about Itokawa. In *AIAA/AAS astrodynamics specialist conference and exhibit*.
- Scheeres, D.J., and F. Marzari. 2002. Spacecraft dynamics in the vicinity of a comet. *Journal of the Astronautical Science* 50 (1): 35–52.
- Seeni, Aravind, Schäfer, Bernd, and Hirzinger, Gerd. 2010. *Robot mobility systems for planetary surface exploration - state-of-the-art and future outlook: a literature survey*, 189–208. Aerospace Technology Advancements.
- Seydel, Rüdiger. 2009. *Practical bifurcation and stability analysis*. Bd. 5. Springer Science & Business Media.
- Stubbs, Timothy J., Richard R. Vondrak, and William M. Farrell. 2006. A dynamic fountain model for lunar dust. *Advances in Space Research* 37 (1): 59–66.
- Vladimirov, et al. 2005. *Physics and Applications of Complex Plasmas*. Imperial College Press.
- Wang, J., Y. Cao, R. Kafafy, J. Pierru, and V.K. Decyk. 2006. Simulations of ion thruster plume-spacecraft interactions on parallel supercomputer. *IEEE Transactions on Plasma Science* 34 (5): 2148–2158.
- Wang, Joseph, and Hastings, D.E. 1992. Ionospheric plasma flow over large high-voltage space platforms. II: The formation and structure of plasma wake. *Physics of Fluids B: Plasma Physics* 4 (6): 1615–1629
- Wang, Joseph, and Yuan Hu. 2018. The breakdown of the fluid approximation for electrons in a plasma wake. *Journal of Geophysical Research: Space Physics* 123 (10): 8797–8805.
- Whipple, Elden C. 1981. Potentials of surfaces in space. *Reports on progress in Physics* 44 (11): 1197–1250.
- Withrow-Maser, Shannah, Koning, Witold, Kuang, Winnie, and Johnson, Wayne R. 2020. Recent efforts enabling martian rotorcraft missions. In *NASA Technical Reports Server (NTRS)*. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20200000788.pdf>.
- Yu, W., J. Wang, and D. Han. 2016. *2016*. In *AIAA space: Numerical modeling of dust dynamics around small asteroids*.
- Yu, Wiliam, Daoru Han, and Joseph Wang. 2019. Numerical simulations of dust dynamics around small asteroids. *IEEE Transactions on Plasma Science* 47 (8): 3724–3730.

Chapter 3

Tracking and Thrust Vectoring of E-Sail-Based Spacecraft for Solar Activity Monitoring



Marco Bassetto

Abstract The Electric Solar Wind Sail (or E-sail) is a propellantless propulsion system conceived by Dr. Janhunen in 2004. An E-sail extracts momentum from the charged particles constituting the solar wind by means of long and electrically charged tethers, which are deployed and kept stretched by spinning the spacecraft about a symmetry axis. Trajectory analysis of an E-sail-based spacecraft is usually performed assuming that the tether arrangement resembles that of a rigid disc. However, this assumption may be inaccurate since the actual shape of each tether is affected by the chaotic interaction between the solar wind dynamic pressure and the centrifugal force due to the spacecraft spin. The first goal of this chapter is therefore to describe the thrust and torque vectors of a non-flat E-sail. Then, the orbital and the attitude dynamics of a spinning E-sail are analyzed separately due to the marked separation between their characteristic timescales, showing that the torque acting on the E-sail induces a perturbation on the orientation of the thrust vector. Accordingly, the modulation of the tether electrical voltage is proposed and investigated as a possible attitude control strategy. An effective control law is first obtained as a function of spacecraft attitude and time, and then validated through numerical simulations. Finally, two heliocentric mission scenarios (useful, for example, for the monitoring of near-Earth objects or the surveillance of solar activity) are analyzed, where the thrust vectoring of the E-sail is exploited for the generation of Earth-following orbits or the maintenance of a heliostationary equilibrium point.

3.1 Introduction

The Electric Solar Wind Sail (E-sail) is an innovative propellantless propulsion system conceived by Janhunen (2004). An E-sail extracts momentum from the solar wind flow, which electrostatically interacts with long conducting tethers which are

M. Bassetto (✉)

Department of Civil and Industrial Engineering, Aerospace Division, University of Pisa, Via
Girolamo Caruso 8, 56122 Pisa, Italy
e-mail: marco.bassetto@ing.unipi.it

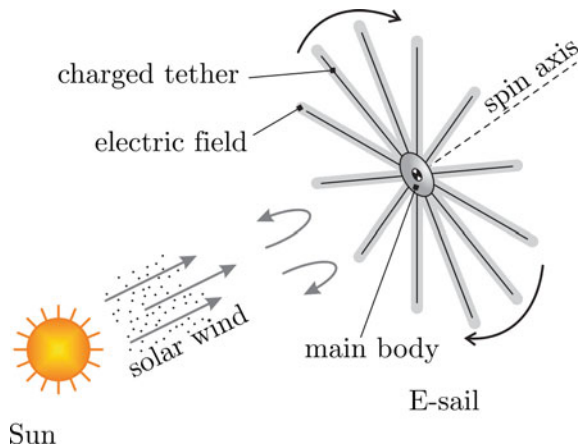
kept charged by an on-board electron gun (Janhunen 2011; Janhunen and Sandroos 2007); see Fig. 3.1.

The same physical principle is also used by the plasma brake, which is a promising option for reducing the decay time of satellites in low-Earth orbits at their end of life (Bassetto et al. 2018b; Janhunen 2010b; Niccolai et al. 2018b; Orsini et al. 2018). The tethers are deployed and kept stretched by rotating the spacecraft about a symmetry axis (Fulton and Schaub 2018; Janhunen et al. 2010, 2013), so that the E-sail qualitatively takes the shape of a spoked wheel (Quarta et al. 2016b). Along with the conventional solar sail, the E-sail is a promising propellantless propulsion system, even though it needs electrical power to produce the required electric field. However, unlike conventional solar sails, a very interesting property of the E-sail is that its thrust magnitude scales as the inverse of the Sun–spacecraft distance (Janhunen 2010a).

In order to get preliminary simulation results, the E-sail thrust vector is often modeled in a simplified way that assumes the tether arrangement resembles that of a rigid disc of given radius (Mengali et al. 2008; Niccolai et al. 2017a, b, 2018a; Quarta and Mengali 2010, 2016b; Yamaguchi and Yamakawa 2013). Using such an ideal configuration, Huo et al. (2018) have recently obtained an analytical description of the E-sail thrust vector using an interesting geometrical approach. However, the Huo model may be inaccurate as the actual shape of each tether is affected by the chaotic interaction between the solar wind dynamical pressure and the centrifugal force due to the spacecraft spin. Accordingly, the tethers are not perfectly straight, and the uncertainty associated with their actual shape makes it difficult to obtain an estimate of the thrust and torque vectors, which are necessary information for both trajectory analysis and attitude control design.

The first goal of this chapter is therefore to discuss a refined mathematical model to describe the E-sail thrust and torque vectors. The mathematical model presented in Sect. 3.2 is a useful improvement over existing ones, as it allows the influence of the tether arrangement on the E-sail performance to be quantified without the use of numerical algorithms (Bassetto et al. 2018a). In particular, Sect. 3.2 first describes

Fig. 3.1 Spinning E-sail conceptual sketch. Adapted from Bassetto et al. (2018a)



the thrust and torque vectors provided by an E-sail with a generic shape, attitude relative to the Sun, and electrical voltage. The general expressions of the thrust and torque vectors are then specialized to the case of an axially symmetric E-sail (which occurs when the tethers are uniformly distributed about the spin axis and have the same shape) with a uniform electrical voltage. Afterwards, an approximate expression of the equilibrium shape of a single tether is analytically determined with the simplifying assumption that the sail spin axis is parallel to the local radial direction. What this shows is that the equilibrium shape of each tether can be described by a natural logarithmic function when the sail spin rate is sufficiently high. Finally, the expressions of the thrust and torque vectors are further simplified assuming that the E-sail maintains the assigned equilibrium shape even when the sail spin axis is not parallel to the local radial line, a reasonable hypothesis if the angle between them does not exceed a few degrees.

Starting from that model, Sect. 3.3 introduces the differential equations that describe the motion of a spacecraft under the action of the propulsive force and torque provided by a rigid E-sail with a uniform electrical voltage (Bassetto et al. 2019b). The orbital motion and the attitude dynamics are analyzed separately due to the marked separation between their characteristic timescales. It is shown that the torque acting on the conducting tethers induces a perturbation on the orientation of the thrust vector, thus reducing the maneuvering capabilities of an E-sail-based spacecraft.

Section 3.4 proposes an effective control law that can remove the disturbance torque due to the tether bending by suitably adjusting the tether electrical voltage (Bassetto et al. 2019a). It is shown that the proposed solution requires a small electrical voltage modulation and guarantees the maintenance of the required thrust magnitude. Moreover, Sect. 3.4 proves that the modulation of the tether electrical voltage is also a feasible option for actively controlling and maintaining the spacecraft attitude (Bassetto et al. 2020). The proposed control law, which is analytically derived as a function of the spacecraft attitude and time, is validated through numerical simulations.

Finally, Sect. 3.5 analyzes two heliocentric mission scenarios, useful for the monitoring of near-earth objects (NEOs) or the surveillance of solar activity, in which the E-sail is used for the generation of earth-following orbits (EFOs) (Bassetto et al. 2019c) or the maintenance of a heliostationary condition (Bassetto et al. 2019b). In particular, the former scenario is obtained when the apse line of the spacecraft osculating orbit precedes at an angular rate equal to the Earth's mean motion. In this case the problem is addressed using a locally optimal formulation, in which the E-sail control parameters are chosen such as to maximize the time derivative of the osculating argument of perihelion of the spacecraft orbit. The latter, instead, is obtained when the spacecraft velocity relative to the Sun is zero and the propulsive acceleration is used for balancing the local Sun's gravitational field. In this case, it is shown that a proportional controller is sufficient for stabilizing the E-sail radial dynamics in the vicinity of the (nominal) equilibrium point.

3.2 E-sail Concept and Modeling

Consider an E-sail-based spacecraft and assume the vehicle to be modeled as an axially symmetric rigid body spinning about its symmetry axis \hat{n} at an angular velocity $\omega \triangleq \omega \hat{n}$ of constant magnitude ω . The E-sail consists of $N \geq 2$ tethers, which are modeled as planar cables belonging to the plane (\hat{i}_k, \hat{n}) , where \hat{i}_k (with $k = \{0, 1, \dots, N - 1\}$) is orthogonal to \hat{n} ; see Fig. 3.2.

The displacement of the generic tether with respect to the spacecraft main body can be evaluated by introducing a principal body reference frame $\mathcal{T}_B(S; x_B, y_B, z_B)$ with the origin S at the spacecraft center of mass and unit vectors $\{\hat{i}_B, \hat{j}_B, \hat{k}_B\}$, defined as

$$\hat{k}_B \triangleq \hat{n} \quad , \quad \hat{i}_B \triangleq \hat{i}_0 \quad , \quad \hat{j}_B \triangleq \hat{n} \times \hat{i}_0 \quad (3.1)$$

Note that the plane (\hat{i}_B, \hat{k}_B) contains the first tether (labeled with $k = 0$), whereas the unit vector \hat{i}_k can be written as

$$\hat{i}_k = \cos \zeta_k \hat{i}_B + \sin \zeta_k \hat{j}_B \quad (3.2)$$

where ζ_k is the angle, measured counterclockwise from the direction of \hat{i}_B , between x_k and x_B ; see Fig. 3.2. Likewise, ζ_k is the angle between the plane containing the k -th tether and the plane identified by the unit vectors \hat{i}_B and \hat{k}_B . Assume now that the shape of the generic tether can be described, in the plane (\hat{i}_k, \hat{k}_B) , through a continuously differentiable function $f_k = f_k(x_k) : [x_{r_k}, x_{t_k}] \in \mathbb{R}$, where $x_{r_k} \geq 0$ (or $x_{t_k} > 0$) is the distance of the tether root (or tip) from the spacecraft spin axis z_B ; see Fig. 3.3.

Fig. 3.2 E-sail geometrical arrangement. Adapted from Bassetto et al. (2018a)

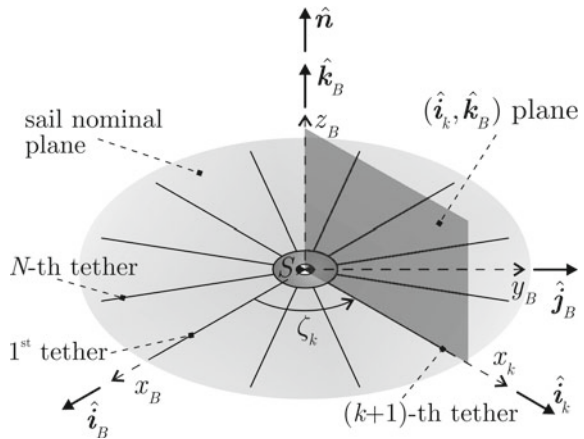


Fig. 3.3 Generic tether displacement. Adapted from Bassetto et al. (2018a)



The position vector \mathbf{d}_k of an infinitesimal arc-length ds_k of the k -th conducting tether is given by

$$\mathbf{d}_k = x_k \hat{\mathbf{i}}_k + f_k \hat{\mathbf{k}}_B \tag{3.3}$$

with

$$ds_k = \sqrt{1 + (f'_k)^2} dx_k \tag{3.4}$$

where $f'_k \triangleq df_k/dx_k$ is the local tether slope. From Eqs. (3.3) and (3.4), the expression of the unit vector $\hat{\mathbf{s}}_k$ tangent to the generic tether at point (x_k, f_k) is

$$\hat{\mathbf{s}}_k \triangleq \frac{d\mathbf{d}_k}{ds_k} = \frac{dx_k \hat{\mathbf{i}}_k + df_k \hat{\mathbf{k}}_B}{\sqrt{1 + (f'_k)^2} dx_k} = \frac{\hat{\mathbf{i}}_k + f'_k \hat{\mathbf{k}}_B}{\sqrt{1 + (f'_k)^2}} \tag{3.5}$$

which can be rewritten using Eq. (3.2) as a function of $\{\hat{\mathbf{i}}_B, \hat{\mathbf{j}}_B, \hat{\mathbf{k}}_B\}$ as

$$\hat{\mathbf{s}}_k = \frac{\cos \zeta_k \hat{\mathbf{i}}_B + \sin \zeta_k \hat{\mathbf{j}}_B + f'_k \hat{\mathbf{k}}_B}{\sqrt{1 + (f'_k)^2}}. \tag{3.6}$$

3.2.1 Thrust and Torque Vectors of a Three-Dimensional E-sail

The aim of this section is to obtain an analytical expression of both the thrust and torque vectors generated by a spinning E-sail of a given three-dimensional shape, under the main assumption that each tether belongs to a plane containing the spacecraft spin axis z_B . Both the total force and the total torque generated by the E-sail can be computed starting from the elementary force $d\mathbf{F}_k$ due to the solar wind dynamical pressure acting on an infinitesimal arc-length ds_k of the generic conducting tether. According to recent work by Janhunen et al. (2010) and Toivanen and Janhunen (2013, 2017), when the Sun–spacecraft distance r is on the order of $r_\oplus = 1$ au, the thrust $d\mathbf{F}_k$ gained by ds_k is

$$d\mathbf{F}_k = \sigma_k \mathbf{u}_{\perp k} ds_k \quad (3.7)$$

where $\mathbf{u}_{\perp k}$ is the component of the solar wind velocity \mathbf{u} perpendicular to $\hat{\mathbf{s}}_k$, whereas

$$\sigma_k \triangleq \sigma_\oplus \left(\frac{r_\oplus}{r} \right) \quad \text{with} \quad \sigma_\oplus \triangleq 0.18 \max(0, V_k - V_w) \sqrt{\epsilon_0 m_p n_\oplus} \quad (3.8)$$

in which V_k (ranging in the interval [20, 40] kV) is the tether voltage, V_w is the electric potential of the solar wind ions, with a typical value of about 1 kV (Janhunen et al. 2010), ϵ_0 is the vacuum permittivity, m_p is the proton mass, and $n_\oplus = 5 \times 10^6 \text{ m}^{-3}$ is the average solar wind number density at $r = r_\oplus$. Assuming a purely radial solar wind stream, that is, $\mathbf{u} \triangleq u \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the Sun–spacecraft unit vector and $u \simeq 400 \text{ km/s}$ is the solar wind speed, the term $\mathbf{u}_{\perp k}$ in Eq. (3.7) becomes

$$\mathbf{u}_{\perp k} = u(\hat{\mathbf{s}}_k \times \hat{\mathbf{r}}) \times \hat{\mathbf{s}}_k = u[\hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{s}}_k) \hat{\mathbf{s}}_k] \quad (3.9)$$

Moreover, according to Fig. 3.4, the Sun–spacecraft unit vector $\hat{\mathbf{r}}$ can be written as a function of $\{\hat{\mathbf{i}}_B, \hat{\mathbf{j}}_B, \hat{\mathbf{k}}_B\}$ as

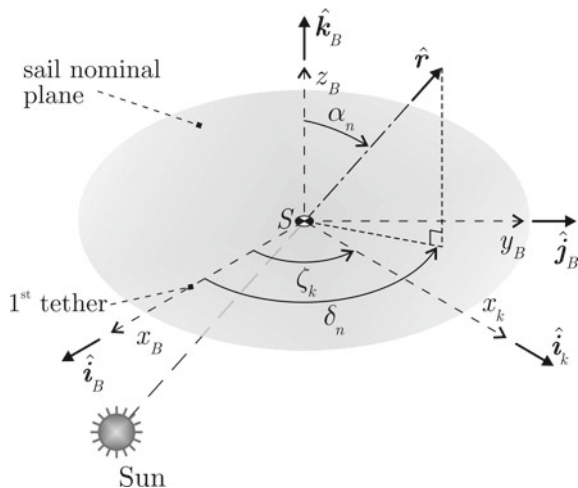
$$\hat{\mathbf{r}} = \sin \alpha_n \cos \delta_n \hat{\mathbf{i}}_B + \sin \alpha_n \sin \delta_n \hat{\mathbf{j}}_B + \cos \alpha_n \hat{\mathbf{k}}_B \quad (3.10)$$

where

$$\alpha_n = \arccos(\hat{\mathbf{r}} \cdot \hat{\mathbf{k}}_B) \in [0, \pi] \text{ rad} \quad (3.11)$$

referred to as pitch angle, is the angle between $\hat{\mathbf{r}}$ and $\hat{\mathbf{k}}_B$, while

Fig. 3.4 Pitch (α_n) and clock (δ_n) angles. Adapted from Bassetto et al. (2018a)



$$\delta_n \triangleq \begin{cases} \arccos\left(\frac{\hat{r} \cdot \hat{i}_B}{|\hat{r} \times \hat{k}_B|}\right) & \text{if } \hat{r} \cdot \hat{j}_B \geq 0 \\ 2\pi - \arccos\left(\frac{\hat{r} \cdot \hat{i}_B}{|\hat{r} \times \hat{k}_B|}\right) & \text{otherwise} \end{cases} \quad (3.12)$$

is the clock angle, which ranges in the interval $[0, 2\pi]$ rad and is measured counter-clockwise starting from \hat{i}_B between x_B and the projection of \hat{r} on the sail nominal plane (x_B, y_B); see Fig. 3.4.

Note that δ_n is undefined when $\alpha_n = 0$, that is, when the E-sail is Sun facing. In particular, the Sun-facing configuration occurs when the sail nominal plane (x_B, y_B) is orthogonal to local radial unit vector \hat{r} and, accordingly, $\hat{k}_B \equiv \hat{r}$. Taking into account Eqs. (3.6) and (3.10), the dot product $\hat{r} \cdot \hat{s}_k$ in Eq. (3.9) can be rearranged as

$$\hat{r} \cdot \hat{s}_k = \frac{\hat{r} \cdot \hat{i}_k + f'_k \hat{r} \cdot \hat{k}_B}{\sqrt{1 + (f'_k)^2}} = \frac{\cos(\delta_n - \zeta_k) \sin \alpha_n + f'_k \cos \alpha_n}{\sqrt{1 + (f'_k)^2}} \quad (3.13)$$

Therefore, with the aid of Eqs. (3.4), (3.6), (3.9), and (3.13), the elementary thrust $d\mathbf{F}_k$, provided by Eq. (3.7), can be rewritten as

$$d\mathbf{F}_k = \sigma_k u \left[\hat{r} - \frac{\cos(\delta_n - \zeta_k) \sin \alpha_n + f'_k \cos \alpha_n}{1 + (f'_k)^2} \left(\cos \zeta_k \hat{i}_B + \sin \zeta_k \hat{j}_B + f'_k \hat{k}_B \right) \right] \sqrt{1 + (f'_k)^2} dx_k \quad (3.14)$$

3.2.1.1 Total Force

Starting from Eq. (3.14), the total force $d\mathbf{F}_k$ acting on the infinitesimal arc-length ds_k of the k -th tether is

$$d\mathbf{F}_k = d\mathcal{A}_k \hat{\mathbf{r}} + d\mathcal{B}_k \hat{\mathbf{i}}_B + d\mathcal{C}_k \hat{\mathbf{j}}_B + d\mathcal{D}_k \hat{\mathbf{k}}_B \quad (3.15)$$

where

$$d\mathcal{A}_k = \sigma_k u \sqrt{1 + (f'_k)^2} dx_k \quad (3.16)$$

$$d\mathcal{B}_k = -\sigma_k u \cos \zeta_k \frac{\cos(\delta_n - \zeta_k) \sin \alpha_n + f'_k \cos \alpha_n}{\sqrt{1 + (f'_k)^2}} dx_k \quad (3.17)$$

$$d\mathcal{C}_k = -\sigma_k u \sin \zeta_k \frac{\cos(\delta_n - \zeta_k) \sin \alpha_n + f'_k \cos \alpha_n}{\sqrt{1 + (f'_k)^2}} dx_k \quad (3.18)$$

$$d\mathcal{D}_k = -\sigma_k u f'_k \frac{\cos(\delta_n - \zeta_k) \sin \alpha_n + f'_k \cos \alpha_n}{\sqrt{1 + (f'_k)^2}} dx_k \quad (3.19)$$

Therefore, the force \mathbf{F}_k acting on the generic conducting tether is

$$\mathbf{F}_k = \int_{x_{r_k}}^{x_{l_k}} d\mathbf{F}_k = \mathcal{A}_k \hat{\mathbf{r}} + \mathcal{B}_k \hat{\mathbf{i}}_B + \mathcal{C}_k \hat{\mathbf{j}}_B + \mathcal{D}_k \hat{\mathbf{k}}_B \quad (3.20)$$

with

$$\mathcal{A}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{A}_k, \mathcal{B}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{B}_k, \mathcal{C}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{C}_k, \mathcal{D}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{D}_k \quad (3.21)$$

whereas the total force \mathbf{F} acting on the E-sail (composed of $N \geq 2$ tethers) is given by

$$\mathbf{F} = \sum_{k=0}^{N-1} \mathbf{F}_k = \mathcal{A} \hat{\mathbf{r}} + \mathcal{B} \hat{\mathbf{i}}_B + \mathcal{C} \hat{\mathbf{j}}_B + \mathcal{D} \hat{\mathbf{k}}_B \quad (3.22)$$

where

$$\mathcal{A} \triangleq \sum_{k=0}^{N-1} \mathcal{A}_k, \mathcal{B} \triangleq \sum_{k=0}^{N-1} \mathcal{B}_k, \mathcal{C} \triangleq \sum_{k=0}^{N-1} \mathcal{C}_k, \mathcal{D} \triangleq \sum_{k=0}^{N-1} \mathcal{D}_k. \quad (3.23)$$

3.2.1.2 Total Torque

The torque $d\mathbf{T}_k$ given by an infinitesimal arc-length ds_k of the k -th conducting tether is

$$d\mathbf{T}_k = d\mathbf{r}_k \times d\mathbf{F}_k \quad (3.24)$$

Taking into account the expressions of $d\mathbf{r}_k$ and $d\mathbf{F}_k$ given by Eqs. (3.3) and (3.15), respectively, and using Eq. (3.10), $d\mathbf{T}_k$ can be written as a function of $\{\hat{\mathbf{i}}_B, \hat{\mathbf{j}}_B, \hat{\mathbf{k}}_B\}$ as

$$d\mathbf{T}_k = d\mathcal{E}_k \hat{\mathbf{i}}_B + d\mathcal{F}_k \hat{\mathbf{j}}_B + d\mathcal{G}_k \hat{\mathbf{k}}_B \quad (3.25)$$

where

$$\begin{aligned} d\mathcal{E}_k = & \left\{ x_k \sin \zeta_k \left[\sigma_k u \cos \alpha_n - \frac{f'_k \sigma_k u (\sin \alpha_n \cos (\delta_n - \zeta_k) + f'_k \cos \alpha_n)}{1 + (f'_k)^2} \right] \right. \\ & \left. + f_k \sin \zeta_k \frac{\sigma_k u (\sin \alpha_n \cos (\delta_n - \zeta_k) + f'_k \cos \alpha_n)}{1 + (f'_k)^2} - f_k \sigma_k u \sin \alpha_n \sin \delta_n \right\} \\ & \sqrt{1 + (f'_k)^2} dx_k \end{aligned} \quad (3.26)$$

$$\begin{aligned} d\mathcal{F}_k = & \left\{ -x_k \cos \zeta_k \left[\sigma_k u \cos \alpha_n - \frac{f'_k \sigma_k u (\sin \alpha_n \cos (\delta_n - \zeta_k) + f'_k \cos \alpha_n)}{1 + (f'_k)^2} \right] \right. \\ & \left. - f_k \cos \zeta_k \frac{\sigma_k u (\sin \alpha_n \cos (\delta_n - \zeta_k) + f'_k \cos \alpha_n)}{1 + (f'_k)^2} + f_k \sigma_k u \sin \alpha_n \cos \delta_n \right\} \\ & \sqrt{1 + (f'_k)^2} dx_k \end{aligned} \quad (3.27)$$

$$d\mathcal{G}_k = \sigma_k u x_k \sin \alpha_n \sin (\delta_n - \zeta_k) \sqrt{1 + (f'_k)^2} dx_k \quad (3.28)$$

Therefore, the torque \mathbf{T}_k acting on the generic tether is

$$\mathbf{T}_k = \int_{x_{rk}}^{x_{lk}} d\mathbf{T}_k = \mathcal{E}_k \hat{\mathbf{i}}_B + \mathcal{F}_k \hat{\mathbf{j}}_B + \mathcal{G}_k \hat{\mathbf{k}}_B \quad (3.29)$$

with

$$\mathcal{E}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{E}_k, \mathcal{F}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{F}_k, \mathcal{G}_k \triangleq \int_{x_{r_k}}^{x_{l_k}} d\mathcal{G}_k \quad (3.30)$$

whereas the total torque \mathbf{T} acting on the E-sail is

$$\mathbf{T} = \sum_{k=0}^{N-1} \mathbf{T}_k = \mathcal{E} \hat{\mathbf{i}}_B + \mathcal{F} \hat{\mathbf{j}}_B + \mathcal{G} \hat{\mathbf{k}}_B \quad (3.31)$$

where

$$\mathcal{E} \triangleq \sum_{k=0}^{N-1} \mathcal{E}_k, \mathcal{F} \triangleq \sum_{k=0}^{N-1} \mathcal{F}_k, \mathcal{G} \triangleq \sum_{k=0}^{N-1} \mathcal{G}_k \quad (3.32)$$

Equations (3.22) and (3.31) are the expressions of the total force and torque acting on an E-sail with a given tether shape, length, and angular separation between tethers. These results will be applied to the noteworthy case of a Sun-facing E-sail (Mengali et al. 2013; Quarta and Mengali 2016a), thus obtaining a set of closed-form relations. Some simplifying assumptions need to be introduced to get a more tractable form of both \mathbf{F} and \mathbf{T} , as is thoroughly discussed in the next section.

3.2.2 Thrust and Torque Vectors of an Axially Symmetric E-sail

This section makes the previous general expressions of the thrust and torque vectors specific to the case of an axially symmetric E-sail. In this case, the tethers are assumed to both have the same shape and electrical voltage and to be uniformly distributed about the z_B -axis, that is

$$\zeta_k = \frac{2\pi k}{N} \quad (3.33)$$

As such, the E-sail is symmetric about the z_B -axis and the shape of its tethers can be described through a suitable differentiable function $f = f(x)$, where the x -axis is orthogonal to z_B and (x, z_B) defines the plane where the generic tether lies; see Fig. 3.3. Using the general mathematical model discussed in Sect. 3.2.1, the thrust (\mathbf{F}) and torque (\mathbf{T}) vectors acting on an axially symmetric E-sail of given shape can be expressed in analytical form as a function of the E-sail attitude. In fact, assuming $r = r_{\oplus}$, the vectors \mathbf{F} and \mathbf{T} become

$$\mathbf{F} = \frac{1}{2} N L \sigma u \left[(2 - \mathcal{P}) \hat{\mathbf{r}} + (3\mathcal{P} - 2) (\hat{\mathbf{r}} \cdot \hat{\mathbf{k}}_B) \hat{\mathbf{k}}_B \right] \quad (3.34)$$

$$\mathbf{T} = \frac{1}{2} \mathcal{M} N L^2 \sigma u (\hat{\mathbf{k}}_B \times \hat{\mathbf{r}}) \quad (3.35)$$

where $\mathcal{P} \in [0, 1]$ and $\mathcal{M} \in [0, 1]$ are two dimensionless coefficients related to the tether shape f through the equations

$$\mathcal{P} \triangleq \frac{1}{L} \int_0^{x_r} \frac{dx}{\sqrt{1 + (f')^2}} \quad (3.36)$$

$$\mathcal{M} \triangleq \frac{1}{L^2} \int_0^{x_r} \frac{f \left[1 + 2(f')^2 \right] + x f'}{\sqrt{1 + (f')^2}} dx \quad (3.37)$$

where L is given by

$$L \triangleq \int_0^{x_r} \sqrt{1 + (f')^2} dx \quad (3.38)$$

and the abscissa of the root section is neglected (i.e., $x_r = 0$). It is worth noting that \mathbf{T} induces a pitch oscillation resembling that of a spherical pendulum.

In the special case of a flat shape, that is, when $f = f' = 0$ and all tethers belong to the (x_B, y_B) plane, Eqs. (3.36)–(3.38) give $\mathcal{P} = 1$, $\mathcal{M} = 0$, and $L = x_r$. In that case, the torque is zero independent of the E-sail attitude, whereas Eq. (3.34) reduces to

$$\mathbf{F} = \frac{1}{2} N L \sigma u \left[\hat{\mathbf{r}} + (\hat{\mathbf{r}} \cdot \hat{\mathbf{k}}_B) \hat{\mathbf{k}}_B \right] \quad (3.39)$$

consistently with the results discussed by Huo et al. (2018). The magnitude of \mathbf{F} when $f = f' = 0$ is

$$F = \frac{1}{2} N L \sigma u \sqrt{1 + 3 \cos^2 \alpha_n} \quad (3.40)$$

which depends on the Sun–spacecraft distance r through the parameter σ defined in Eq. (3.8).

3.2.3 Tether Equilibrium Shape

The problem of describing the actual E-sail equilibrium shape is substantially simplified when the spacecraft spin axis is aligned with the solar wind velocity vector. In that case, each tether can be thought of as being aligned with the force field and belonging to a plane containing the spacecraft spin axis. An estimation of the tether equilibrium shape of a Sun-facing E-sail can be obtained using the approach discussed by Toivanen and Janhunen (2017). Assuming a spinning E-sail, Toivanen and Janhunen (2017) describe the tether equilibrium shape with an integral equation, which is solved numerically. In particular, using an analytical approximation of the tether shape, Toivanen and Janhunen (2017) also obtain closed-form expressions for both the thrust and torque arising from the solar wind momentum transfer. Their results essentially state that the tethers form a cone near the spacecraft, while they are substantially flattened around the tip region by the centrifugal force. A new approximation of the tether equilibrium shape will be achieved in the next section using an analytical procedure.

In this section, instead, it is shown that the tether tip slope may be found in closed form with some simplifying hypotheses. The procedure to perform such a calculation starts by assuming that the equilibrium shape of each tether results from a combination between the centrifugal force due to the spacecraft spin about \hat{n} and the electrical interaction of the solar wind particles with the conducting tether. Since each tether is assumed to be at equilibrium, the sum between the centrifugal and the electrical forces is balanced by the constraint reaction at the root section of the cable. It is worth noting that this simplified model neglects the inertial forces due to the E-sail acceleration. Actually, the tether shape also depends on the instantaneous acceleration of the spacecraft, which, in its turn, is affected by the current shape of the E-sail. However, such a problem is not solvable with an analytical approach, as its solution requires an iterative numerical procedure, which is very expensive from a computational point of view. The removal of the inertial forces in the presented model is equivalent to assuming that the E-sail velocity is constant, a reasonable approximation within sufficiently small time intervals. Also note that at a Sun–sail distance $r = r_{\oplus}$, the ratio of the electrical to the gravitational force per unit length is equal to

$$\frac{\sigma_{\oplus} u}{\rho \mu_{\odot} / r_{\oplus}^2} \simeq 6.2737 \quad (3.41)$$

where ρ is the tether linear mass density (approximately equal to 10^{-5} kg/m for a μm -diameter aluminum tether), whereas $\sigma_{\oplus} \simeq 9.3 \times 10^{-13}$ kg/m/s when $V = 20$ kV; see Eq. (3.8). Therefore, from Eq. (3.1), the gravitational effects on the E-sail are small when compared to the electrical forces acting on the tethers.

The elementary centrifugal force $d\mathbf{F}_{\omega_k}$ acting on ds_k can be written recalling that x_k is the distance of ds_k from the spacecraft spin axis z_B (see Fig. 3.3), that is

$$d\mathbf{F}_{\omega_k} = \rho \omega^2 x_k ds_k \hat{\mathbf{i}}_k = \rho \omega^2 x_k \sqrt{1 + (f'_k)^2} dx_k \hat{\mathbf{i}}_k \quad (3.42)$$

where $\hat{\mathbf{i}}_k$ is given by Eq. (3.2) as a function of $\{\hat{\mathbf{i}}_B, \hat{\mathbf{j}}_B\}$. Enforcing the Sun-facing condition $\alpha_n = \delta_n = 0$ into Eqs. (3.15)–(3.19), the sum $d\mathbf{F}_{s_k}$ between $d\mathbf{F}_k$ and $d\mathbf{F}_{\omega_k}$ becomes

$$d\mathbf{F}_{s_k} = \sigma_k u \left[\frac{\rho \omega^2 x_k}{\sigma_k u} - \frac{f'_k}{1 + (f'_k)^2} \hat{\mathbf{i}}_k + \frac{1}{1 + (f'_k)^2} \hat{\mathbf{k}}_B \right] \sqrt{1 + (f'_k)^2} dx_k \quad (3.43)$$

where $\hat{\mathbf{i}}_k$, given by Eq. (3.2), is the unit vector obtained from the projection of $d\mathbf{F}_{s_k}$ on the E-sail nominal plane (x_B, y_B) . Without loss of generality, the notation may be simplified by dropping the subscript k in the variables $\{x_k, f'_k, \sigma_k, \hat{\mathbf{i}}_k\}$ of Eq. (3.43).

Assume now the generic tether to have no bending stiffness, so that only an internal tension acts tangentially to its neutral axis. In this case, according to Toivanen and Janhunen (2017), the direction of the vector tangent to the tether at a generic point P of abscissa $x \in [x_r, x_t]$ is parallel to the direction of the integral of $d\mathbf{F}_s$ from x to x_t (i.e., the integral of the total force from P to the tether tip). Therefore, from Eq. (3.43), the tether slope f' at point P is the solution of the following integral–differential equation

$$f'(x) = \frac{\sigma u \int_x^{x_t} \frac{d\xi}{\sqrt{1 + (f')^2}}}{\rho \omega^2 \int_x^{x_t} \xi \sqrt{1 + (f')^2} d\xi - \sigma u \int_x^{x_t} \frac{f' d\xi}{\sqrt{1 + (f')^2}}} \quad (3.44)$$

where the numerator (or denominator) in the right-hand side is the component along the z_B -axis (or x_B -axis) of the resultant force acting on the tether arc between P and the tip, that is

$$F_x(x) \triangleq \rho \omega^2 \int_x^{x_t} \xi \sqrt{1 + (f')^2} d\xi - \sigma u \int_x^{x_t} \frac{f' d\xi}{\sqrt{1 + (f')^2}} \quad (3.45)$$

$$F_z(x) \triangleq \sigma u \int_x^{x_t} \frac{d\xi}{\sqrt{1 + (f')^2}} \quad (3.46)$$

Introduce now the dimensionless abscissa $h \triangleq x/x_t$, with $h \in [h_r, 1]$, where $h_r \triangleq x_r/x_t \geq 0$ is the value at the root section. Equation (3.44) can be conveniently rewritten as

$$f'(h) = \frac{\int_h^1 \frac{d\xi}{\sqrt{1+(f')^2}}}{K \int_h^1 \xi \sqrt{1+(f')^2} d\xi - \int_h^1 \frac{f' d\xi}{\sqrt{1+(f')^2}}} \quad (3.47)$$

where $K > 0$ is a dimensionless shaping parameter, defined as

$$K \triangleq \frac{\rho \omega^2 x_t}{\sigma u} \quad (3.48)$$

which relates the tether equilibrium shape of a Sun-facing E-sail to the ratio of the centrifugal ($\rho \omega^2 x_t$) to the electrical (σu) effects. The tether slope at the tip, that is, the exact value of $f'_t \triangleq f'(h = 1)$ can be obtained from Eq. (3.47) using a limiting procedure, viz.

$$f'_t = \lim_{h \rightarrow 1} f'(h) = \frac{1}{K [1 + (f'_t)^2] - f'_t} \quad (3.49)$$

Equation (3.49) can be rewritten as

$$\left(f'_t - \frac{1}{K}\right) \left[1 + (f'_t)^2\right] = 0 \quad (3.50)$$

the only real solution of which is

$$f'_t = \frac{1}{K} = \frac{\sigma u}{\rho \omega^2 x_t} \quad (3.51)$$

As expected, the tether slope at the tip sharply reduces as the E-sail spin rate increases. The variation of f'_t with $\{x_t, \omega\}$ is shown in Fig. 3.5 for $\rho = 10^{-5}$ kg/m and $V = 20$ kV. In particular, $f'_t \leq 0.1$ (or $K \geq 10$) when $\omega \geq 5$ rph (with $1\text{rph} \simeq 1.7453 \times 10^{-3}$ rad/s) and $x_t \geq 5$ km, which implies a tether slope at the tip less than 6° .

Having obtained the exact value of f'_t , it is now possible to calculate the function $f'(x)$ (or $f'(h)$). To that end, a recursive procedure is necessary, which, starting from the tether tip and backward proceeding towards the root, numerically solves Eq. (3.47) for a given value of K . The results of such a procedure are summarized in Fig. 3.6 for some values of the shaping parameter K . In particular, Fig. 3.6 shows that $f'_t = 1/K$, in agreement with Eq. (3.51).

Note also that the tether slope at the root becomes $f'(0) \simeq 2f'_t = 2/K$ when the shaping parameter is sufficiently large (that is, when $K \geq 5$) and $h_r = 0$. In particular, the latter condition amounts to neglecting the main body width and to assuming the root section to be attached to the z_B -axis. In that case, $1/K \leq f' \leq 2/K$ or, equivalently

Fig. 3.5 Tip slope f'_t as a function of the spin rate ω and the spin axis-tip distance x_t ; see Eq. (3.51). Adapted from Bassetto et al. (2018a)

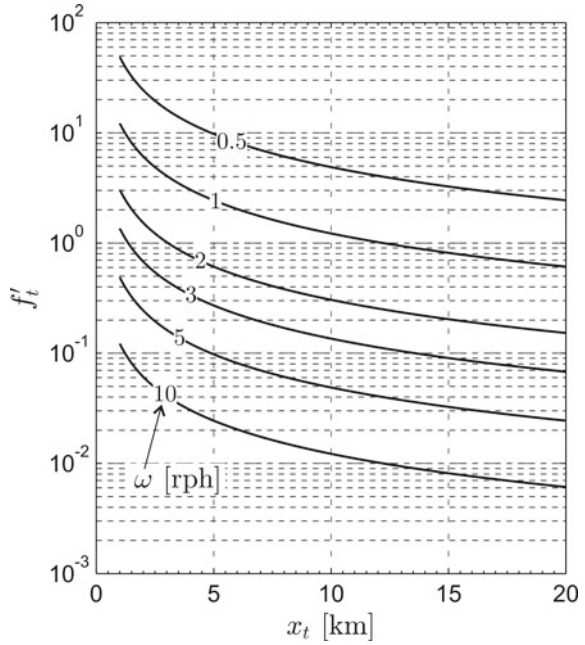
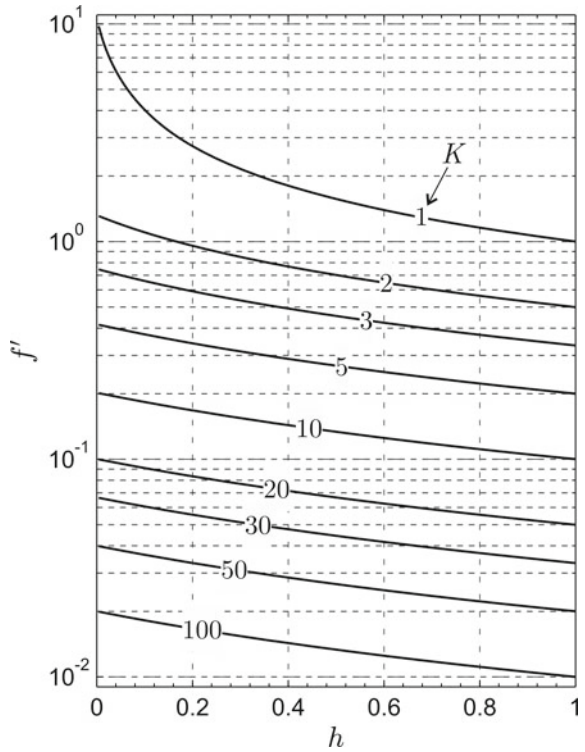


Fig. 3.6 Tether slope f' as a function of the dimensionless abscissa $h \triangleq x/x_t$ and the shaping parameter K ; see Eq. (3.48). Adapted from Bassetto et al. (2018a)



$$\frac{K^2 + 1}{K^2} \leq 1 + (f')^2 \leq \frac{K^2 + 4}{K^2} \tag{3.52}$$

which implies

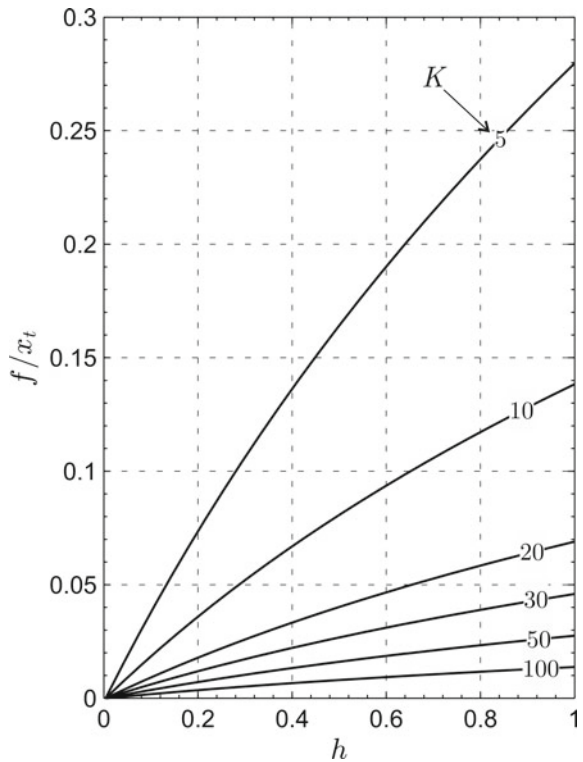
$$1 + (f')^2 \simeq 1 \tag{3.53}$$

The tether shape may be obtained by means of a numerical integration, and the results are summarized in Fig. 3.7 assuming $h_r = 0$.

Notably, an accurate analytical approximation can be obtained for a sufficiently large value of the shaping parameter, for example when $K \geq 5$. In that case, substituting Eq. (3.53) into Eq. (3.47), the result is

$$f'(h) \simeq \frac{\int_h^1 d\xi}{K \int_h^1 \xi d\xi - \int_h^1 f' d\xi} = \frac{2(1-h)/K}{1-h^2 - \frac{2[f_t - f(h)]}{Kx_t}} \tag{3.54}$$

Fig. 3.7 Tether shape as a function of $h \triangleq x/x_t$ and K obtained through numerical integration. Adapted from Bassetto et al. (2018a)



Since $\max\{2[f_t - f(h)]/(Kx_t)\} \simeq 0.11$ (see Fig. 3.7), the last relation may be further simplified as

$$f'(h) \simeq \frac{2}{K(1+h)} \tag{3.55}$$

Notably, the approximation (3.55) gives the exact value at tether tip (i.e., $f'_t = 1/K$), and also captures the approximate value at tether root (i.e., $f'(0) = 2/K$), in agreement with the estimate obtained previously. Figure 3.8 compares the analytical approximation of Eq. (3.55) (dashed line) with the numerical solution (solid line), showing that the two results are nearly coincident when $K \geq 5$.

Accordingly, an accurate analytical solution of the tether shape can be found from Eq. (3.55). Indeed, using a variable separation and integrating both sides, it may be verified that

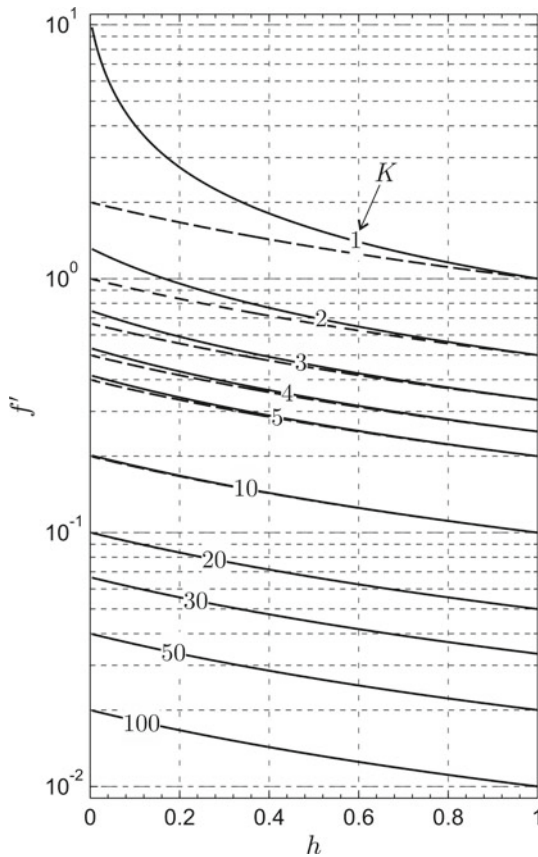


Fig. 3.8 Tether slope f' as a function of h and K : numerical (solid line) vs. analytical approximation (dashed line). Adapted from Bassetto et al. (2018a)

$$f(h) = \frac{2x_t}{K} \ln\left(\frac{1+h}{1+h_r}\right) \text{ with } h \in [h_r, 1] \quad (3.56)$$

or, using Eq. (3.48)

$$f(x) = \frac{2\sigma u}{\rho\omega^2} \ln\left(\frac{x+x_t}{x_r+x_t}\right) \text{ with } x \in [x_r, x_t] \quad (3.57)$$

Equation (3.57) proves the importance of the logarithmic function for describing the equilibrium shape of a Sun-facing E-sail. Its actual accuracy is better appreciated with the aid of Fig. 3.9, which plots Eq. (3.56) with $h_r = 0.01$. The obtained results are nearly coincident with those reported in Fig. 3.7, which correspond to a numerical integration of the actual tether slope.

This result is qualitatively in accordance with the numerical simulations by Toivanen and Janhunen (2017), which show that the tethers form a cone near the spacecraft, whereas they are flattened by the centrifugal force near the tip region. Actually, the analytical approximation by Toivanen and Janhunen (2017) estimates a parabolic shape of the tethers, with the effect of a null slope at their tips. The discrepancy between the two models is consistent with the assumption that in the

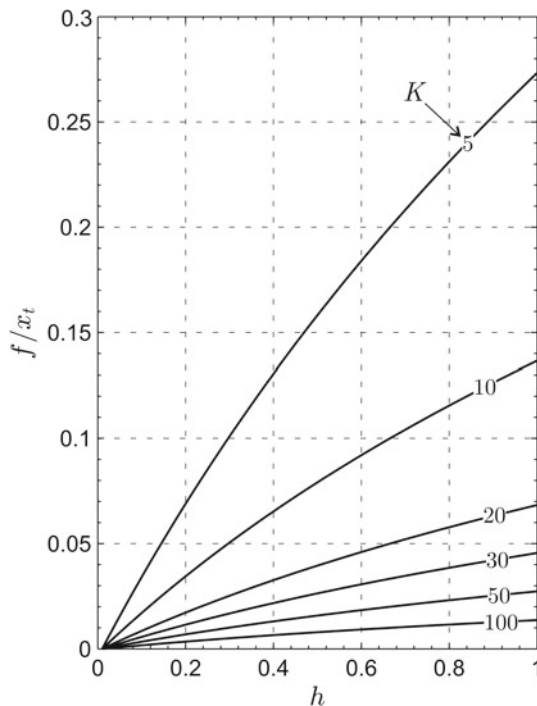


Fig. 3.9 Tether approximate shape as a function of the dimensionless abscissa $h \triangleq x/x_t$ and K when $h_r = 0.01$; see Eq. (3.56). Adapted from Bassetto et al. (2018a)

paper by Toivanen and Janhunen (2017) the tips of the main tethers host remote units connected to an external rim in order to provide mechanical stability to the sail. In the presented case, instead, the remote units are not included in the model with the aim of decreasing the stress at tether root. As such, a nonzero tip slope is an expected result.

3.2.4 Approximate \mathcal{P} and \mathcal{M} for a Logarithmic Shape Tether

Section 3.2.3 introduced the dimensionless shape coefficient K , which relates the tether equilibrium shape of a Sun-facing E-sail to the ratio of the centrifugal ($\rho\omega^2x_t$) to the electrical (σu) effects; see Eq. (3.48). Section 3.2.3 also provides an analytical approximation of the tether shape, which is valid as long as $K \geq 5$, or

$$\omega \geq \omega_{\min} \triangleq \sqrt{\frac{5\sigma u}{\rho x_t}} \tag{3.58}$$

which means that the E-sail spin rate ω must be sufficiently high. The analytical expression of the tether equilibrium shape found in Sect. 3.2.3 (see Eq. (3.57)) is here reported assuming $x_r = 0$, viz.

$$f(x) = b_l x_t \ln\left(1 + \frac{x}{x_t}\right) \text{ with } x \in [0, x_t] \tag{3.59}$$

where b_l is defined as

$$b_l \triangleq \frac{2\sigma u}{\rho \omega^2 x_t} = \frac{2}{K} \tag{3.60}$$

However, the spin rate cannot exceed a maximum value ω_{\max} related to the tether yield strength τ_{\max} , that is

$$\omega \leq \omega_{\max} \triangleq \sqrt{\frac{2\tau_{\max}}{\rho x_t^2}} \tag{3.61}$$

Using a μm -diameter aluminum tether (Seppänen et al. 2013) with $\rho \simeq 10^{-5} \text{ kg/m}$ and $\tau_{\max} = 0.1275\text{N}$, the allowable pairs $\{\omega, x_t\}$ are shown in Fig. 3.10 when $x_t \in [1, 10]$ km. For example, assuming $\omega = 10$ rph, Fig. 3.10 shows that the maximum value of x_t is about 9 km. In that case, Fig. 3.11 shows the variation of $\{L, \mathcal{P}, \mathcal{M}\}$ with $x_t \in [1, 9]$ km according to Eqs. (3.36)–(3.38).

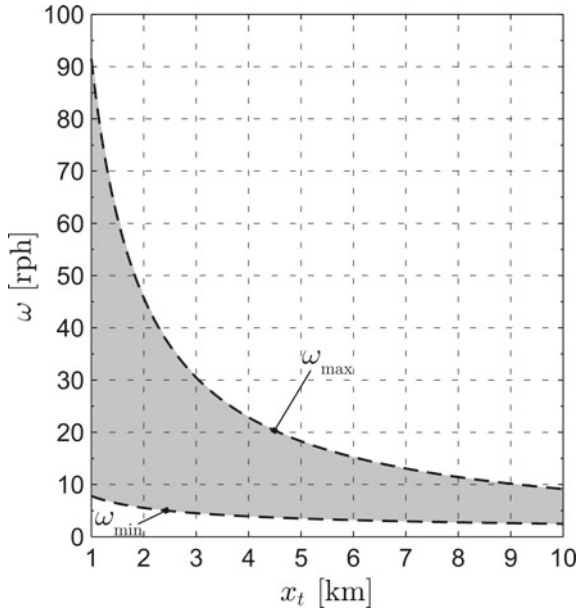


Fig. 3.10 Allowable spin rates as a function of x_t for a μm -diameter aluminum tether (Seppänen et al. 2013). Adapted from Bassetto et al. (2019b)

Even though the coefficients \mathcal{P} and \mathcal{M} must be calculated numerically, they can also be accurately estimated with an analytical approximation. As a result, closed-form expressions of the E-sail propulsive characteristics can be easily obtained, which are very useful for both trajectory simulation and preliminary mission analysis purposes. In fact, the condition $\omega \geq \omega_{\min}$, with ω taken from Eq. (3.60) and ω_{\min} from Eq. (3.58), implies $b_l \leq 0.4$. Observing that

$$f' = \frac{b_l x_t}{x_t + x} \leq b_l \quad (3.62)$$

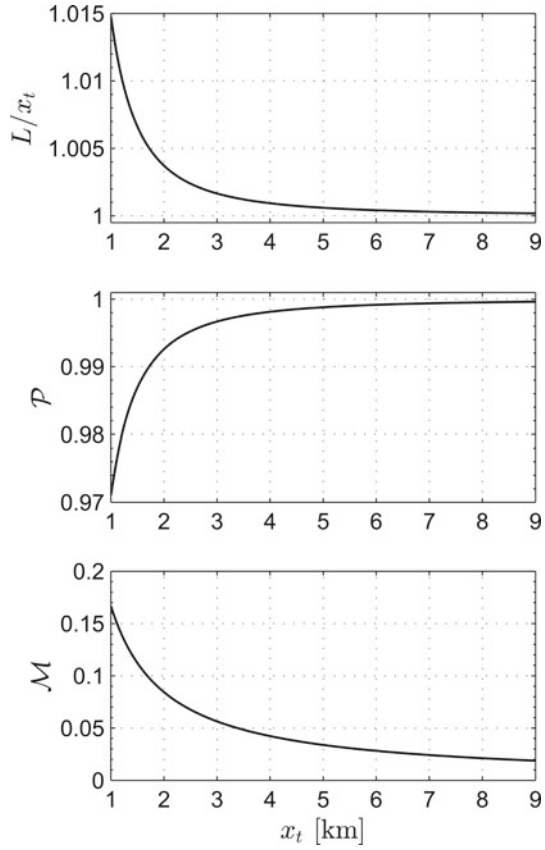
the contribution of $(f')^2$ in Eqs. (3.36)–(3.38) may be neglected so as to obtain

$$L \simeq x_t, \mathcal{P} \simeq 1, \mathcal{M} \simeq \ln(2)b_l \quad (3.63)$$

in accordance with the graphs of Fig. 3.11 and consistently with Eq. (3.53). Substituting now Eqs. (3.63) into Eqs. (3.34) and (3.35), the approximate expression of the thrust vector reduces to Eq. (3.39) (with the magnitude given by Eq. (3.40)), whereas the torque vector becomes

$$\mathbf{T} = \frac{1}{2} \ln(2) b_l N L^2 \sigma u (\hat{\mathbf{k}}_B \times \hat{\mathbf{r}}) \quad (3.64)$$

Fig. 3.11 Variation of $\{L, \mathcal{P}, \mathcal{M}\}$ with x_t when $\omega = 10$ rph. Adapted from Bassetto et al. (2019b)



the magnitude of which is

$$\|T\| = \frac{1}{2} \ln(2) b_l N L^2 \sigma u \sin \alpha_n \tag{3.65}$$

Finally, from Eq. (3.63) the function $f(x)$ describing the shape of each tether can be simplified as

$$f(x) \simeq \frac{2\sigma u}{\rho\omega^2} \ln\left(1 + \frac{x}{L}\right) \text{ with } x \in [0, L]. \tag{3.66}$$

3.3 E-sail Dynamics

3.3.1 Orbital Dynamics

Consider an E-sail-based spacecraft that covers a heliocentric parking orbit of given characteristics. The spacecraft is modeled as a point mass subjected to the gravitational force of the Sun and to the E-sail thrust. The spacecraft state is defined by a set of non-singular modified equinoctial orbital elements (MEOEs) (Walker 1986; Walker et al. 1985) $\{p, f, g, h, k, L\}$, which are related to the classical orbital elements $\{a, e, i, \Omega, \omega, \nu\}$ of the osculating orbit by the following relationships

$$\begin{aligned} p &= a(1 - e^2), \quad f = e \cos(\Omega + \omega), \quad g = e \sin(\Omega + \omega), \\ h &= \tan(i/2) \cos \Omega, \quad k = \tan(i/2) \sin \Omega, \\ L &= \nu + \Omega + \omega \end{aligned} \quad (3.67)$$

where a is the semimajor axis, e is the orbital eccentricity, i is the orbital inclination, Ω is the right ascension of the ascending node, ω is the argument of perihelion, and ν is the true anomaly. The spacecraft heliocentric motion is described by the vectorial differential equation (Betts 2000)

$$\dot{\mathbf{x}} = \mathbb{A}[\mathbf{a}]_{\mathcal{T}_{\text{RTN}}} + \mathbf{c} \quad (3.68)$$

where $\mathbf{x} \triangleq [p \ f \ g \ h \ k \ L]^T$ is the state vector, $\mathbb{A} \in \mathbb{R}^{6 \times 3}$ is the state matrix, defined as

$$\mathbb{A} \triangleq \sqrt{\frac{p}{\mu_{\odot}}} \begin{bmatrix} 0 & \frac{2p}{q} & 0 \\ \sin L & \frac{(q+1)}{q} \cos L + f & \frac{g(k \cos L - h \sin L)}{q} \\ -\cos L & \frac{(q+1)}{q} \sin L + g & \frac{f(h \sin L - k \cos L)}{q} \\ 0 & 0 & \frac{(1+h^2+k^2)}{q} \cos L \\ 0 & 0 & \frac{2q}{(1+h^2+k^2)} \sin L \\ 0 & 0 & \frac{h \sin L - k \cos L}{q} \end{bmatrix} \quad (3.69)$$

where $q \triangleq (1 + f \cos L + g \sin L)$, whereas $\mathbf{c} \in \mathbb{R}^{6 \times 1}$ is given by

$$\mathbf{c} \triangleq \sqrt{\frac{\mu_{\odot}}{p^3}} q^2 [0 \ 0 \ 0 \ 0 \ 0 \ 1]^T \quad (3.70)$$

in which $\mu_{\odot} \simeq 1.327 \times 10^{11} \text{ km}^3/\text{s}^2$ is the Sun's gravitational parameter. In Eq. (3.68), \mathbf{a} is the spacecraft propulsive acceleration vector, the components of which must be expressed in the radial-transverse-normal reference frame \mathcal{T}_{RTN} , centered at the spacecraft center of mass S , of unit vectors

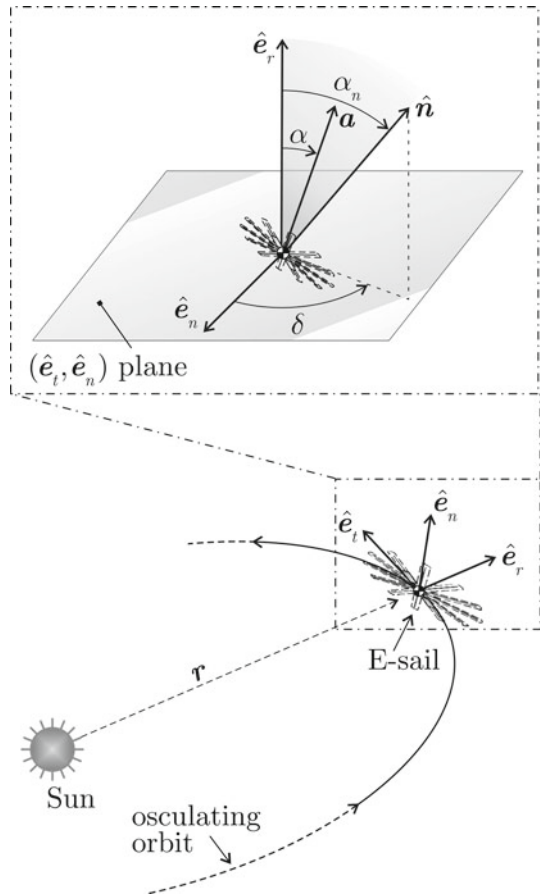
$$\hat{\mathbf{e}}_r \triangleq \hat{\mathbf{r}}, \hat{\mathbf{e}}_n \triangleq \frac{\mathbf{r} \times \mathbf{v}}{\|\mathbf{r} \times \mathbf{v}\|}, \hat{\mathbf{e}}_t \triangleq \hat{\mathbf{e}}_n \times \hat{\mathbf{e}}_r \tag{3.71}$$

where $\hat{\mathbf{r}}$ is the Sun-sail position vector, while \mathbf{v} is the spacecraft absolute velocity; see Fig. 3.12.

Note that Eq. (3.68) is free from singularities since $q \equiv p/r$, being p the semilatus rectum of the spacecraft osculating orbit. The spacecraft propulsive acceleration vector \mathbf{a} can be obtained from Eq. (3.39) by computing the ratio of the thrust vector \mathbf{F} to the spacecraft mass m , viz.

$$\mathbf{a} \triangleq \frac{\mathbf{F}}{m} = \frac{a_c}{2} [\hat{\mathbf{e}}_r + (\hat{\mathbf{e}}_r \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}}] \tag{3.72}$$

Fig. 3.12 Reference frame and E-sail characteristic angles. Adapted from Bassetto et al. (2019c)



where, consistently with the nomenclature adopted in Sect. 3.2, $\hat{\mathbf{n}}$ is the unit vector perpendicular to the E-sail nominal plane pointing in the opposite direction to the Sun, while

$$a_c \triangleq \frac{N L \sigma_{\oplus} u}{m} \quad (3.73)$$

is the characteristic acceleration of the E-sail, that is, the maximum feasible propulsive acceleration when $r = r_{\oplus}$. In particular, such a performance level is achieved when the sail attitude is Sun facing, that is, when $\alpha_n = 0$. Indeed, when $\hat{\mathbf{n}} \equiv \hat{\mathbf{r}}$ and $r = r_{\oplus}$, then

$$\mathbf{a} = a_c \hat{\mathbf{e}}_r \quad (3.74)$$

with reference to Fig. 3.12, when $\alpha_n \neq 0$, \mathbf{a} belongs to the plane $\{\hat{\mathbf{e}}_r, \hat{\mathbf{n}}\}$ and the angle α between $\hat{\mathbf{e}}_r$ and \mathbf{a} , referred to as thrust cone angle, is given by

$$\alpha = \arccos\left(\frac{1 + \cos^2 \alpha_n}{\sqrt{1 + 3 \cos^2 \alpha_n}}\right) \quad (3.75)$$

the maximum value of which, reached when $\alpha_n \simeq 54.74^\circ$, is approximately equal to 19.47° . It is also interesting to evaluate the dependence of the propulsive acceleration magnitude on α_n . This can be done by computing the dimensionless propulsive acceleration γ , defined as

$$\gamma \triangleq \frac{\|\mathbf{a}\| r}{a_c r_{\oplus}} = \frac{\sqrt{1 + 3 \cos^2 \alpha_n}}{2} \quad (3.76)$$

In this context, Fig. 3.13 shows γ and α as a function of α_n . Finally, the components of \mathbf{a} in \mathcal{T}_{RTN} are

$$[\mathbf{a}]_{\mathcal{T}_{\text{RTN}}} = \begin{bmatrix} a_r \\ a_t \\ a_n \end{bmatrix} \quad (3.77)$$

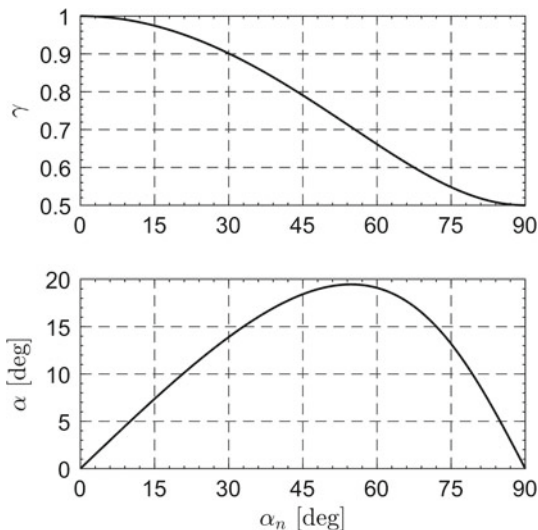
with

$$a_r \triangleq \mathbf{a} \cdot \hat{\mathbf{e}}_r = \frac{a_c}{2} \left(\frac{r_{\oplus}}{r}\right) (1 + \cos^2 \alpha_n) \quad (3.78)$$

$$a_t \triangleq \mathbf{a} \cdot \hat{\mathbf{e}}_t = -\frac{a_c}{2} \left(\frac{r_{\oplus}}{r}\right) \sin \alpha_n \cos \alpha_n \sin \delta \quad (3.79)$$

$$a_n \triangleq \mathbf{a} \cdot \hat{\mathbf{e}}_n = \frac{a_c}{2} \left(\frac{r_{\oplus}}{r}\right) \sin \alpha_n \cos \alpha_n \cos \delta \quad (3.80)$$

Fig. 3.13 Dimensionless propulsive acceleration and cone angle as a function of α_n . Adapted from Bassetto et al. (2019c)



where $\delta \in [0, 2\pi]$ rad is the angle (measured counterclockwise) between the projection of \hat{n} on the local horizontal plane (i.e., the plane (\hat{e}_t, \hat{e}_n) perpendicular to the Sun-spacecraft line) and \hat{e}_n ; see Fig. 3.12.

3.3.2 Attitude Dynamics

This section deals with the problem of analyzing the attitude dynamics of a spinning E-sail-based spacecraft. Indeed, when the spacecraft attitude is perturbed from the Sun-facing configuration, that is, when the z_B -axis slightly differs from the Sun-spacecraft direction ($\hat{k}_B \neq \hat{r}$), the spacecraft experiences a nonzero propulsive torque, which may be described using the approximate expression (3.64). The attitude motion of the spacecraft is here studied without any type of control. Later in this chapter we will address the problem of investigating a strategy capable of maintaining and controlling the E-sail attitude by generating a suitable control torque. In particular, it will be shown that the attitude control can be performed through a suitable modulation of the electrical voltage of each tether.

Under the assumption that the spacecraft (including the E-sail) behaves like a rigid body, the dimensionless coefficient \mathcal{M} (see Eq. (3.37)) and the spacecraft inertia tensor are both constant. The effects of the torque due to the tether inflection on the spacecraft dynamics can therefore be analyzed by a numerical integration of the classical Euler equations. To that end, the components of \mathbf{T} in the body reference frame are written as a function of the three Euler angles $\{\phi, \theta, \psi\}$, which define the orientation of \mathcal{T}_B with respect to an inertial reference frame $\mathcal{T}_I(S; x_I, y_I, z_I)$ of unit vectors $\{\hat{i}_I, \hat{j}_I, \hat{k}_I\}$, where $\hat{k}_I \equiv \hat{r}$, while \hat{i}_I points towards a fixed direction in

space. Using a rotational sequence $3(\psi) \rightarrow 1(\phi) \rightarrow 2(\theta)$ (Wertz 1978) to describe the orientation of \mathcal{T}_B relative to \mathcal{T}_I , the components of $\hat{\mathbf{r}}$ in the body reference frame are

$$[\mathbf{r}]_{\mathcal{T}_B} = \begin{bmatrix} -\cos \phi \sin \theta \\ \sin \phi \\ \cos \phi \cos \theta \end{bmatrix} \quad (3.81)$$

whereas the components of the torque vector in \mathcal{T}_B are given by

$$[\mathbf{T}]_{\mathcal{T}_B} = -\frac{1}{2} \ln(2) b_l N L^2 \sigma u \begin{bmatrix} \sin \phi \\ \cos \phi \sin \theta \\ 0 \end{bmatrix} \quad (3.82)$$

In this context, Sect. 3.2.1 shows that the sail attitude can also be expressed as a function of the pitch angle α_n and the clock angle δ_n ; see Eqs. (3.11) and (3.12). In particular, α_n and δ_n are related to the Euler angles through the following equations

$$\cos \alpha_n = \cos \phi \cos \theta \quad (3.83)$$

$$\sin \alpha_n \sin \delta_n = \sin \phi \quad (3.84)$$

$$\sin \alpha_n \cos \delta_n = -\cos \phi \sin \theta \quad (3.85)$$

With the assumed rotational sequence $3(\psi) \rightarrow 1(\phi) \rightarrow 2(\theta)$, the kinematic equations of a rigid E-sail-based spacecraft are

$$\dot{\phi} = \Omega_x \cos \theta + \Omega_z \sin \theta \quad (3.86)$$

$$\dot{\theta} = \Omega_y - (\Omega_z \cos \theta - \Omega_x \sin \theta) \tan \phi \quad (3.87)$$

$$\dot{\psi} = (\Omega_z \cos \theta - \Omega_x \sin \theta) \sec \phi \quad (3.88)$$

where $\{\Omega_x, \Omega_y, \Omega_z\}$ are the components of the spacecraft angular velocity $\boldsymbol{\Omega}$ in \mathcal{T}_B . Although, in principle, the spacecraft attitude dynamics is affected by the rotation of \mathcal{T}_B with respect to the Sun–sail line due to the vehicle orbital motion, this effect is negligible because the spacecraft mean motion (about 1.99×10^{-7} rad/s for a heliocentric circular orbit of radius equal to 1 au) is usually several orders of magnitude smaller than the typical values of $\{\Omega_x, \Omega_y, \Omega_z\}$. This is the reason why the orbital and the attitude motions may be studied separately. Accordingly, the Euler equations are

$$\dot{\Omega}_x = -h\Omega_z^2 \sin \phi + \lambda\Omega_y\Omega_z \quad (3.89)$$

$$\dot{\Omega}_y = -h\Omega_z^2 \sin \theta \cos \phi - \lambda\Omega_x\Omega_z \quad (3.90)$$

$$\dot{\Omega}_z = 0 \quad \text{from which} \quad \Omega_z = \omega \quad (3.91)$$

with

$$h \triangleq \frac{\ln(2)NL(\sigma u)^2}{\rho\omega^4 I_t}, \lambda \triangleq \frac{I_t - I_z}{I_t} \quad (3.92)$$

where I_t and I_z are the spacecraft longitudinal and transverse moments of inertia, respectively. Note that h is a dimensionless parameter that depends on the tether shape through ω , see Eq. (3.66), and on r and V through the design parameter σ ; see Eq. (3.8).

3.3.2.1 Linearized Attitude Dynamics

Assuming α_n to be sufficiently small, which implies $\{\phi, \theta\} \ll 1$, Eq. (3.81) reduces to

$$[\mathbf{r}]_{\mathcal{T}_B} = \begin{bmatrix} -\theta \\ \phi \\ 1 \end{bmatrix} \quad (3.93)$$

while, using Eqs. (3.64) and (3.93), the components of the propulsive torque vector in \mathcal{T}_B become

$$[\mathbf{T}]_{\mathcal{T}_B} = -\frac{1}{2} \ln(2) b_l N L^2 \sigma \begin{bmatrix} \phi \\ \theta \\ 0 \end{bmatrix} \quad (3.94)$$

Accordingly, the linearized kinematic equations are Longuski et al. (2005)

$$\dot{\phi} = \Omega_x + \theta\Omega_z \quad (3.95)$$

$$\dot{\theta} = \Omega_y - \phi(\Omega_z - \theta\Omega_x) \quad (3.96)$$

$$\dot{\psi} = \Omega_z - \theta\Omega_x \quad (3.97)$$

which can be further simplified observing that $\Omega_z \gg \theta\Omega_x$, viz.

$$\dot{\phi} = \Omega_x + \theta \Omega_z \quad (3.98)$$

$$\dot{\theta} = \Omega_y - \phi \Omega_z \quad (3.99)$$

$$\dot{\psi} = \Omega_z \quad (3.100)$$

Bearing in mind Eq. (3.91), the linearized Euler equations for the axially symmetric E-sail are given by

$$\dot{\Omega}_x = -h\omega^2\phi + \lambda\omega\Omega_y \quad (3.101)$$

$$\dot{\Omega}_y = -h\omega^2\theta - \lambda\omega\Omega_x \quad (3.102)$$

From Eq. (3.91), the value of Ω_z is constant and is equivalent to the nominal spin rate ω . Therefore, Eq. (3.100) implies that $\psi = \psi_0 + \omega t$, being ψ_0 the initial value of ψ and t the time. Using the dimensionless time $\tilde{t} \triangleq t\omega$ and angular velocities $\{\tilde{\Omega}_x, \tilde{\Omega}_y\}$, defined as

$$\tilde{\Omega}_x \triangleq \frac{\Omega_x}{\omega}, \quad \tilde{\Omega}_y \triangleq \frac{\Omega_y}{\omega} \quad (3.103)$$

Therefore, Eqs. (3.98), (3.99), (3.101) and (3.102) can be equivalently written in matrix form as

$$\frac{d\mathbf{x}}{d\tilde{t}} = \mathbb{B}\mathbf{x} \quad \text{with} \quad \mathbb{B} \triangleq \begin{bmatrix} 0 & \lambda & -h & 0 \\ -\lambda & 0 & 0 & -h \\ 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \quad (3.104)$$

where $\mathbf{x} \triangleq [\tilde{\Omega}_x \ \tilde{\Omega}_y \ \phi \ \theta]^T$ is the dimensionless state vector. Using the Routh–Hurwitz stability criterion and taking into account that $h > 0$ and $\lambda < 0$, it is possible to verify that the linear differential system of Eq. (3.104) is stable and is characterized by a pair of imaginary poles. Therefore, the presence of a torque due to a pitch angle α_n different from zero does not affect the stability of the E-sail linearized attitude motion.

3.3.2.2 Numerical Simulations

The stability of the nonlinear E-sail-based spacecraft attitude dynamics has been investigated by an extensive simulation campaign for some E-sail configurations. In particular, Table 3.1 reports the characteristics of five possible E-sail arrangements

Table 3.1 Major specifications of five possible E-sail arrangements

E-sail	L (km)	N	ω (rph)
1	2	500	43.46
2	4	250	21.73
3	6	168	14.49
4	8	126	10.86
5	10	100	8.69

characterized by $(NL) = 1000$ km, $\omega = 0.95\omega_{\max}$ [see Eq. (3.61)], and $V = 20$ kV (Seppänen et al. 2013).

Using the configurations 1 and 5 as exemplary cases, Eqs. (3.86)–(3.91) have been numerically integrated with initial conditions

$$\{\alpha_{\eta_0}, \delta_{\eta_0}\} \triangleq \{10, 90\}^\circ, \quad \{\Omega_{x_0}, \Omega_{y_0}, \Omega_{z_0}\} \triangleq \{0, 0, \omega\} \quad (3.105)$$

and assuming $\lambda = -0.5$ and $I_t = 1000$ kg m². The simulation results are illustrated in Figs. 3.14, 3.15, 3.16 and 3.17. The pitch angle shows a periodic time variation (the maximum value of which coincides with α_{n_0}) due to a nutation motion of the spacecraft symmetry axis; see Figs. 3.14 and 3.15. Both the amplitude and frequency of the nutation oscillations are higher for an E-sail with longer tethers. Moreover, introducing the components of the unit vector $\hat{\mathbf{k}}_B$ in the inertial reference frame \mathcal{T}_I (that is, $[\hat{\mathbf{k}}_B]_{\mathcal{T}_I} \triangleq [k_x, k_y, k_z]^T$), Figs. 3.14 and 3.15 show that the E-sail symmetry axis is subjected to a precession motion, too. Again, the frequency of oscillation is higher for an E-sail with longer tethers.

Fig. 3.14 Time variation of α_n , k_x , and k_y for E-sail 1 defined in Table 3.1. Adapted from Bassetto et al. (2019a)

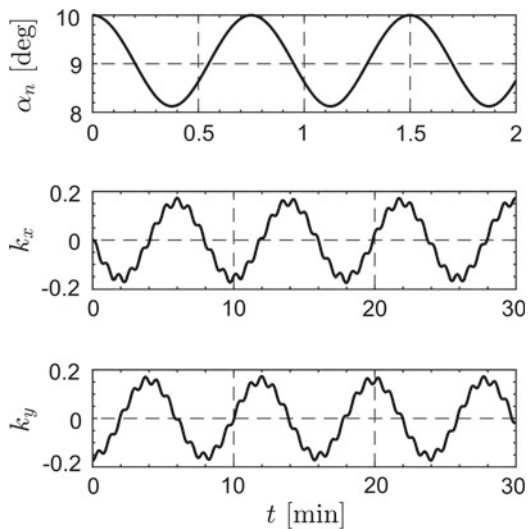


Fig. 3.15 Time variation of α_n , k_x , and k_y for E-sail 5 defined in Table 3.1. Adapted from Bassetto et al. (2019a)

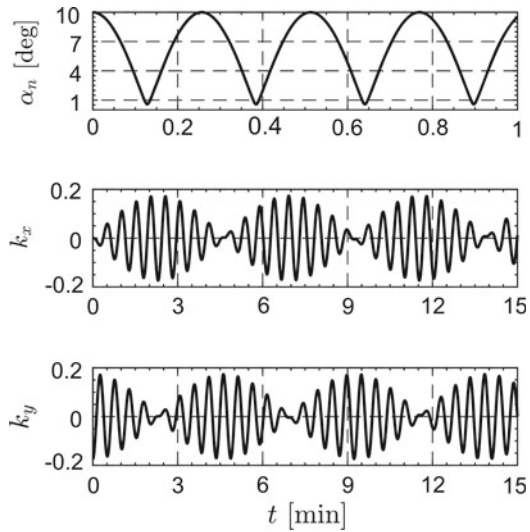
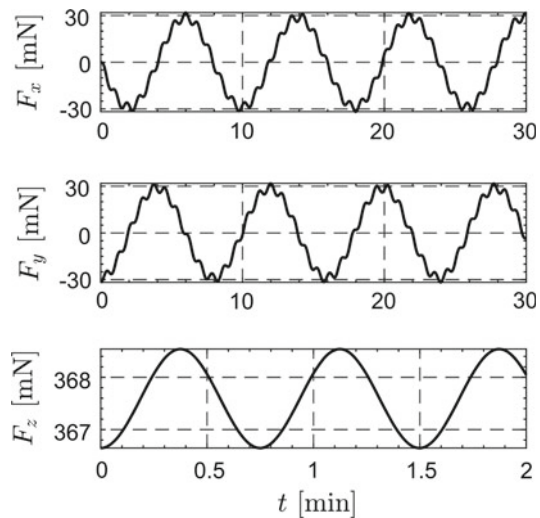


Fig. 3.16 Time variation of $[F]_{\mathcal{T}_I}$ for E-sail 1 defined in Table 3.1. Adapted from Bassetto et al. (2019a)

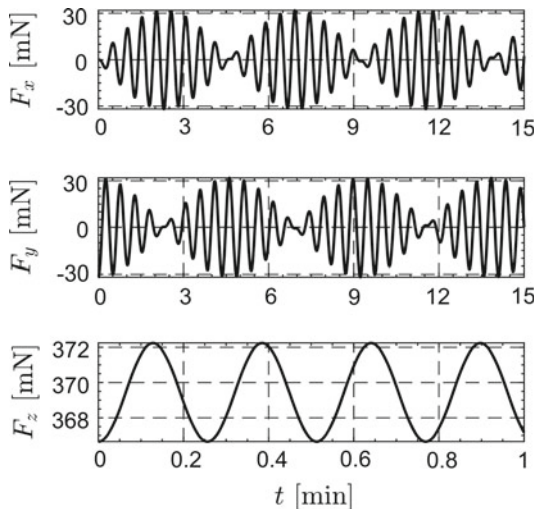


The long-term propulsive effect due the torque acting on the spacecraft is better appreciated by representing the time evolution of the thrust vector components in the inertial frame. This is possible using the following equations

$$F_x = \frac{1}{2} N L \sigma u (3\mathcal{P} - 2)[c\varphi c\theta (s\theta c\psi + s\varphi c\theta s\psi)] \quad (3.106)$$

$$F_y = \frac{1}{2} N L \sigma u (3\mathcal{P} - 2)[c\varphi c\theta (s\theta s\psi - s\varphi c\theta c\psi)] \quad (3.107)$$

Fig. 3.17 Time variation of $[F]_{\mathcal{T}_I}$ for E-sail 5 defined in Table 3.1. Adapted from Bassetto et al. (2019a)



$$F_z = \frac{1}{2} N L \sigma u [2 - \mathcal{P} + (3\mathcal{P} - 2)(c\varphi c\theta)^2] \quad (3.108)$$

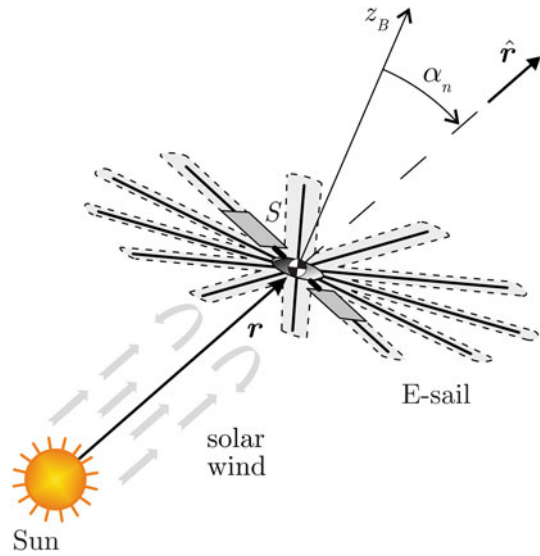
where $s \triangleq \sin$, $c \triangleq \cos$, while $\{F_x, F_y, F_z\}$ are the components of \mathbf{F} in \mathcal{T}_I . Both Figs. 3.16 and 3.17 show that the long-term thrust direction is radial, with small-amplitude and short-period oscillations due to the nutation motion. An active control system is therefore necessary to counteract the effect of the external torque, allowing the spacecraft to give a long-term nonzero transverse thrust.

3.4 Attitude Maintenance and Control

The results presented so far show that the torque provided by an E-sail is zero when its attitude is Sun facing. In that configuration, Sect. 3.2.3 proves that the equilibrium shape of each tether is well approximated by a natural logarithmic arc when the sail spin rate is sufficiently high, and the electrical voltage is uniform. The more complex case of an E-sail that generates a transverse thrust (and, accordingly, a nonzero torque vector) is discussed in Sect. 3.2.2 under the basic assumption that the E-sail maintains a rigid shape independent of the pitch angle α_n ; see Fig. 3.18.

Such an assumption is reasonable if the pitch angle is small and, therefore, the tether arrangement is not far from the equilibrium shape found in a pure Sun-facing configuration. In fact, a transverse thrust is necessary for orbital maneuvers since it allows the orbit angular momentum to be varied. A transverse thrust component can be obtained by inclining the sail nominal plane with respect to the local radial direction. This problem has already been addressed by Toivanen et al. (2015), according to whom the sail attitude can be controlled by modulating the tether voltage

Fig. 3.18 E-sail conceptual sketch. Adapted from Bassetto et al. (2019a)



synchronously with the sail rotation. However, tether bending causes the onset of a disturbance torque, which induces a perturbation on the orientation of the spacecraft spin axis. The latter experiences an undamped precession-nutation motion, which tends to align the thrust vector along the Sun–spacecraft direction. In particular, the numerical integration of Euler’s attitude equations (see Sect. 3.3.2) shows that the amplitude and the frequency of these two harmonic motions are affected by the initial conditions, the spacecraft inertia tensor, the number of tethers, and the electrical voltage. Because the external torque causes the long period thrust to be oriented in the radial direction, this perturbative effect must be removed.

The aim of this section is first to discuss a simple control law that counteracts the generation of the external torque by suitably adjusting the tether electrical voltage. Janhunen and Toivanen (2018) have recently presented an algorithm to control the sail attitude. The effectiveness of such an algorithm is confirmed by a full end-to-end simulation in which the tethers are modeled as elastic wires, while the solar wind characteristics are taken from historical satellite data. The same problem is here addressed in an analytical way. In essence, the idea is to ideally divide the sail plane into two symmetrical parts delimited by the straight line passing through the torque vector, and to assign a given value of electrical voltage to each part. Note that, from an operational perspective, the crossing of such a control plane may be detected by simply measuring the current value of δ_n , as is better illustrated in the next section. The value of the two control voltages is chosen such as to remove the external torque and, at the same time, to maintain the nominal thrust vector. The obtained solution shows that the required variation of electrical voltage is some orders of magnitude smaller than its reference value. Then, we propose a strategy for changing the E-sail attitude. Previous work on this subject dates back to the work by Janhunen (2013), who discussed the problem of creating and modifying the E-sail spin rate with small

photonic sails applied to the tips of the main tethers. Later, Toivanen and Janhunen (2013) showed that the sail attitude can be controlled by modulating the voltage of each individual tether to produce a torque for thrust vectoring. In this case, the control is obtained by modulating the electrical voltage of each individual tether as a function of the time and the sail attitude, assuming that the spacecraft maintains an axially symmetric shape. Again, the E-sail is virtually divided into two parts, each one characterized by a precise level of electrical voltage. To that end, a control plane (perpendicular to the sail nominal plane) is defined to locate where the tether voltage must change. The amplitude of the voltage modulation is obtained in an indirect way by enforcing the sail attitude to track a desired time history. The control strategy is validated through numerical simulations, which show that the required torque can be generated with a small modulation of the electrical voltage with respect to its reference value.

3.4.1 Attitude Maintenance for Tracking Purposes

The numerical simulations shown in Sect. 3.3.2 prove that the torque acting on the E-sail tends to align its spin axis with the radial direction. Therefore, such a torque must be removed so that the E-sail can generate a long-term nonzero transverse thrust. This is possible by properly adjusting the tether voltage, according to a control law that is now discussed. In the following analysis the assumption is made of small pitch angles (that is, $\alpha_n \leq 10^\circ$), which implies that the tethers maintain the equilibrium shape found in the Sun-facing configuration; see Sect. 3.3.2. In that case, the tether shape is accurately described by a natural logarithmic arc, provided the E-sail spins at a sufficiently high rate, in accordance with Eq. (3.58). The second approximation here introduced is that the E-sail shape does not change when the electrical voltage is slightly modified from its nominal value.

A simple control strategy is now proposed, which consists in changing the design parameter σ with respect to its nominal value by changing the tether electrical voltage; see Eq. (3.8). Using the results presented in Sect. 3.2.4, the torque generated by an axially symmetric E-sail with a uniform tether electrical voltage is given by the following equation

$$\mathbf{T}_d = \frac{1}{2} \ln(2) b_l N L^2 \sigma u \left(\hat{\mathbf{k}}_B \times \hat{\mathbf{r}} \right) \quad (3.109)$$

where the subscript d stands for “disturbance”, being \mathbf{T}_d an unwanted torque. From Sect. 3.2.1, the expression of \mathbf{T}_d can also be written in the body reference frame as

$$\mathbf{T}_d = \mathcal{E} \hat{\mathbf{i}}_B + \mathcal{F} \hat{\mathbf{j}}_B + \mathcal{G} \hat{\mathbf{k}}_B \quad (3.110)$$

where $\{\mathcal{E}, \mathcal{F}, \mathcal{G}\}$ are given by Eq. (3.32). Therefore, the problem is to look for a control law that can cancel the three components $\{\mathcal{E}, \mathcal{F}, \mathcal{G}\}$ of the resultant torque. Note that $\mathcal{G} = 0$ is a sufficient condition to ensure a constant spin rate because of the assumption of axial symmetry. A reduction of the spin rate is undesirable because it would produce a greater tether bending, thus reducing the propulsive E-sail performance, and increasing the magnitude of the disturbance torque. On the other hand, an increase of ω must be avoided to prevent the tether yield strength to be exceeded.

In order to face the problem, the tethers are first divided into two subsets by the plane (referred to as control plane) passing through T_d and $\hat{\mathbf{k}}_B$, as is schematically illustrated in Fig. 3.19 by the dotted line. All the tethers in the same subset have the same electrical voltage, but those belonging to the half-plane containing the projection of the position vector $\hat{\mathbf{r}}$ on the sail nominal plane ($\hat{\mathbf{r}}_{\parallel}$) have a higher potential (σ_{\max}) than those belonging to the other half-plane (σ_{\min}). In particular, the value of σ_k depends on the sign of the scalar product $\hat{\mathbf{i}}_k \cdot \hat{\mathbf{r}}$, where $\hat{\mathbf{i}}_k$ and $\hat{\mathbf{r}}$ are given by Eqs. (3.2) and (3.10), respectively. The sign of $\hat{\mathbf{i}}_k \cdot \hat{\mathbf{r}}$ may be conveniently rewritten considering that $\sin \alpha_n \geq 0$, viz.

$$\text{sign}\{\hat{\mathbf{i}}_k \cdot \hat{\mathbf{r}}\} = \text{sign}\{\cos(\delta_n - \zeta_k)\} \quad (3.111)$$

where ζ_k is given by Eq. (3.33). Therefore, $\sigma_k = \sigma_{\max}$ when $\cos(\delta_n - \zeta_k) > 0$, whereas $\sigma_k = \sigma_{\min}$ when $\cos(\delta_n - \zeta_k) < 0$. Note that σ_k must be switched from σ_{\min} to σ_{\max} (or, vice versa, from σ_{\max} to σ_{\min}) when the k -th tether crosses the control plane, that is, when $\cos(\delta_n - \zeta_k) = 0$. Note also that the overall thrust does not change if $\Delta\sigma \triangleq (\sigma_{\max} - \sigma_{\min}) = 2(\sigma - \sigma_{\min})$. In that case, the E-sail can maintain a fixed attitude with respect to the Sun-spacecraft line without affecting the total propulsive acceleration.

With reference to the configuration of Fig. 3.19, the torque balance requires the following equilibrium conditions to be met

$$\mathcal{E} \triangleq \int_0^{x_r|_{\sigma_{\min}}} \sum_{k=0}^{N/2-1} d\mathcal{E}_k|_{\sigma_{\min}} + \int_0^{x_r|_{\sigma_{\max}}} \sum_{k=N/2}^{N-1} d\mathcal{E}_k|_{\sigma_{\max}} = 0 \quad (3.112)$$

$$\mathcal{F} \triangleq \int_0^{x_r|_{\sigma_{\min}}} \sum_{k=0}^{N/2-1} d\mathcal{F}_k|_{\sigma_{\min}} + \int_0^{x_r|_{\sigma_{\max}}} \sum_{k=N/2}^{N-1} d\mathcal{F}_k|_{\sigma_{\max}} = 0 \quad (3.113)$$

$$\mathcal{G} \triangleq \int_0^{x_r|_{\sigma_{\min}}} \sum_{k=0}^{N/2-1} d\mathcal{G}_k|_{\sigma_{\min}} + \int_0^{x_r|_{\sigma_{\max}}} \sum_{k=N/2}^{N-1} d\mathcal{G}_k|_{\sigma_{\max}} = 0 \quad (3.114)$$

Fig. 3.19 Sketch of a generic E-sail for attitude maintenance determination. Adapted from Bassetto et al. (2019a)

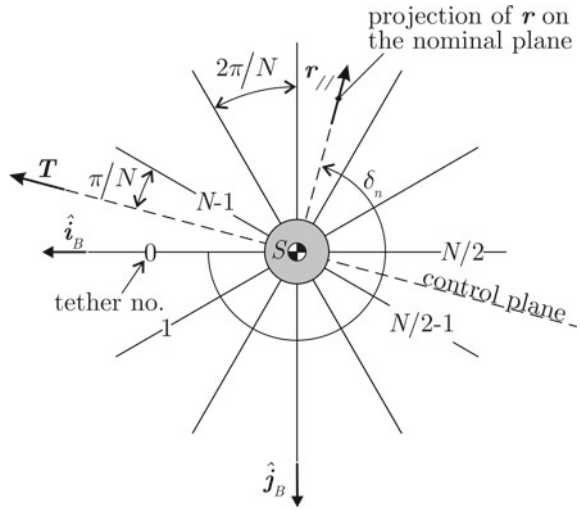


Table 3.2 Summations of trigonometric functions in Eqs. (3.112)–(3.114)

	$\sum_{k=0}^{N/2-1} \dots$	$\sum_{k=N/2}^{N-1} \dots$
$\sin \zeta_k$	$\cot(\pi/N)$	$-\cot(\pi/N)$
$\cos \zeta_k$	1	-1
$\sin \zeta_k \cos \zeta_k$	0	0
$\sin^2 \zeta_k$	$N/4$	$N/4$
$\cos^2 \zeta_k$	$N/4$	$N/4$

where the whole expressions of $d\mathcal{E}_k$, $d\mathcal{F}_k$, and $d\mathcal{G}_k$ are reported in Sect. 3.2.1; see Eqs. (3.26)–(3.28). With the aid of Table 3.2, the summations in Eqs. (3.112)–(3.114) give the following results

$$\sum_{k=0}^{N/2-1} d\mathcal{E}_k |_{\sigma_{\min}} = \left[\sigma_{\min} u \cos \alpha_n \cot\left(\frac{\pi}{N}\right) \frac{x + f_1 f_1'}{\sqrt{1 + (f_1')^2}} - \frac{N}{4} \sigma_{\min} u \sin \alpha_n \sin \delta_n \frac{f_1 + x f_1' + 2 f_1 (f_1')^2}{\sqrt{1 + (f_1')^2}} \right] dx \quad (3.115)$$

$$\sum_{k=N/2}^{N-1} d\mathcal{E}_k |_{\sigma_{\max}} = \left[-\sigma_{\max} u \cos \alpha_n \cot\left(\frac{\pi}{N}\right) \frac{x + f_2 f_2'}{\sqrt{1 + (f_2')^2}} - \frac{N}{4} \sigma_{\max} u \sin \alpha_n \sin \delta_n \frac{f_2 + x f_2' + 2 f_2 (f_2')^2}{\sqrt{1 + (f_2')^2}} \right] dx \quad (3.116)$$

$$\sum_{k=0}^{N/2-1} d\mathcal{F}_k |_{\sigma_{\min}} = \left[-\sigma_{\min} u \cos \alpha_n \frac{x + f_1 f_1'}{\sqrt{1 + (f_1')^2}} + \frac{N}{4} \sigma_{\min} u \sin \alpha_n \cos \delta_n \frac{f_1 + x f_1' + 2 f_1 (f_1')^2}{\sqrt{1 + (f_1')^2}} \right] dx \quad (3.117)$$

$$\sum_{k=N/2}^{N-1} d\mathcal{F}_k |_{\sigma_{\max}} = \left[\sigma_{\max} u \cos \alpha_n \frac{x + f_2 f_2'}{\sqrt{1 + (f_2')^2}} + \frac{N}{4} \sigma_{\max} u \sin \alpha_n \cos \delta_n \frac{f_2 + x f_2' + 2 f_2 (f_2')^2}{\sqrt{1 + (f_2')^2}} \right] dx \quad (3.118)$$

$$\sum_{k=0}^{N/2-1} d\mathcal{G}_k |_{\sigma_{\min}} = \sigma_{\min} u x \sin \alpha_n \left[\sin \delta_n - \cos(\delta_n) \cot\left(\frac{\pi}{N}\right) \right] \sqrt{1 + (f_1')^2} dx \quad (3.119)$$

$$\sum_{k=N/2}^{N-1} d\mathcal{G}_k |_{\sigma_{\max}} = \sigma_{\max} u x \sin \alpha_n \left[-\sin \delta_n + \cos(\delta_n) \cot\left(\frac{\pi}{N}\right) \right] \sqrt{1 + (f_2')^2} dx \quad (3.120)$$

where, with reference to Fig. 3.19, the clock angle is given by

$$\delta_n = \frac{3\pi}{2} - \frac{\pi}{N} \quad (3.121)$$

whereas f_1 (or f_2) describes shape of the k -th tether in the plane $(\hat{\mathbf{i}}_k, \hat{\mathbf{k}}_{\mathbf{B}})$ for $k = \{0, 1, \dots, N/2-1\}$ (or $k = \{N/2, N/2+1, \dots, N-1\}$). Note that $f_1 \triangleq f(\sigma = \sigma_{\min})$ and $f_2 \triangleq f(\sigma = \sigma_{\max})$. The assumption that the E-sail shape does not change when the electrical voltage is slightly modified from its nominal value entails that $f_1 = f_2 = f$, that is, the E-sail preserves its nominal shape. The latter is given by Eq. (3.57) when $x_r = 0$, which requires the spacecraft spin rate to be sufficiently high and, as such, the tether slope to be sufficiently small, or

$$\max(f')^2 = b_l^2 \simeq 0 \quad (3.122)$$

where b_l is given by Eq. (3.60). Moreover, consistently with Eq. (3.63), the assumption is made that $x_t |_{\sigma_{\min}} = x_t |_{\sigma_{\max}} \simeq L$.

Bearing in mind that $\sigma_{\max} = (2\sigma - \sigma_{\min})$, Eq. (3.114) is satisfied for any value of σ_{\min} , whereas Eqs. (3.112) and (3.113) are equivalent to each other (i.e., they have the same solution). In particular, from Eq. (3.121) it is possible to verify that

$$\frac{\sin \delta_n}{\cot(\pi/N)} \equiv \cos \delta_n = -\sin\left(\frac{\pi}{N}\right) \tag{3.123}$$

The three integrals obtained by substituting Eq. (3.115) into Eq. (3.112) reduce to

$$\int_0^L (x + ff') \, dx = \frac{L^2}{2} \tag{3.124}$$

$$\int_0^L (f + xf') \, dx = L^2 b_l \ln(2) \tag{3.125}$$

$$\int_0^L (xf) \, dx = \frac{b_l L^3}{4} \tag{3.126}$$

and the solution of Eq. (3.112) is

$$\sigma_{\min} = \sigma \left[1 - \frac{N \sin(\pi/N) b_l \ln(2)}{2} \tan \alpha_n \right] \tag{3.127}$$

which implies that

$$(\sigma - \sigma_{\min}) \propto \tan \alpha_n \tag{3.128}$$

Instead, Toivanen and Janhunen (2017) have shown that the voltage modulation necessary for removing the torque is proportional to a function of α_n , that is

$$(\sigma - \sigma_{\min}) \propto \sin \alpha_n (1 + \sin \alpha_n) \tag{3.129}$$

Therefore, when α_n is sufficiently small, the right-hand side of both Eqs. (3.128) and (3.129) may be approximated with α_n . As a result, Eq. (3.127) is in agreement with Toivanen and Janhunen (2017) as long as $\alpha_n \ll 1$.

For exemplary purposes, consider the E-sail configurations 1–5, the characteristics of which are defined in Table 3.1. Figure 3.20 shows the required σ_{\min}/σ as a function of α_n .

Note that σ_{\min}/σ is a monotonic decreasing function of α_n and its value reduces as long as an E-sail with longer tethers is used. This is an expected result, since the external torque is higher when, for a given thrust magnitude, the tethers have a greater length. In particular, for a fixed value of the product NL , the external torque is proportional to L^2 , being $\omega_{\max} \propto L^{-1}$; see Eqs. (3.60), (3.61), and (3.64). Note also that the required σ_{\min} and σ_{\max} are not much different from their nominal

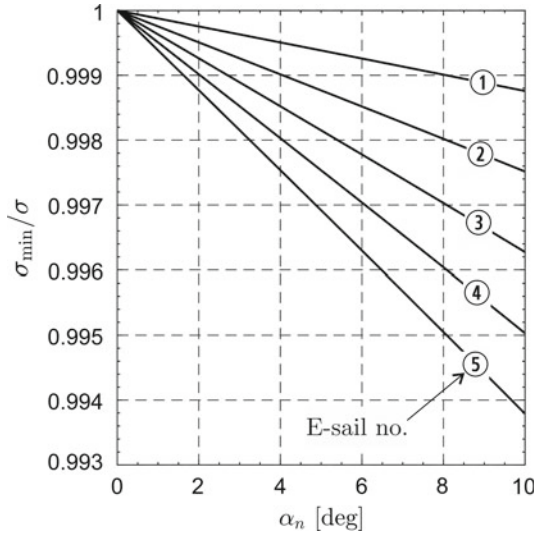


Fig. 3.20 Required σ_{\min}/σ as a function of α_n . See Table 3.1 for E-sail characteristics. Adapted from Bassetto et al. (2019a)

value σ . Hence, the previously enforced hypothesis that $f_1 = f_2 = f$ is realistic. For example, for a 5-type E-sail, and assuming $\alpha_n = 10^\circ$, $\sigma_{\min} \simeq 0.9938\sigma$ and $\sigma_{\max} \simeq 1.0062\sigma$. This is an interesting result because, as long as $\Delta\sigma/\sigma \ll 1$, the assumption of axial symmetry can be retained even when the electrical voltage of the two half-planes is different. In that case, it is convenient to define the required value of $\Delta\sigma$ per unit of generated torque, that is

$$\kappa \triangleq \frac{\Delta\sigma}{T_d} = \frac{2 \sin(\pi/N)}{uL^2 \cos \alpha_n} \tag{3.130}$$

where

$$T_d \triangleq ||T_d|| = \frac{1}{2} \ln(2) b_l N L^2 \sigma u \sin \alpha_n \tag{3.131}$$

Note that κ depends on the E-sail geometrical characteristics and on the current sail attitude. Since κ is independent of the E-sail shape, a control torque obtained by individually modulating the tether voltage can also be generated in the (particular) case of a flat E-sail. The next section will use Eq. (3.130) to derive a control law able to change the sail attitude and, at the same time, to remove the disturbance torque due to the tether bending.

3.4.2 Attitude Control for Thrust Vectoring

The creation of a transverse thrust component requires the sail nominal plane (x_B, y_B) to be oriented such as to reach a suitable value of the pitch angle α_n ; see Fig. 3.21. The aim of this section is to show how the spacecraft attitude may be adjusted with a control torque T_c generated through a suitable modulation of the (generic) tether electric potential. To that end, recall that the E-sail is assumed to maintain an axially symmetric shape, which is independent of both the pitch angle α_n and the tether voltage V . The control torque T_c is the result of two vectors: (1) the attitude variation torque (T_a), which ensures the required change of α_n to be obtained, and (2) the vector $-T_d$ (with T_d given by Eq. (3.109)), which removes the disturbance torque produced by the tether bending, that is

$$T_c \triangleq T_a - T_d \tag{3.132}$$

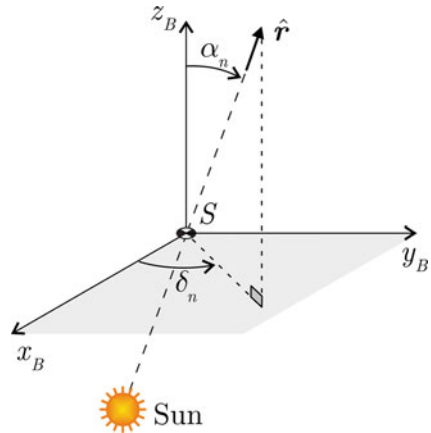
The value of $\Delta\sigma \triangleq (\sigma_{\max} - \sigma_{\min}) > 0$, necessary for generating the desired control torque T_c , is

$$\Delta\sigma = \kappa T_c \tag{3.133}$$

where κ is given by Eq. (3.130). The two vectors on the right-hand side of Eq. (3.132) will be analyzed separately.

Consider first the attitude variation torque T_a . Its magnitude T_a is found by enforcing a desired time evolution of the sail pitch angle. Without loss of generality, assume that the E-sail is initially placed in a Sun-facing configuration, such that its nominal plane (x_B, y_B) is orthogonal to the Sun-sail direction; see Fig. 3.21. The attitude maneuver is better visualized with the aid of two additional reference frames. More precisely, let $T_I(S; x_I, y_I, z_I)$ be an inertial reference frame of unit

Fig. 3.21 Sketch of an E-sail at a generic given attitude. Adapted from Bassetto et al. (2020)



vectors $\{\hat{i}_I, \hat{j}_I, \hat{k}_I\}$ with $\hat{i}_I \equiv \hat{i}_{B_0}$ and $\hat{k}_I \equiv \hat{k}_{B_0} \equiv \hat{r}$, where \hat{i}_{B_0} (or \hat{k}_{B_0}) is the unit vector \hat{i}_B (or \hat{k}_B) at the initial time $t = 0$. Also introduce an auxiliary reference frame $\mathcal{T}_A(S; x_A, y_A, z_A)$ of unit vectors $\{\hat{i}_A, \hat{j}_A, \hat{k}_A\}$ with $\hat{j}_A \equiv \hat{j}_I$. In particular, \mathcal{T}_A is a non-inertial reference frame, which rotates about the y_A -axis (fixed in the inertial space) of an angle equal to the pitch angle α_n , until the attitude maneuver is completed; see the sketch in Fig. 3.22.

The rotation of the sail nominal plane (x_B, y_B) is chosen to track a desired time variation of the pitch angle, that is

$$\dot{\tilde{\alpha}}_n \triangleq A \left[1 - \frac{4}{t_f^2} \left(t - \frac{t_f}{2} \right)^2 \right] \Pi \quad (3.134)$$

where the symbol \sim denotes the desired value, t_f the attitude maneuver time, A is a suitable constant, and

$$\Pi \triangleq 1(t) - 1(t - t_f) \quad (3.135)$$

in which $1(t)$ is the step function. Note that the initial and final equilibrium conditions at the beginning and at the end of the attitude maneuver, that is, $\tilde{\alpha}_n(0) = \tilde{\alpha}_n(t_f) = 0$, are automatically met. In this context, Eq. (3.134) can be integrated over time to get

$$\tilde{\alpha}_n = \frac{2A}{t_f} \left(t^2 - \frac{2}{3} \frac{t^3}{t_f} \right) \Pi \quad (3.136)$$

where the constant A is obtained by enforcing the final condition $\tilde{\alpha}_n(t_f) \triangleq \tilde{\alpha}_{n_f}$, viz.

$$A = \frac{3\tilde{\alpha}_{n_f}}{2t_f} \quad (3.137)$$

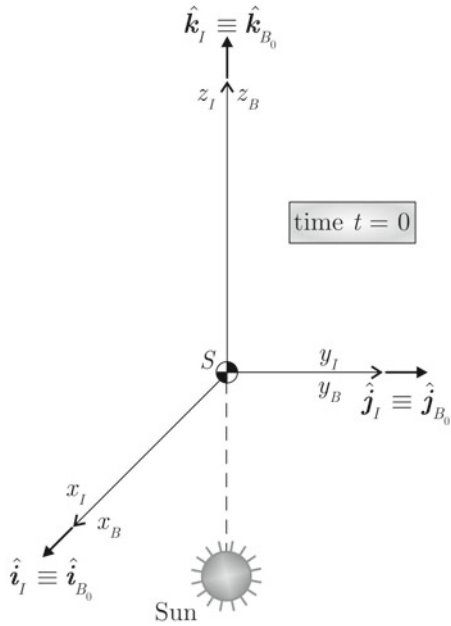
from which

$$\dot{\tilde{\alpha}}_n = \frac{6\tilde{\alpha}_{n_f}}{t_f} \left[\frac{t}{t_f} - \left(\frac{t}{t_f} \right)^2 \right] \Pi \quad (3.138)$$

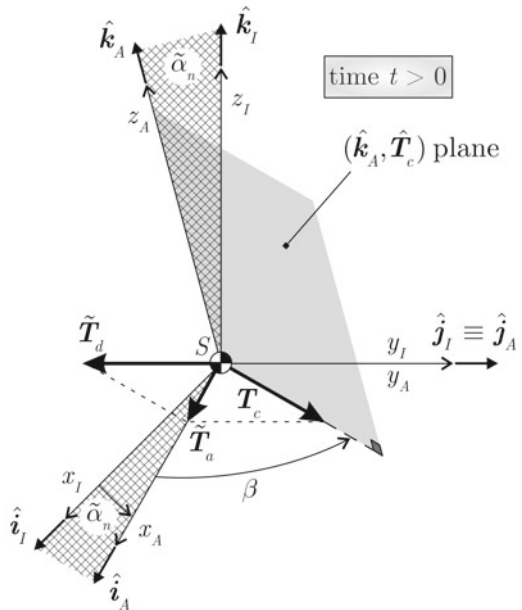
Figure 3.23 shows the desired time variation of the pitch angle and its derivative, corresponding to Eqs. (3.136) and (3.138), respectively.

The attitude variation torque T_a is chosen such that the rate of change of α_n equals that given by Eq. (3.138). Let $\tilde{\Omega}_A = \tilde{\alpha}_n \hat{j}_A$ be the angular velocity with which the auxiliary frame \mathcal{T}_A rotates about its y_A -axis. If the pitch angle were able to meet the desired time variation of Eq. (3.136), the principal body frame would exactly follow the auxiliary frame during the whole attitude maneuver and, accordingly, the

Fig. 3.22 Sketch of the attitude maneuver with the control torque. Adapted from Bassetto et al. (2020)



(a) E-sail orientation at $t = 0$.



(b) Attitude maneuver.

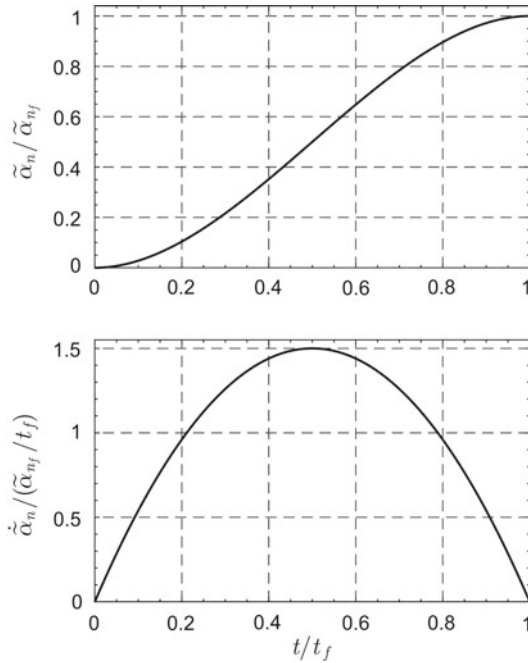


Fig. 3.23 Desired time variation of the pitch angle during the E-sail attitude maneuver. Adapted from Bassetto et al. (2020)

E-sail spin axis z_B would always coincide with z_A . The desired spacecraft angular momentum vector $\tilde{\mathbf{H}}$ in the auxiliary frame is therefore $\tilde{\mathbf{H}} = I_z \omega \hat{\mathbf{k}}_A$, where I_z is the spacecraft longitudinal moment of inertia. As a result, the required attitude variation torque $\tilde{\mathbf{T}}_A$ is given by

$$\tilde{\mathbf{T}}_A = \boldsymbol{\Omega}_A \times \tilde{\mathbf{H}} = T_a \hat{\mathbf{t}}_A \tag{3.199}$$

where

$$T_a \triangleq I_z \omega \dot{\tilde{\alpha}}_n = \frac{6 \tilde{\alpha}_{n_f} I_z \omega}{t_f} \left[\frac{t}{t_f} - \left(\frac{t}{t_f} \right)^2 \right] \Pi \tag{3.140}$$

Moreover, under the previous assumption that $z_B \equiv z_a$, the onset of a pitch angle different from zero also causes a disturbance torque to act along the y_A -axis, that is, $\tilde{\mathbf{T}}_d = -T_d \hat{\mathbf{j}}_A$, where T_d is given by Eq. (3.131). From Eq. (3.132), the control torque is

$$\mathbf{T}_c \simeq \tilde{\mathbf{T}}_a - \tilde{\mathbf{T}}_d \tag{3.141}$$

or

$$\mathbf{T}_c \simeq \frac{6\tilde{\alpha}_{n_f} I_z \omega}{t_f} \left[\frac{t}{t_f} - \left(\frac{t}{t_f} \right)^2 \right] \Pi \hat{\mathbf{i}}_A + \frac{NL(\sigma u)^2 \ln(2) \sin \alpha_n}{\rho \omega^2} \hat{\mathbf{j}}_A \tag{3.142}$$

While the attitude variation torque $\tilde{\mathbf{T}}_a$ is an explicit function of time, $\tilde{\mathbf{T}}_d$ depends on α_n , which must therefore be measured during the attitude maneuver. Because $\tilde{\mathbf{T}}_a$ and $\tilde{\mathbf{T}}_d$ are perpendicular to each other, the magnitude of the control torque is

$$T_c \triangleq \|\mathbf{T}_c\| = \sqrt{\tilde{T}_a^2 + \tilde{T}_d^2} \tag{3.143}$$

In analogy with the previously described approach for counterbalancing the disturbance torque alone, the sail nominal plane is now split into two parts by the control plane, which is orthogonal to the (x_A, y_A) plane and passes through \mathbf{T}_c . Again, the tethers belonging to the half-planes separated by the control plane are set to two different levels of electric potential, that is, according to Eq. (3.133), $\sigma \pm \kappa T_c/2$. The orientation of the control plane is defined by the angle $\beta = \beta(t, \alpha_n) \in [0, 2\pi)$ rad, which is the angle, measured counterclockwise from x_A , between \mathbf{T}_c and the x_A -axis. With reference to Fig. 3.24, the angle β is obtained as

$$\beta \triangleq \arctan\left(\frac{T_d}{T_a}\right) \tag{3.144}$$

where T_d and T_a are given by Eqs. (3.131) and (3.140), respectively.

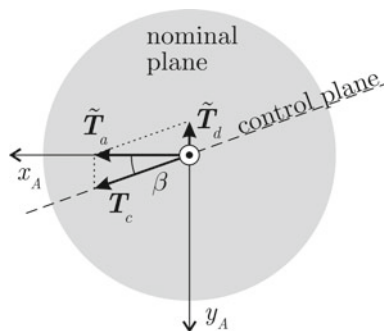


Fig. 3.24 Control torque and orientation of the control plane trace. Adapted from Bassetto et al. (2020)

Note that when the attitude maneuver ends, $T_a \rightarrow 0$ and $\beta \rightarrow \pi/2$ rad. The value of σ_k to be assigned to each tether depends on the angular displacement γ_k of the generic tether relative to the control plane, that is

$$\gamma_k \triangleq \omega t + \zeta_k - \beta \quad (3.145)$$

Therefore, the resultant control law is

$$\sigma_k \triangleq \begin{cases} \sigma - \frac{\kappa T_c}{2} & \text{if } \gamma_k \in [0, \pi) \text{ rad} \\ \sigma + \frac{\kappa T_c}{2} & \text{if } \gamma_k \in [\pi, 2\pi) \text{ rad} \end{cases} \quad (3.146)$$

where σ and κ are given by Eqs. (3.8) and (3.130), respectively, while T_c must be computed from Eq. (3.143). Finally, β is obtained as

$$\beta \triangleq \begin{cases} \arctan\left(\frac{T_d}{T_a}\right) & \text{if } t \leq t_f \\ \frac{\pi}{2} \text{ rad} & \text{if } t > t_f \end{cases}. \quad (3.147)$$

3.4.2.1 Numerical Simulations

The effectiveness of the proposed control law is now checked by numerically integrating the Euler attitude equations. To that end, the torque components generated by the E-sail are written in the principal body frame \mathcal{T}_B . The kinematic equations of a rigid E-sail are those given by Eqs. (3.86)–(3.88), while the classical Euler equations are

$$\dot{\Omega}_x = \lambda \Omega_y \Omega_z + \mathcal{E}/I_t \quad (3.148)$$

$$\dot{\Omega}_y = -\lambda \Omega_x \Omega_z + \mathcal{F}/I_t \quad (3.149)$$

$$\dot{\Omega}_z = \mathcal{G}/I_z \quad (3.150)$$

where λ is defined in Eq. (3.92), $\{\Omega_x, \Omega_y, \Omega_z\}$ are the components of the spacecraft angular velocity about its center of mass, and $\{\mathcal{E}, \mathcal{F}, \mathcal{G}\}$ are the components of the total torque generated by the E-sail in the principal body frame \mathcal{T}_B . Note that $\{\mathcal{E}, \mathcal{F}, \mathcal{G}\}$ include not only the contribution to the total torque given by the modulation of the electrical voltage, but also the effects of the tether bending. Bearing in mind that $b_l^2 \simeq 0$ (see Eq. (3.122)), the expressions of $d\mathcal{E}_k$, $d\mathcal{F}_k$, and $d\mathcal{G}_k$, taken from Sect. 2.1, are reported below for $(f')^2 = 0$

$$\begin{aligned} d\mathcal{E}_k &= \sigma_k u \left[(f - x f') \sin \zeta_k \sin \alpha_n \cos (\delta_n - \zeta_k) - f \sin \alpha_n \sin \delta_n \right. \\ &\quad \left. + x \sin \zeta_k \cos \alpha_n \right] dx \end{aligned} \quad (3.151)$$

$$\begin{aligned} d\mathcal{F}_k &= \sigma_k u \left[(x f' - f) \cos \zeta_k \sin \alpha_n \cos (\delta_n - \zeta_k) + f \sin \alpha_n \cos \delta_n \right. \\ &\quad \left. - x \cos \zeta_k \cos \alpha_n \right] dx \end{aligned} \quad (3.152)$$

$$d\mathcal{G}_k = \sigma_k u x \sin \alpha_n \sin (\delta_n - \zeta_k) dx \quad (3.153)$$

The torque components from the k -th tether are obtained by integrating Eqs. (3.151)–(3.153) along the tether length. The result is

$$\begin{aligned} \mathcal{E}_k &= u L^2 \sigma_k \{ b_l \sin \alpha_n [\cos (\delta_n - \zeta_k) \sin \zeta_k (\ln 8 - 2) - \sin \delta_n (\ln 4 - 1)] \\ &\quad + (\cos \alpha_n \sin \zeta_k) / 2 \} \end{aligned} \quad (3.154)$$

$$\begin{aligned} \mathcal{F}_k &= u L^2 \sigma_k \{ b_l \sin \alpha_n [-\cos (\delta_n - \zeta_k) \cos \zeta_k (\ln 8 - 2) + \cos \delta_n (\ln 4 - 1)] \\ &\quad - (\cos \alpha_n \cos \zeta_k) / 2 \} \end{aligned} \quad (3.155)$$

$$\mathcal{G}_k = u L^2 \sigma_k [\sin \alpha_n \sin (\delta_n - \zeta_k) / 2] \quad (3.156)$$

The special case of flat E-sail can be retrieved by simply setting $b_l = 0$. For exemplary purposes assume that $I_z = 3I_t = 3000 \text{ kg m}^2$, $L = 2 \text{ km}$, and $\omega \simeq 0.0758 \text{ rad/s}$, from which $b_l \simeq 6.471 \times 10^{-3}$. The initial conditions are $\phi(0) = \theta(0) = \psi(0) = \Omega_x(0) = \Omega_y(0) = 0$ and $\Omega_z(0) = 0$. Equations (3.86)–(3.88) and (3.148)–(3.150) have been numerically integrated by implementing the control law described by Eqs. (3.146) and (3.147) with $\tilde{\alpha}_{n_f} = 5^\circ$ and $t_f = 2 \text{ min}$. Two E-sail configurations have been considered, the former with $N = 16$, the latter with $N = 32$. In both cases, the simulations are reported in a time span greater than t_f to verify whether the E-sail can maintain its final attitude with the proposed control law.

The results are shown in Figs. 3.25, 3.26, 3.27 and 3.28. The time evolutions of the sail pitch angle during and after the attitude maneuver (see Fig. 3.25) exhibit an oscillatory behavior around a value close to the desired angle of 5° . It can also be observed that the mean asymptotic values tend to slightly decrease over time. Such a behavior is due to the combined effect of the torque component generated along the z_B -axis (see Fig. 3.26), which causes a small increase of the sail spin rate from 0.0758 rad/s to 0.0761 rad/s and the presence of the disturbance torque T_d . Indeed, the observed decrease of α_n no longer occurs when the torque along z_B is set equal to zero (i.e., $\mathcal{G} = 0$) and/or when the E-sail takes a perfectly flat shape (i.e., $b_l = 0$). This is clearly illustrated in Fig. 3.27, where the time evolution of α_n is shown for

$N = 16$ in three different cases characterized by $\{b_l \neq 0, \mathcal{G} \neq 0\}$, $\{b_l = 0, \mathcal{G} \neq 0\}$, and $\{b_l \neq 0, \mathcal{G} = 0\}$. Figure 3.26 also shows the components \mathcal{E} and \mathcal{F} , the values of which never exceed 0.25 Nm. Finally, Fig. 3.28 reports the time evolutions of $\Delta\sigma/\sigma$. Note that $\Delta\sigma = 0$ at $t = 0$ because of the assumption that the E-sail starts its attitude maneuver from a Sun-facing configuration. During the attitude variation, the maximum value of $\Delta\sigma/\sigma$ is about 0.0651 (or 0.0327) when $N = 16$ (or $N = 32$). Therefore, for both E-sail configurations, the electrical voltage modulation is a few percent only of its nominal value. The values of $\Delta\sigma$ for $t > t_f$ allow an E-sail to maintain its final attitude, in accordance with the results presented in Sect. 3.4.1.

For a given E-sail configuration, it is interesting to compare the order of magnitude of $\Delta\sigma$ required for varying the spacecraft attitude with that necessary for maintaining it. To that end, since

$$\kappa T_a \propto \sin\left(\frac{\pi}{N}\right), \kappa T_d \propto N \sin\left(\frac{\pi}{N}\right) \tag{3.157}$$

it turns out that $T_a/T_d \propto N^{-1}$. This is confirmed by the results shown in Fig. 3.28, in which the required $\Delta\sigma$ during the attitude maneuver is maximum in the configuration with the smallest number of tethers ($N = 16$). In other words, when N is sufficiently small, the voltage modulation necessary for removing the disturbance torque becomes negligible when compared to that required for changing the spacecraft attitude. In those cases, the assumption of a flat sail (corresponding to a negligible disturbance torque due to the tether bending) becomes a reasonable approximation, useful for simplifying the mission analysis.

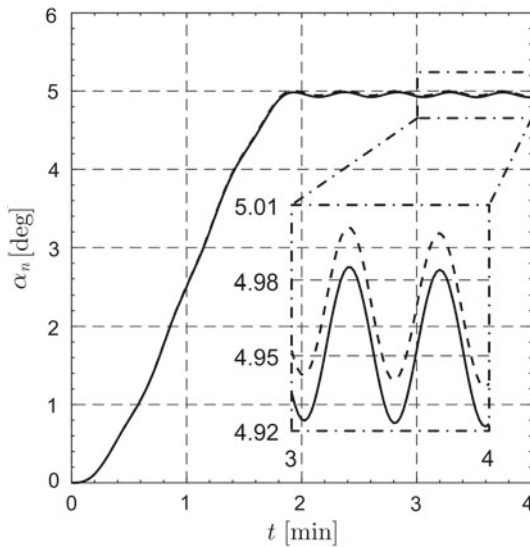


Fig. 3.25 Time evolution of the pitch angle α_n during attitude maneuver when $N = 16$ (solid line) and $N = 32$ (dotted line). Adapted from Bassetto et al. (2020)

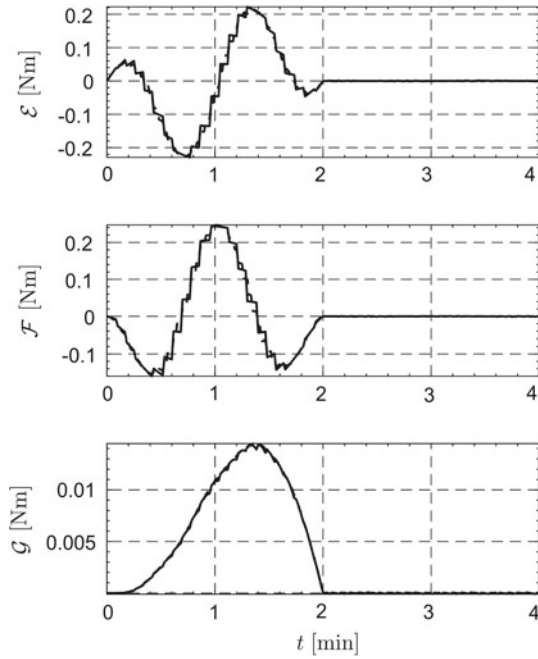


Fig. 3.26 Components of E-sail torque in \mathcal{T}_B during attitude maneuver when $N = 16$ (solid line) and $N = 32$ (dotted line). Adapted from Bassetto et al. (2020)

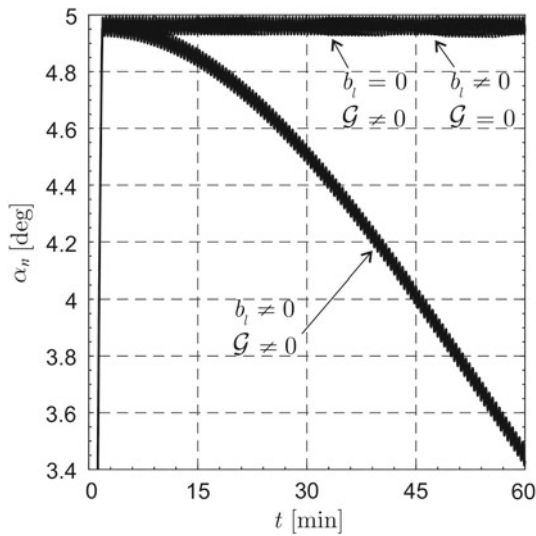


Fig. 3.27 Time evolution of α_n for $N = 16$ when $\{b_l \neq 0, \mathcal{G} \neq 0\}$, $\{b_l = 0, \mathcal{G} \neq 0\}$, and $\{b_l \neq 0, \mathcal{G} = 0\}$. Adapted from Bassetto et al. (2020)

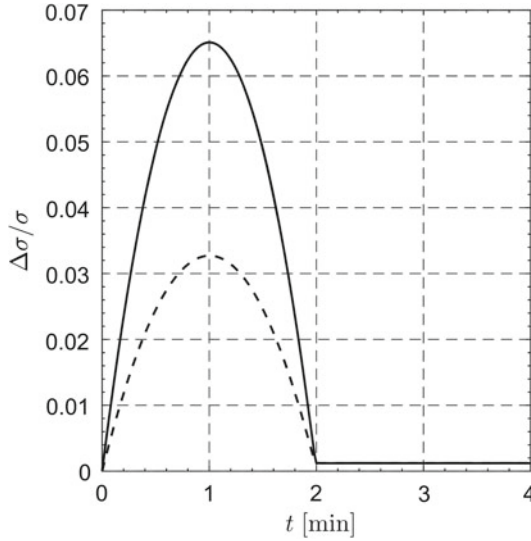


Fig. 3.28 Time evolution of $\Delta\sigma/\sigma$ during attitude maneuver when $N = 16$ (solid line) and $N = 32$ (dotted line). Adapted from Bassetto et al. (2020)

3.5 Mission Applications

Because propellantless propulsion systems can provide a continuous thrust without mass consumption, they turn out to be needful when a spacecraft must accomplish high-energy or long-term deep-space missions. The use of propellantless propulsion systems is indeed fundamental for tracking non-Keplerian orbits (Baig and McInnes 2010; Forward 1991; Heiligers and McInnes 2015; McKay et al. 2011; Zeng et al. 2014), generating artificial Lagrangian points (Aliasi et al. 2013; McInnes 1999), or maintaining a heliostationary position (Dandouras et al. 2004; McInnes 2003; Mengali and Quarta 2007). In the latter case, the spacecraft is first required to reach a point with zero absolute velocity (McInnes 2003; Mengali and Quarta 2007), and then to exploit its propulsive acceleration for balancing the solar gravitational attraction and maintaining that position. Possible scientific missions for a heliostationary spacecraft (Dandouras et al. 2004) include observations of the Sun, the monitoring of NEOs, or the release of a small solar probe along a rectilinear trajectory (Quarta and Mengali 2011, 2013). Solar activity monitoring and NEO surveillance could also be performed through the generation of an EFO (Heiligers and McInnes 2015), which is obtained when the apse line of the spacecraft osculating orbit follows the Earth during its revolution around the Sun. The basic idea behind this concept is to place the spacecraft into an eccentric orbit belonging to the ecliptic plane, the apse line of which precedes at a mean angular rate $\bar{\omega} \triangleq 2\pi$ rad/year. This amounts to having a phasing angle ϕ between the apse line of the osculating orbit and the Sun–Earth line that fluctuates (in time) with zero mean value; see Fig. 3.29.

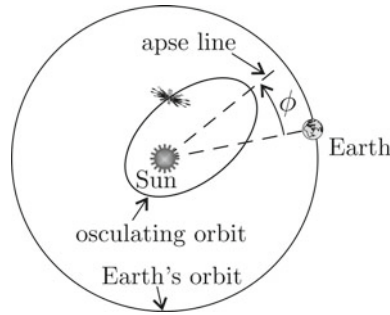


Fig. 3.29 Phasing angle ϕ in an EFO. Adapted from Bassetto et al. (2019c)

The maintenance of a heliostationary position and the generation of an EFO are investigated in the following parts of this section, with the assumption that the spacecraft is equipped with an E-sail.

3.5.1 Spinning E-sail in Heliostationary Condition for Solar Activity Monitoring

The problem of maintaining a heliostationary position is especially involved from a control point of view when an E-sail is used as primary propulsion system. In this case, in fact, the heliostationary condition is known to be unstable (Niccolai et al. 2018a) and, accordingly, a small error in the insertion causes the spacecraft to move away from the prescribed reference position. The aim of this section is therefore to study a feedback control system capable of stabilizing the dynamics of an E-sail around a heliostationary position at one astronomical unit from the Sun (i.e., $r = r_{\oplus}$). In such a situation, the spacecraft absolute velocity is zero, the thrust vector is parallel to the Sun–spacecraft line (that is, $\hat{k}_B \equiv \hat{r}$), and the propulsive acceleration magnitude balances the local Sun’s gravitational field. According to Eqs. (3.39), (3.40), and (3.64), and bearing in mind Eq. (3.8), the heliostationary condition is described by

$$\mathbf{T} = 0, \quad \frac{\|\mathbf{F}\|}{m} = \frac{N L \sigma_{\oplus} u}{m} = \frac{\mu_{\odot}}{r_{\oplus}^2} \tag{3.158}$$

where m is the total spacecraft mass, which is assumed to be constant. It is worth noting that the dynamics of a spinning E-sail exhibit a marked separation between the orbital and the attitude motion; see Sect. 3.3.2. Indeed, a deviation from the Sun-facing condition (that is, the presence of a small pitch angle $\alpha_n \neq 0$) generates a propulsive torque that induces a marginally stable oscillatory motion with a period on the order of few minutes; see Figs. 3.14, 3.15, 3.16 and 3.17. Such a peculiarity

allows the attitude dynamics and the orbital motion of a spinning E-sail to be studied separately.

The instability of the heliostationary condition can be checked starting from Eqs. (3.8), (3.40), and (3.158), according to which the propulsive acceleration about the reference point is

$$\frac{\|\mathbf{F}\|}{m} = \frac{N L \sigma_{\oplus} u}{m} \left(\frac{r_{\oplus}}{r} \right) = \frac{\mu_{\odot}}{r_{\oplus}^2} \left(\frac{r_{\oplus}}{r} \right) \quad (3.159)$$

from which the spacecraft equation of motion is

$$\ddot{r} = -\frac{\mu_{\odot}}{r^2} + \frac{\|\mathbf{F}\|}{m} = \frac{\mu_{\odot}}{r_{\oplus}^2} \left[\frac{r_{\oplus}}{r} - \left(\frac{r_{\oplus}}{r} \right)^2 \right] \quad (3.160)$$

If we define $v \triangleq (r/r_{\oplus} - 1)$ as the dimensionless error in radial distance, the linearization of Eq. (3.160) yields

$$\ddot{v} - \frac{\mu_{\odot}}{r_{\oplus}^3} v = 0 \quad (3.161)$$

which describes an unstable motion. Accordingly, a control system is required to make the heliostationary equilibrium point stable.

A simple solution is to change the tether electrical voltage V (and so, the value of σ_{\oplus}) as a function of the Sun–spacecraft distance. The voltage can therefore be conveniently adjusted so as to induce a variation of the propulsive acceleration in the form of a proportional control law, viz.

$$\frac{\|\mathbf{F}\|}{m} = \frac{\mu_{\odot}}{r_{\oplus}^2} \left(\frac{r_{\oplus}}{r} \right) (1 - k_p v) \quad (3.162)$$

where k_p is a constant, dimensionless, parameter. Since the magnitude of the propulsive acceleration is proportional to the tether voltage V (see Eqs. (3.8) and (3.40)), the maximum percentage variation of V is

$$\frac{\Delta V_{\max}}{\bar{V}} = k_p (v_{\max} - v_{\min}) \quad (3.163)$$

where v_{\max} (or v_{\min}) is the maximum (or minimum) value of v , while \bar{V} is the nominal value of the tether voltage. Substituting Eq. (3.162) into (3.160), the linearized dynamics about the reference position becomes

$$\ddot{v} + \frac{\mu_{\odot}}{r_{\oplus}^3} (k_p - 1) v = 0 \quad (3.164)$$

which describes a harmonic motion when $k_p > 1$, that is

$$v(t) = v_0 \cos(\omega_n t) + \frac{\dot{v}_0}{\omega_n} \sin(\omega_n t) \quad (3.165)$$

where $\omega_n \triangleq \sqrt{\mu_\odot(k_p - 1)/r_\oplus^3}$ is the natural frequency and $\{v_0, \dot{v}_0\}$ are the initial conditions. The oscillation period is $2\pi/\omega_n$, which is smaller (or greater) than 1 year if $k_p > 2$ (or $1 < k_p < 2$), whereas $\omega_n \triangleq \sqrt{\mu_\odot/r_\oplus^3}$ if $k_p = 2$. In the special case when $\dot{v}_0 = 0$, Eqs. (3.163) and (3.165) give $\Delta V_{\max}/\bar{V} = 2k_p v_0$, that is, the maximum percentage variation of the tether voltage is proportional (through k_p) to the error v_0 .

Because the linearized dynamics described by Eq. (3.164) have imaginary poles, the nonlinear model can be proved to be locally stable around the equilibrium point $v(0) = 0$ and $\dot{v}(0) = 0$ by defining an appropriate Lyapunov function. To that end, consider the nonlinear equation of motion

$$\ddot{v} = \frac{\mu_\odot}{r_\oplus^3} \frac{(1 - k_p)v - k_p v^2}{(1 + v)^2} \quad (3.166)$$

and let \mathbf{q} be defined as $[q_1 \ q_2]^T$, where $q_1 \triangleq v$ and $q_2 \triangleq \dot{v}$. Note that Eq. (3.166) can be rewritten as

$$\begin{cases} \dot{q}_1 = q_2 \\ \dot{q}_2 = \frac{\mu_\odot}{r_\oplus^3} \frac{(1 - k_p)q_1 - k_p q_1^2}{(1 + q_1)^2} \triangleq g(q_1) \end{cases} \quad (3.167)$$

Then introduce the candidate Lyapunov function

$$\mathcal{V}(\mathbf{q}) = - \int g(q_1) dq_1 + \frac{1}{2} q_2^2 \quad (3.168)$$

with $\mathcal{V}(0) = 0$, from which

$$\mathcal{V}(\mathbf{q}) = \frac{\mu_\odot}{r_\oplus^3} \left[\frac{q_1}{1 + q_1} + k_p q_1 - (1 + k_p) \ln(1 + q_1) \right] + \frac{1}{2} q_2^2 \quad (3.169)$$

It may be verified that, when $k_p > 1$, $\mathcal{V}(\mathbf{q}) > 0 \forall \mathbf{q} \neq 0$ and $\dot{\mathcal{V}}(\mathbf{q}) = 0 \forall \mathbf{q}$. This implies that $\mathcal{V}(\mathbf{q})$ is a Lyapunov function and the origin is a locally stable point when a proportional feedback control law is implemented.

The radial oscillations described by Eq. (3.165) can also be damped out with a proportional-derivative control system. In that case, the tether voltage is modulated in such a way that the propulsive acceleration magnitude is in the form

$$\frac{\|\mathbf{F}\|}{m} = \frac{\mu_{\odot}}{r_{\oplus}^2} \left(\frac{r_{\oplus}}{r} \right) \left(1 - k_p v - \frac{k_d \dot{v}}{\sqrt{\mu_{\odot}/r_{\oplus}^3}} \right) \quad (3.170)$$

where $k_d > 0$ is a constant, dimensionless, parameter. In this case, the linearized dynamics is

$$\ddot{v} + \frac{\mu_{\odot}}{r_{\oplus}^3} \left[(k_p - 1)v + \frac{k_d \dot{v}}{\sqrt{\mu_{\odot}/r_{\oplus}^3}} \right] = 0 \quad (3.171)$$

which describes a second-order system with damping factor $\zeta \triangleq k_d / (2\sqrt{k_p - 1})$ and natural frequency ω_n . For example, when $k_d = 2\sqrt{k_p - 1}$ (that is, $\zeta = 1$), the time variation of the radial error is

$$v(t) = \exp(-\omega_n t) [v_0 + (\dot{v}_0 + \omega_n v_0)t] \quad (3.172)$$

and the maximum variation of tether voltage is $\Delta V_{\max}/\bar{V} = v_0$ if $\dot{v}_0 = 0$. Note that, in this case, the stability of the complete system follows from the asymptotic stability of the linearized equation of motion.

3.5.2 Thrust Vectoring for the Generation of EFOs

With reference to Fig. 3.12, the E-sail propulsive acceleration vector is

$$\mathbf{a} = \tau \frac{a_c}{2} \left(\frac{r_{\oplus}}{r} \right) [\hat{\mathbf{e}}_r + (\hat{\mathbf{e}}_r \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] \quad (3.173)$$

where the switching variable $\tau \in \{0; 1\}$ is now introduced to model the possibility of turning off (with $\tau = 0$) the electron gun in such a way as to obtain a coasting arc.

The problem of generating an EFO is here addressed by finding the control parameters α_n and τ that maximize the time derivative of the argument of perihelion ω at any time t . In this case, since the spacecraft motion takes place on the ecliptic plane (i.e., its trajectory is two-dimensional), the normal component of the propulsive acceleration must be $a_n = 0$, while the radial (a_r) and transversal (a_t) components become

$$a_r = \tau \frac{a_c}{2} \left(\frac{r_{\oplus}}{r} \right) (1 + \cos^2 \alpha_n) \quad (3.174)$$

$$a_t = \tau \frac{a_c}{2} \left(\frac{r_{\oplus}}{r} \right) \sin \alpha_n \cos \alpha_n \quad (3.175)$$

where now the E-sail pitch angle ranges in the interval $[-90, 90]^\circ$. In this simplified scenario, the time derivative of ω is (Battin 1999)

$$\dot{\omega} = \sqrt{\frac{p}{\mu_\odot}} \left[-\frac{\cos v}{e} a_r + \frac{2 + e \cos v}{e(1 + e \cos v)} \sin v a_t \right] \quad (3.176)$$

where p is the semilatus rectum of the osculating orbit, v is the true anomaly, $u \triangleq \omega + v$ is the argument of latitude, and $\{a_r, a_t\}$ are given by Eqs. (3.174)–(3.175). The locally optimal values α_n^* and τ^* can be determined by first computing the stationary points $\{\alpha_{n_1}, \alpha_{n_2}\}$ of the function $\dot{\omega}(\alpha_n)$. To that end, consider the equation

$$\frac{\partial \dot{\omega}}{\partial \alpha_n} = 0 \quad (3.177)$$

from which

$$\alpha_{n_1} = \arcsin \left(\sqrt{\frac{k_\omega^2 - k_\omega \sqrt{k_\omega^2 + 1} + 1}{2(k_\omega^2 + 1)}} \right) \quad (3.178)$$

$$\alpha_{n_2} = \alpha_{n_1} - \frac{\pi}{2} \quad (3.179)$$

where

$$k_\omega \triangleq -\frac{(1 + e \cos v) \cos v}{(2 + e \cos v) \sin v} \quad (3.180)$$

Even though $\{\alpha_{n_1}, \alpha_{n_2}\}$ are functions of e and v only, the locally optimal sail pitch angle α_n^* must account for the sign of $\dot{\omega}$, which is also a function of the argument of latitude u . Finally, the optimal switching parameter τ^* is equal to 1 (or 0) when the maximum obtainable value of $\dot{\omega}$ is positive (or negative), that is

$$\tau^* = \frac{\text{sign}\{\dot{\omega}(\alpha_n^*)\} + 1}{2} \quad (3.181)$$

where sign is the signum function.

The results are reported in graphical form in Fig. 3.30, where the eccentricity is chosen within the interval $e \in [0.1, 0.9]$, corresponding to closed osculating orbits. Note that, however, Eqs. (3.178)–(3.180) are valid for any value of the orbital eccentricity. In particular, Fig. 3.30 shows that ω cannot be increased when $v \in [-55.1, 55.1]^\circ$ and $e = 0.1$, while the interval of forbidden true anomalies is $v \in [-58.9, 58.9]^\circ$ when $e = 0.9$.

Having determined the locally optimal control law, the semimajor axis a_0 and the eccentricity e_0 of the parking orbit are calculated for a given a_c to satisfy the constraint

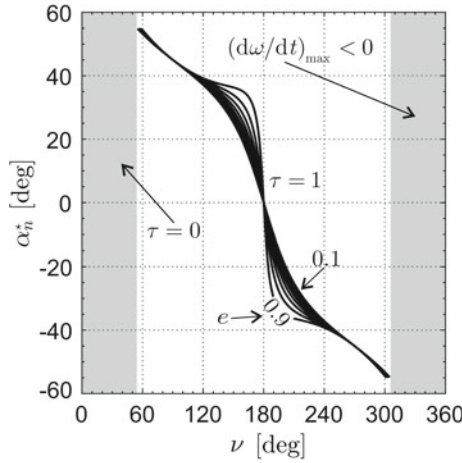
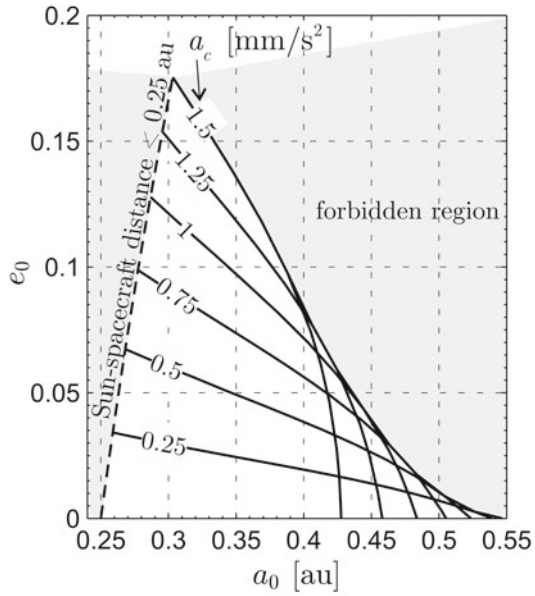


Fig. 3.30 Control parameters $\{\alpha_n^*, \tau^*\}$ for the local maximization of $\dot{\omega}$. Adapted from Bassetto et al. (2019c)

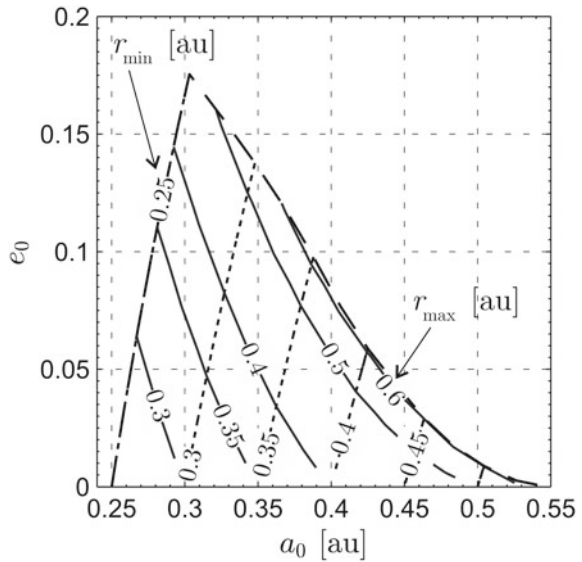
$\bar{\omega} \triangleq 2\pi$ rad/year. In particular, the assumptions are made that the perihelion of the parking orbit is in opposition to the Earth and that the initial true anomaly is $\nu_0 = 0$. Equation (3.68) is numerically integrated with a pitch angle $\alpha_n^* = \alpha_{n_1}$ (or $\alpha_n^* = \alpha_{n_2}$) if $\nu \in [0, \pi]$ rad (or $\nu \in [\pi, 2\pi]$ rad), where $\{\alpha_{n_1}, \alpha_{n_2}\}$ are given by Eqs. (3.178) and (3.179); see also Fig. 3.30. Figure 3.31a shows the possible pairs $\{a_0, e_0\}$ that solve the problem for a given value of $a_c \in [0.25, 1.5]$ mm/s², that is, for a medium-high-performance E-sail. Not all the pairs $\{a_0, e_0\}$ turn out to be admissible and, in fact, the forbidden region in the right-hand side of Fig. 3.31a corresponds to initial conditions that provide an insufficient apsidal precession rate, while the dotted line defines the constraint about the minimum heliocentric distance, set equal to 0.25 au, which is introduced to prevent the E-sail from an excessive thermal load (Quarta and Mengali 2010). The possible combinations of aphelion (r_{\max}) and perihelion (r_{\min}) radii are shown in Fig. 3.31b as a function of $\{a_0, e_0\}$.

For example, consider a spacecraft trajectory with a perihelion radius $r_{\min} = 0.4$ au and an aphelion radius $r_{\max} = 0.6$ au. According to Fig. 3.31a, the parking orbit characteristics are $a_0 \simeq 0.42$ au and $e_0 \simeq 0.06$, while the required characteristic acceleration is $a_c \simeq 1$ mm/s². Figure 3.32 shows the time variation of the control parameters $\{\alpha_n^*, \tau^*\}$, whereas Figs. 3.33, 3.34 and 3.35 show the time variation of $\{a, e, \omega\}$ of the osculating orbit. Note that the functions a and e are periodic with the same frequency, as they take their initial values at the end of each on/off cycle; see Figs. 3.33 and 3.34. The argument of perihelion, instead, is a monotonic increasing function of time; see Fig. 3.35. In this context, Fig. 3.36 shows the time variation of the phasing displacement ϕ . As expected, the mean value of ϕ is zero, whereas its maximum amplitude $|\phi|$ is about 17°. Within the chosen interval of a_c , the simulations show that $|\phi|_{\max}$ is always below 23.5°; see Fig. 3.37. In the limiting case as a_c tends to zero, the apse line does not rotate at all ($\omega = \omega_0$) and $|\phi|_{\max}$ tends to 180°.

Fig. 3.31 Parking orbits characteristics and Sun-spacecraft extreme distances for an EFO with a medium-low-performance E-sail. Adapted from Bassetto et al. (2019c)



(a) Parking orbit characteristics.



(b) Perihelion (r_{min}) and aphelion (r_{max}) radii.

Fig. 3.32 Optimal pitch angle and switching variable for an EFO with $a_c = 1 \text{ mm/s}^2$, $r_{\min} = 0.4 \text{ au}$, and $r_{\max} = 0.6 \text{ au}$. Adapted from Bassetto et al. (2019c)

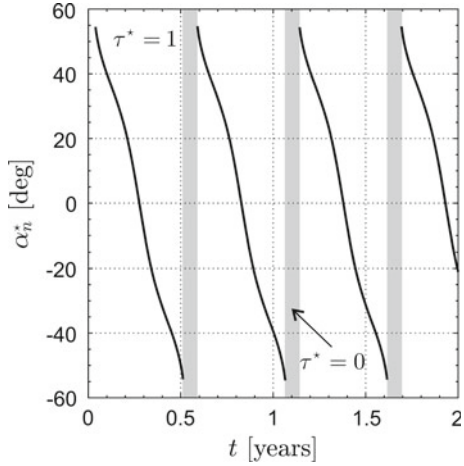
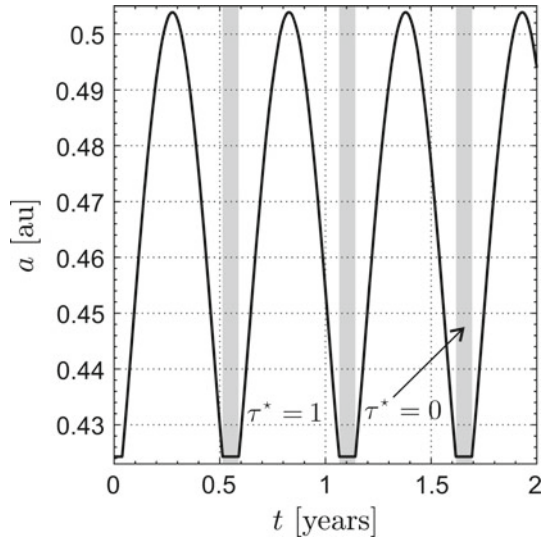


Fig. 3.33 Time variation of a for an EFO with $a_c = 1 \text{ mm/s}^2$, $r_{\min} = 0.4 \text{ au}$, and $r_{\max} = 0.6 \text{ au}$. Adapted from Bassetto et al. (2019c)



3.6 Conclusions

This chapter has first addressed the problem of determining the thrust and torque vectors provided by an E-sail of given shape as a function of the spacecraft attitude. The general expressions of the thrust and torque vectors have been specialized to the case of a Sun-facing E-sail, showing that the equilibrium shape of each tether may be approximated by a natural logarithmic arc when its spin rate is sufficiently high. With the assumption that the E-sail maintains such an equilibrium shape, analytical expressions of the thrust and torque vectors have then been derived as a function of

Fig. 3.34 Time variation of e for an EFO with $a_c = 1 \text{ mm/s}^2$, $r_{\min} = 0.4 \text{ au}$, and $r_{\max} = 0.6 \text{ au}$. Adapted from Bassetto et al. (2019c)

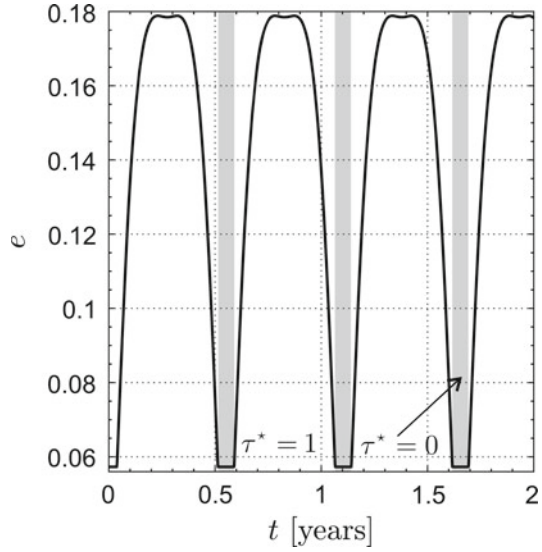
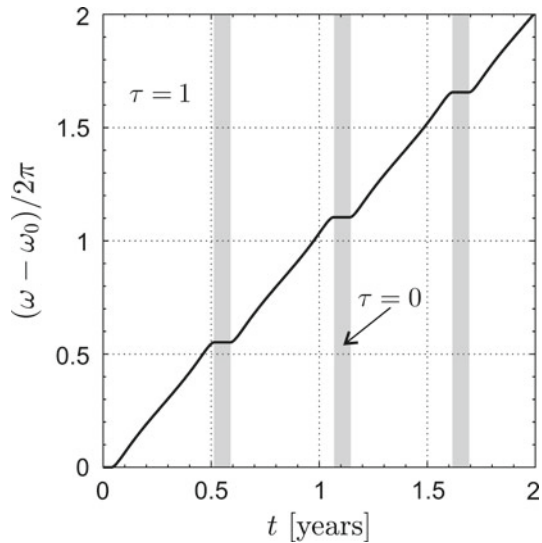


Fig. 3.35 Time variation of ω for an EFO with $a_c = 1 \text{ mm/s}^2$, $r_{\min} = 0.4 \text{ au}$, and $r_{\max} = 0.6 \text{ au}$. Adapted from Bassetto et al. (2019c)



the sail orientation. The proposed mathematical model allows the performance of an E-sail to be quantified in closed form, and, as such, to be easily implemented in a simulation code to get preliminary mission results.

The differential equations of the orbital and attitude dynamics of an E-sail-based spacecraft are introduced in the central part of this chapter.

Then, an analytical control law has been investigated to counteract the disturbance torque due to the tether bending, which induces a perturbation on the orientation of

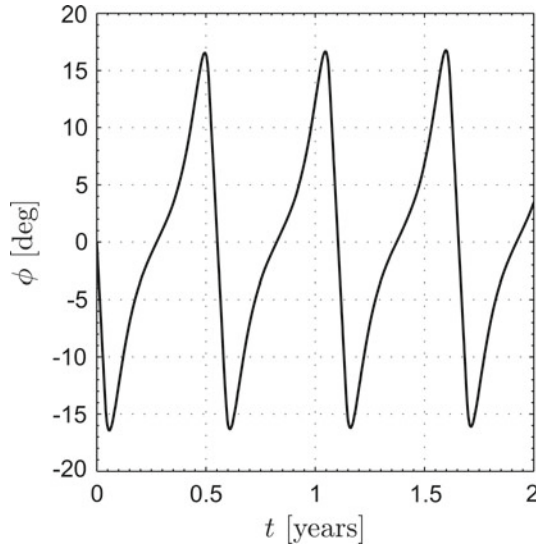


Fig. 3.36 Phase displacement ϕ for an EFO with $a_c = 1 \text{ mm/s}^2$, $r_{\min} = 0.4 \text{ au}$, and $r_{\max} = 0.6 \text{ au}$. Adapted from Bassetto et al. (2019c)

the spacecraft spin axis. In this case, the E-sail has been assumed to maintain its equilibrium shape found in the Sun-facing configuration, a reasonable hypothesis if the sail pitch angle is sufficiently small, and the tether voltage slightly differs from its nominal value. The strategy discussed, which consists in a modulation of the tether electrical voltage without modifying the total thrust, is simple and effective. In particular, the results have shown that the tether electrical voltage requires a very small variation (on the order of 1% or less) with respect to its nominal level, thus allowing the E-sail to maintain a fixed attitude and, as such, to generate a long-term nonzero transverse thrust. This chapter has also proved that the attitude of an E-sail-based spacecraft may be changed by modulating the electrical voltage of each tether. To that end, an analytical control law has been proposed, in which the voltage level of each tether is expressed as a function of the time and the sail pitch angle. Also, in this case a small variation of the voltage level is sufficient for most practical purposes. The proposed method for controlling and maintaining the spacecraft attitude is easy to implement and offers good performance in terms of reorientation time.

Finally, two scientific mission scenarios have been investigated with the aim of testing the E-sail performance in a heliocentric context for solar activity monitoring and NEO surveillance. The generation of earth-following orbits has been performed through a locally optimal control law, which has been analytically computed as a function of the spacecraft state variables, in such a way that the apsidal precession rate is maximized at any time. In particular, the orbital parameters of the parking orbit have been calculated a posteriori as a function of the E-sail characteristic acceleration

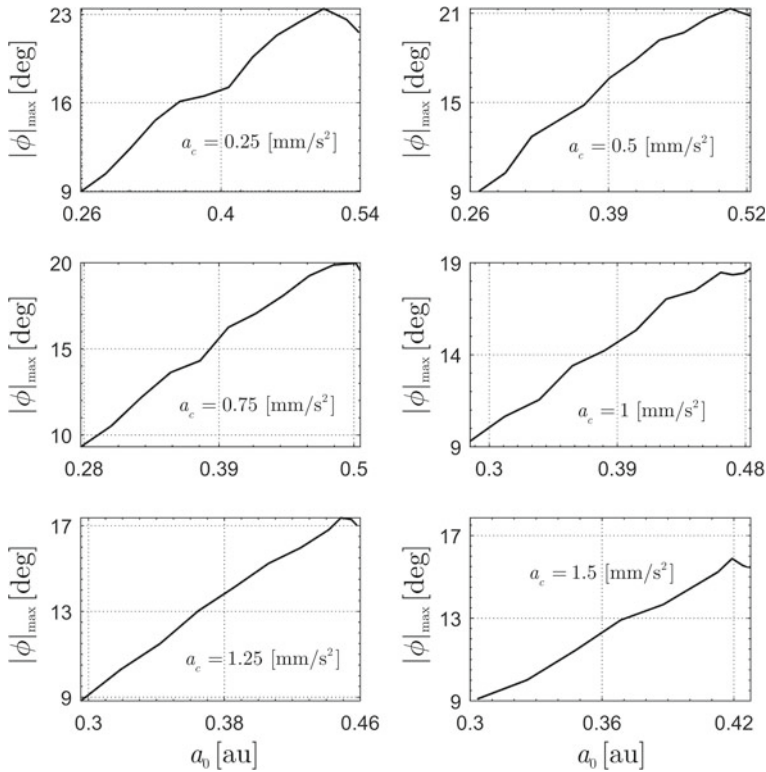


Fig. 3.37 Maximum phasing angle $|\phi|_{\max}$ as a function of $\{a_0, e_0\}$ and $a_c \in [0.25, 1.5] \text{ mm/s}^2$. Adapted from Bassetto et al. (2019c)

in order to satisfy the mission requirements. Finally, the dynamics of a spinning E-sail around a heliostationary position at one astronomical unit from the Sun has been investigated. It has been shown that, with a suitable modulation of the tether electrical voltage, the spacecraft center of mass moves along the Sun–spacecraft line around its nominal position. When a simple proportional controller is used, the maximum variation of tether voltage is proportional to the error in orbit insertion. The resulting spacecraft motion is an undamped harmonic oscillation with a period on the order of some years.

Abbreviations

- E-sail** Electric solar wind sail
- EFO** Earth-following orbit
- NEO** Near-earth object

References

- Aliasi, G., G. Mengali, and A.A. Quarta. 2013. Artificial Lagrange points for solar sail with electrochromic material panels. *Journal of Guidance, Control, and Dynamics* 36 (5): 1544–1550. <https://doi.org/10.2514/1.58167>.
- Baig, S., and C.R. McInnes. 2010. Light-levitated geostationary cylindrical orbits are feasible. *Journal of Guidance, Control, and Dynamics* 33 (3): 782–793. <https://doi.org/10.2514/1.46681>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2018a. Thrust and torque vector characteristics of axially-symmetric E-sail. *Acta Astronautica* 146: 134–143. <https://doi.org/10.1016/j.actaastro.2018.02.035>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2019a. Attitude dynamics of an electric sail model with a realistic shape. *Acta Astronautica* 159: 250–257. <https://doi.org/10.1016/j.actaastro.2019.03.064>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2019b. Stability and control of spinning E-sail in heliostationary orbit. *Journal of Guidance, Control, and Dynamics* 42 (2): 425–431. <https://doi.org/10.2514/1.G003788>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2020. E-sail attitude control with tether voltage modulation. *Acta Astronautica* 166: 350–357. <https://doi.org/10.1016/j.actaastro.2019.10.023>.
- Bassetto, M., A.A. Quarta, and G. Mengali. 2019c. Locally-optimal electric sail transfer. In *Proceedings of the institution of mechanical engineers, Part G: Journal of aerospace engineering*, vol. 233, no. 1, 166–179. <https://doi.org/10.1177/0954410017728975>.
- Bassetto, M., et al. 2018b. Plasma brake approximate trajectory. Part II: Relative motion. *Advances in the Astronautical Sciences* 163: 249–259.
- Battin, R.H. 1999. *An introduction to the mathematics and methods of astrodynamics*, Revised Edition, Chap. 10, 484–489. Reston, VA: AIAA. ISBN: 1-56347-342-9.
- Betts, J.T. 2000. Very low-thrust trajectory optimization using a direct SQP method. *Journal of Computational and Applied Mathematics* 120 (1): 27–40.
- Dandouras, I., B. Pirard, and J.Y. Prado. 2004. High performance solar sails for linear trajectories and heliostationary missions. *Advances in Space Research* 34 (1): 198–203. <https://doi.org/10.1016/j.asr.2003.02.055>.
- Forward, R.L. 1991. Statite—A spacecraft that does not orbit. *Journal of Spacecraft and Rockets* 28 (5): 606–611. <https://doi.org/10.2514/3.26287>.
- Fulton, J., and H. Schaub. 2018. Fixed-axis electric sail deployment dynamics analysis using hub-mounted momentum control. *Acta Astronautica* 144: 160–170. <https://doi.org/10.1016/j.actaastro.2017.11.048>.
- Heiligers, J., and C.R. McInnes. 2015. Solar sail heliocentric Earth-following orbits. *Journal of Guidance, Control and Dynamics* 38 (5): 937–944. <https://doi.org/10.2514/1.G000579>.
- Huo, M., G. Mengali, and A.A. Quarta. 2018. Electric sail thrust model from a geometrical perspective. *Journal of Guidance, Control, and Dynamics* 41 (3): 735–741. <https://doi.org/10.2514/1.G003169>.
- Janhunen, P. 2004. Electric sail for spacecraft propulsion. *Journal of Propulsion and Power* 20 (4): 763–764. <https://doi.org/10.2514/1.8580>.
- Janhunen, P. 2010a. Electric solar wind sail in-space propulsion status report. In *European planetary science congress 2010*. Vol. 5. Paper EPSC 2010-297. European Planetology Network and the European Geosciences Union.
- Janhunen, P. 2010b. Electrostatic plasma brake for deorbiting a satellite. *Journal of Propulsion and Power* 26 (2): 370–372. <https://doi.org/10.2514/1.47537>.
- Janhunen, P. 2011. Status report of the electric sail in 2009. *Acta Astronautica* 68 (5–6): 567–570. <https://doi.org/10.1016/j.actaastro.2010.02.007>.
- Janhunen, P. 2013. Photonic spin control for solar wind electric sail. *Acta Astronautica* 83: 85–90. <https://doi.org/10.1016/j.actaastro.2012.10.017>.

- Janhunen, P., and A. Sandroos. 2007. Simulation study of solar wind push on a charged wire: Basis of solar wind electric sail propulsion. *Annales Geophysicae* 25 (3): 755–767. <https://doi.org/10.5194/angeo-25-755-2007>.
- Janhunen, P., and P.K. Toivanen. 2018. A scheme for controlling the E-sail's spin rate by the E-sail effect itself. In *Space Propulsion 2018*. Seville, Spain.
- Janhunen, P., A. A. Quarta, and G. Mengali. 2013. Electric solar wind sail mass budget model. *Geoscientific Instrumentation, Methods and Data Systems* 2 (1): 85–95. <https://doi.org/10.5194/gi-2-85-2013>.
- Janhunen, P., et al. 2010. Invited article: Electric solar wind sail: Toward test missions. *Review of Scientific Instruments* 81 (11): 111301. <https://doi.org/10.1063/1.3514548>.
- Longuski, J.M., et al. 2005. Analytical solutions for thrusting, spinning spacecraft subject to constant forces. *Journal of Guidance, Control, and Dynamics* 28 (6): 1301–1308. <https://doi.org/10.2514/1.12272>.
- McInnes, C. R. 1999. Artificial Lagrange points for a partially reflecting flat solar sail. *Journal of Guidance, Control, and Dynamics* 22 (1): 185–187. <https://doi.org/10.2514/2.7627>.
- McInnes, C. R. (2003). Inverse solar sail trajectory problem. *Journal of Guidance, Control, and Dynamics* 26 (2): 369–371. <https://doi.org/10.2514/2.5057>.
- McKay, R.J., et al. 2011. Survey of highly-non-Keplerian orbits with low-thrust propulsion. *Journal of Guidance, Control, and Dynamics* 34 (3): 645–666. <https://doi.org/10.2514/1.52133>.
- Mengali, G., and A.A. Quarta. 2007. Optimal heliostationary missions of high-performance sailcraft. *Acta Astronautica* 60 (8–9): 676–683. ISSN: 0094-5765. <https://doi.org/10.1016/j.actaastro.2006.07.018>.
- Mengali, G., A.A. Quarta, and G. Aliasi. 2013. A graphical approach to electric sail mission design with radial thrust. *Acta Astronautica* 82 (2): 197–208. <https://doi.org/10.1016/j.actaastro.2012.03.022>.
- Mengali, G., A.A. Quarta, and P. Janhunen. 2008. Electric sail performance analysis. *Journal of Spacecraft and Rockets* 45 (1): 122–129. ISSN: 0022-4650. <https://doi.org/10.2514/1.31769>.
- Niccolai, L., A.A. Quarta, and G. Mengali. 2017a. Electric sail elliptic displaced orbits with advanced thrust model. *Acta Astronautica* 138: 503–511. <https://doi.org/10.1016/j.actaastro.2016.10.036>.
- Niccolai, L., A.A. Quarta, and G. Mengali. 2017b. Two-dimensional heliocentric dynamics approximation of an electric sail with fixed attitude. *Aerospace Science and Technology* 71: 441–446. <https://doi.org/10.1016/j.ast.2017.09.045>.
- Niccolai, L., A.A. Quarta, and G. Mengali. (2018a). Electric sail-based displaced orbits with refined thrust model. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 232 (3): 423–432. <https://doi.org/10.1177/0954410016679195>.
- Niccolai, L., et al. 2018b. Plasma brake approximate trajectory. Part I: Geocentric motion. *Advances in the Astronautical Sciences* 163: 235–247.
- Orsini, L., et al. 2018. Plasma brake model for preliminary mission analysis. *Acta Astronautica* 144: 297–304. <https://doi.org/10.1016/j.actaastro.2017.12.048>.
- Quarta, A.A., and G. Mengali. 2010. Electric sail mission analysis for outer Solar System exploration. *Journal of Guidance, Control, and Dynamics* 33 (3): 740–755. <https://doi.org/10.2514/1.47006>.
- Quarta, A.A., and G. Mengali. 2011. Solar sail capabilities to reach elliptic rectilinear orbits. *Journal of Guidance, Control, and Dynamics* 34 (3): 923–926. <https://doi.org/10.2514/1.51638>.
- Quarta, A.A., and G. Mengali. 2013. Optimal solar sail transfer to linear trajectories. *Acta Astronautica* 82 (2): 189–196. <https://doi.org/10.1016/j.actaastro.2012.03.005>.
- Quarta, A.A., and G. Mengali. 2016a. Analysis of electric sail heliocentric motion under radial thrust. *Journal of Guidance, Control, and Dynamics* 39 (6): 1431–1435. <https://doi.org/10.2514/1.G001632>.
- Quarta, A.A., and G. Mengali. 2016b. Minimum-time trajectories of electric sail with advanced thrust model. *Aerospace Science and Technology* 55: 419–430. <https://doi.org/10.1016/j.ast.2016.06.020>.

- Seppänen, H., et al. 2013. One kilometer (1 km) electric solar wind sail tether produced automatically. *Review of Scientific Instruments* 84 (9). <https://doi.org/10.1063/1.4819795>.
- Toivanen, P. K., and P. Janhunen. 2013. Spin plane control and thrust vectoring of electric solar wind sail. *Journal of Propulsion and Power* 29 (1): 178–185. <https://doi.org/10.2514/1.B34330>.
- Toivanen, P.K., and P. Janhunen. 2017. Thrust vectoring of an electric solar wind sail with a realistic sail shape. *Acta Astronautica* 131: 145–151. <https://doi.org/10.1016/j.actaastro.2016.11.027>.
- Toivanen, P.K., P. Janhunen, and J. Envall. 2015. Electric sail control mode for amplified transverse thrust. *Acta Astronautica* 106: 111–119. <https://doi.org/10.1016/j.actaastro.2014.10.031>.
- Walker, M.J.H. 1986. Erratum: A set of modified equinoctial orbit elements. *Celestial Mechanics* 38 (4): 391–392. <https://doi.org/10.1007/BF01238929>.
- Walker, M.J.H., B. Ireland, and J. Owens. 1985. A set of modified equinoctial orbit elements. *Celestial Mechanics* 36 (4): 409–419. <https://doi.org/10.1007/BF01227493>.
- Wertz, J.R. (ed.). 1978. *Spacecraft attitude determination and control*, 1st ed. <https://doi.org/10.1007/978-94-009-9907-7>. Astrophysics and Space Science Library. Dordrecht, Holland: Springer Netherlands, pp. 760–766. ISBN: 978-90-277-1204-2.
- Yamaguchi, K., and H. Yamakawa. 2013. Study on orbital maneuvers for electric sail with on-off thrust control. *Aerospace Technology Japan, the Japan Society for Aeronautical and Space Sciences* 12: 79–88. <https://doi.org/10.2322/astj.12.79>.
- Zeng, X., K.T. Alfriend, and S.R. Vadali. 2014. Solar sail planar multireversal periodic orbits. *Journal of Guidance, Control, and Dynamics* 37 (2): 674–681. <https://doi.org/10.2514/1.58598>.

Chapter 4

Space Elevator for Space-Resource Mining



Yoji Ishikawa

Abstract The space elevator is an innovative space traffic and transportation system that may be built in the future. The Earth-based space elevator not only significantly reduces the cost of shipping to space compared to conventional rockets, but also enables safe and routine delivery of massive payloads to GEO and beyond. The Earth-based space elevator consists of three main elements: cable, climber, and station. The climber moves up and down the cable between the Earth surface and the top, and if a spacecraft is brought to a high altitude on a long cable by the climber and launched from the height, it can be flown to the orbit of a distant planet. This is because once released from the cable at altitudes above approximately 47,000 km, objects can escape the Earth's gravity. A space elevator can not only be built on the Earth but also on the Moon and other planets such as Mars, and even on asteroids. Cargos and/or crewed spacecraft dispatched by the Earth space elevator can fly to the Moon and Mars to be caught at the end tips of the lunar or Martian space elevators, and then brought down to the surface. Space elevators can be utilized at both ends of a flight when cargo and crew return from the Moon or Mars to the Earth. Asteroids are known to be rich in various kinds of resources, from platinum to nickel to carbon to water. Space elevators built on asteroids will help to mine and transport such resources. Transportation between the Earth, the Moon, Mars, the moons of Mars, and asteroids will be made possible by using space elevators at each location. While constructing a space elevator will take tremendous effort, when it is completed, it will contribute greatly to the resource mining of the solar system.

4.1 Introduction

The space elevator is an innovative future space traffic and transportation system. Expected to be built primarily on Earth, its biggest advantage is that it will reduce the cost of shipping to space by roughly two orders of magnitude compared to

Y. Ishikawa (✉)

Obayashi Future Lab, Technology Division, Obayashi Corporation, Shinagawa Intercity Tower B, 2-15-2 Konan, Minato-Ku, Tokyo 108-8502, Japan
e-mail: ishikawa.yoji@obayashi.co.jp

conventional rockets. A second significant strength is the safe and routine delivery of massive payloads to GEO and beyond. Instead of using fuel, the space elevator employs the Earth's dynamic rotation and power provided through laser beaming or some alternative to send spacecraft. The space elevator can not only deploy artificial satellites to Earth orbit, but also dispatch spacecraft to the Moon and the planets.

The planets that a spacecraft can reach using a space elevator depend on the height of the cable at which the spacecraft is released—the higher the altitude, the faster it will be released. Calculations show it is possible to escape the Earth's gravitational field by releasing the spacecraft from the cable at a height greater than approximately 47,000 km (Edwards and Westling 2002). When the spacecraft is released in a direction in which the Earth revolves around the Sun, its speed becomes the sum of its speed when it is released and the revolution speed of the Earth. The spacecraft will then enter an elliptical orbit, with the Earth as its perigee point. When another planet approaches the path of the orbit, the spacecraft can arrive at that planet. For example, when sending a spacecraft to Mars, by releasing it from the cable at a height of approximately 57,000 km, the spacecraft will achieve a speed that is the sum of the release speed and the velocity of the Earth around the Sun. By acquiring such a high speed, the spacecraft can follow an elliptical orbit (the Hohmann orbit) to make a transition to Mars. This is one of the methods for sending a spacecraft to outer planets such as Mars, but there is another way to allow more rapid transit. As far as travel to Mars is concerned, if the spacecraft climbed to the top of the cable (where the counterweight is located, and the height is, for example, 100,000 km; sometimes called the apex anchor (Swan et al. 2020)) rather than being released at a height of 57,000 km, it would acquire greater energy and reach destinations more rapidly. It will take six to nine months to arrive at Mars when released at 57,000 km; however, only two months are necessary if released at 100,000 km.

Conversely, if the spacecraft was released in a direction exactly opposite to the direction in which the Earth revolves around the Sun, then the spacecraft would acquire a speed that is the speed of the Earth's revolution minus the speed at the time of release. The spacecraft will enter an elliptical orbit with the Earth as its apogee point. If another planet happens to cross the orbit, the spacecraft can reach the planet. This is the method for sending spacecraft to inner planets, that is, Mercury and Venus.

A space elevator can be built not only on the Earth but also on the Moon and other planets such as Mars, and even on asteroids. Cargos and/or crewed spacecraft dispatched by the Earth space elevator can fly to the Moon and Mars to be caught at the end tips of the Lunar or Martian space elevators, and then brought down to the surface. In this way, both acceleration at departure and deceleration at re-entry using rockets become unnecessary, eliminating the need for rocket fuel and reducing transfer cost. Space elevators can be utilized at both ends of a flight when cargo and crew return from the Moon or Mars to the Earth.

Asteroids are known to be rich in various kinds of resources, from platinum to nickel to carbon to water. Space elevators built on asteroids will help to mine and transport such resources.

Space elevators not only lower the cost of transportation but also benefit various fields. They enable easier access to all the celestial bodies in the solar system and facilitate utilization of their resources, zero-gravity manufacturing, Earth observation, communication between the Moon and planets, and so on.

However, establishing a space elevator, no matter where it is built, still requires tremendous efforts in various areas such as technology, finance, and law. Despite this, mining of space resources will inevitably benefit from the reasonable transportation cost of space elevators, and this will become a driving force in the economy of the Earth, Moon, and other planets in the solar system in the near future.

4.2 Earth-Based Space Elevators

4.2.1 Overview

The history of space elevator research is surprisingly long. At the end of the nineteenth century, when there were no rockets yet, Konstantin Tsiolkovsky, known as the father of space engineering, used the Eiffel Tower as an inspiration to propose a space tower that would reach orbit if the tower was extended high enough. In 1960, Yuri N. Artsutanov suggested a precursor of the space elevator to be called a space cableway (Artsutanov 1960). In 1975, Jerome Pearson developed the basic design of a space elevator through a calculation of its mechanics (Pearson 1975). However, no material existed that could withstand the gravitational and centrifugal forces applied to the cable—a space elevator’s most essential component—and so the space elevator was thought to be only a dream.

A major turning point was the discovery of carbon nanotubes by Iijima (1991). In the early 2000s, Bradley Edwards conducted a full-scale study on a NASA project (Edwards and Westling 2002; Edwards 2000, 2002, and 2003), including cable and climber design, power transmission methods to climbers, earth port design, and countermeasures for the harmful space environment. It was a landmark study that outlined the archetypal space elevator. Based on these previous studies, in 2012, Obayashi Corporation published the “Space Elevator Construction Concept” in its public relations magazine *Quarterly Obayashi* at the same time as completing the construction of the world’s tallest free-standing tower, Tokyo Skytree® (Ishikawa et al. 2012). Their concept featured comprehensive architecture and an evolutionary approach for the foreseeable future, and it became one of the fundamental models of the space elevator (Ishikawa et al. 2013, 2016). Another important report came in 2013, with the results of a large-scale study organized by Peter Swan for the International Academy of Astronautics (Swan et al. 2013). It included not only the technical aspects of each component but also the feasibility, legal aspects, and finance. Their follow-up report was published in 2019; it dealt with the space elevator from a systems engineering approach and the basic conclusion of the 350-page report was that a space elevator was feasible, assuming carbon nanotubes could be produced (Swan et al. 2019).

In the next section, as an example of an Earth-based space elevator, Obayashi Corporation's concept is described in detail.

4.2.2 *Obayashi Corporation's Space Elevator Concept*

4.2.2.1 Principles

Space elevators are a promising means of transportation in the future. The transportation cost is estimated to be two orders of magnitude less than that of conventional rockets, and it is also relatively safer. In addition, the ability to move massive payloads routinely will enable new missions at GEO and beyond, such as developing space-based solar power. The Obayashi space elevator design incorporates the need for human transportation as a principal goal.

There are three key points concerning the space elevator. First, space elevators are different from the elevators used on Earth. In principle, the vehicle closest to the space elevator is a monorail. A monorail is a vehicle that travels on or under a single rail and is generally self-propelled, being driven by the rotation of tires or the like that sandwich the rails. In addition, stations are located at both ends and along the rails. Similar to a monorail, a space elevator is composed of a cable (counterpart of the rail), a self-propelled climber (the vehicle), and a space station (monorail stations). A space elevator can be regarded as a monorail rotated by 90° to rise vertically into the sky.

Second, a space elevator is a tower that hangs from space. The center of the space elevator is said to be in geosynchronous orbit (an altitude of approximately 36,000 km) and during the construction process a cable is extended upward and downward from this point. A ground structure is typically assembled from the bottom up, but a space elevator is a tower that hangs from geosynchronous orbit like a spider descending on its thread.

Third, a space elevator uses the principle of a hammer throw (or sling) to launch spacecraft. A hammer throw is an athletic event in which the rotation speed of a hammer is used to propel it in the tangential direction of the rotation, making it fly for a great distance. In the case of a space elevator, the climber does not have to only move up and down along the cable, as the cable is attached to the Earth and revolves with the Earth every 24 hour. The longer the cable, the faster it will revolve at higher positions. If a spacecraft is launched from a high altitude on a long cable, it can be flown to the orbit of a distant planet.

Edwards calculated the altitude required for a spacecraft to reach each planet, as listed in Table 4.1 (Edwards 2002; Swan et al. 2020). Obviously, the higher the altitude at which the spacecraft is released, the farther the planet that can be reached. In addition, by achieving maximum velocity (at the top of the cable) a payload can make faster flights to planets such as Mars in as few as 61 days (Swan et al. 2020). With a cable of a length of approximately 100,000 km, a spacecraft released from its end tip can reach the asteroid belt. Giving it extra velocity will enable it to reach

Table 4.1 Minimum altitude of the cable at which a spacecraft can reach the orbit of a planet once released (Edwards 2002; Swan et al. 2020)

Planet	Minimum altitude on the cable
Mercury	103,348 km
Venus	54,148 km
Earth	–
Mars	56,898 km
Asteroid belts (inner)	67,748 km
Asteroid belts (outer)	98,748 km
Jupiter	119,063 km
Saturn	138,418 km
Uranus	151,383 km
Neptune	156,322 km
Pluto	158,441 km
Solar system escape velocity	163,000 km

more distant planets such as Jupiter and Saturn. Releasing a spacecraft in the opposite direction of the Earth’s revolution allows it to be launched into the orbits of inner planets such as Mercury and Venus. In this way, it is possible to use a space elevator to launch a spacecraft into the orbit of any planet in the solar system. In Obayashi’s concept, the cable length is set to 96,000 km, that is, approximately 100,000 km. The concept is described in detail in this section and an image is shown in Fig. 4.1 (Ishikawa et al. 2012, 2013, 2016).

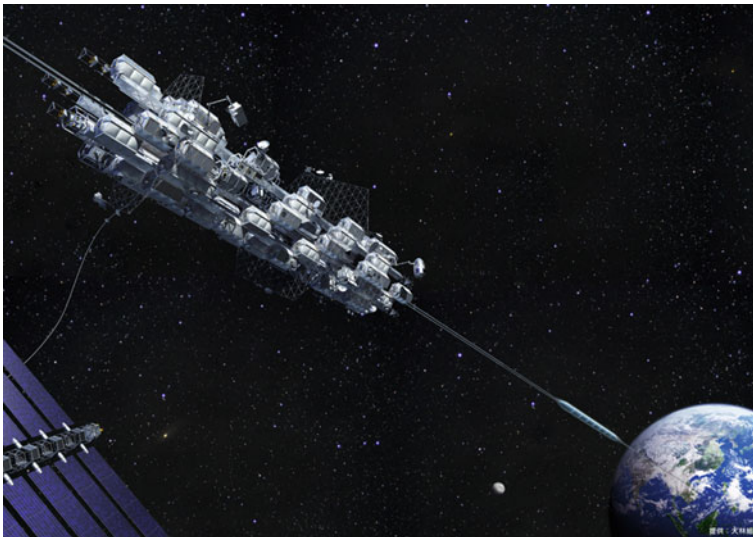


Fig. 4.1 Image of Obayashi Corporation’s space elevator construction concept

4.2.2.2 Components

As mentioned above, a space elevator consists of three main elements: cable, climber, and station. However, Obayashi's space elevator concept includes slightly more complicated components, as shown in Fig. 4.2.

The cable is 96,000 km long and connected to the facility on the Earth's surface, which is referred as an earth port. The cable revolves around the Earth once every 24 hour and, when drawn out, it tends to extend above the equator as a result of centrifugal force. Therefore, it is most efficient to build the earth port in an equatorial region.

The climber moves up and down the cable between the earth port and the top; however, as the climber rises, the gravity felt by the crew in the climber decreases rapidly. The design includes research and training facilities that utilize this low-gravity environment, with a Mars Gravity Center and a Moon Gravity Center at altitudes of approximately 3,900 km and 8,900 km, respectively, that simulate gravity on the surface of Mars (approximately 1/3rd of the Earth's gravity), and on the lunar surface (approximately 1/6th of the Earth's gravity). Such low gravitational fields are difficult to achieve for long periods on the Earth.

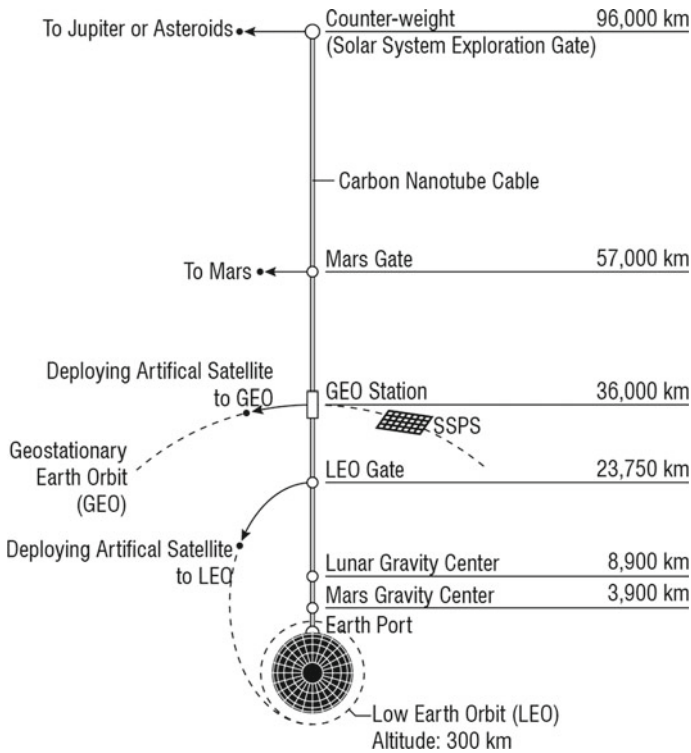


Fig. 4.2 Components in Obayashi Corporation's space elevator concept

When a spacecraft is released from an altitude lower than geosynchronous orbit (approximately 36,000 km), the rotation speed is not high enough for it to escape the Earth’s gravitational field and it will enter an elliptical orbit whose confocal point is the Earth. When released at a height of 23,750 km, an object can reach a height of 300 km at the perigee of its orbit. Thus, a low Earth orbit (LEO) gate is set up at that height to deploy artificial satellites.

In a geosynchronous orbit (approximately 36,000 km), gravitational force and centrifugal force are just balanced out, so a large geostationary orbit station can be constructed here. Geostationary satellites can be deployed into geostationary orbit from here and materials can be transported using the space elevator to space-solar-power satellites.

Once released from the cable at altitudes above approximately 47,000 km, objects can escape the Earth’s gravity. If released from an altitude of approximately 57,000 km, the spacecraft will be able to enter the orbit of Mars, so at this point the Mars Gate is installed. A counterweight is attached to the cable tip at an altitude of 96,000 km to create balance. When released from here, a spacecraft can fly to the asteroid belt and, with a little additional acceleration, to Jupiter and Saturn. The cable tip may be called the Solar System Resource Mining Gate because it is a place to bring back precious metals from asteroids, Jupiter, and the other outer planets.

4.2.2.3 Cable Mechanics

Several types of forces act on the cable, the principal ones being the universal gravitational force (gravity) of the Earth and the centrifugal force (Fig. 4.3).

The universal gravitational force varies in inverse proportion to the square of the distance from the center of the Earth, while the centrifugal force is proportional to the distance. These forces act in opposite directions. The point where the two forces balance out lies at the geosynchronous orbit (the GEO). The universal gravitational force is predominant on below the geosynchronous orbit (Earth side) and the centrifugal force is predominant above it (the space side). Therefore, a significant amount of tension is applied to the cable. The tensile force reaches a maximum at geosynchronous orbit and decreases toward both sides.

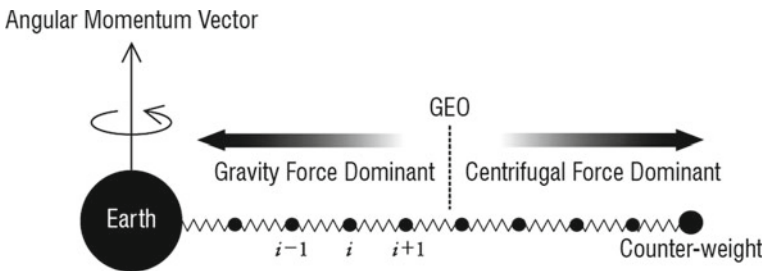


Fig. 4.3 Universal gravitational force and centrifugal force applied to the cable

In the Obayashi concept, the cable material is made of carbon nanotubes and its tensile strength is assumed to be 150 GPa (Demczyk et al. 2002). However, recent discoveries have put forward other materials that might be preferred for tethers: single-crystal graphene and a new 2D category of graphene (Nixon 2020). From an engineering point of view, it is better to make the cable tapered so that the cross-sectional area of the cable is largest at geostationary orbit. Tapered shape can be analytically solved, but here the tension, internal tensile stress, and cross-sectional area were acquired through numerical calculations for a carbon nanotube cable (see Figs. 4.4, 4.5, and 4.6), by considering the masses of all the stations and assuming that the safety factor of the cable is 2. Here, only a fixed (that is, non-moving) load is considered. The two bumpy steps at heights lower than 10,000 km shown in Figs. 4.4 and 4.5 are due to the weights of the Lunar and Mars Gravity Centers.

As the universal gravitational force and centrifugal force balance out, no force is applied to the earth port. However, when a climber (designed to be 100 tons), is attached to the cable and attempts to rise, making it a moving load, the cable is stretched by more than 100 km. In addition, the balance will be lost and the cable will be dragged down. In order to solve this problem and keep the cable stretched constantly, tension (pretension) is applied to the cable in advance in this concept.

In other words, the center of gravity of the cable is shifted slightly to the space side in advance so that the cable is slightly tensioned at the earth port. In addition, a ballast tension control system is introduced at the earth port at the bottom end of the cable so that it can be controlled according to changes in the moving load. Figure 4.7 shows the load balance at geostationary orbit during the initial construction. The difference in the load between the space side and the Earth side at the initial stage is the amount of pretension. At the moment when the first climber grabs the cable, the load at the Earth side increases instantaneously; however, the cable never loses the

Fig. 4.4 Tension on the cable

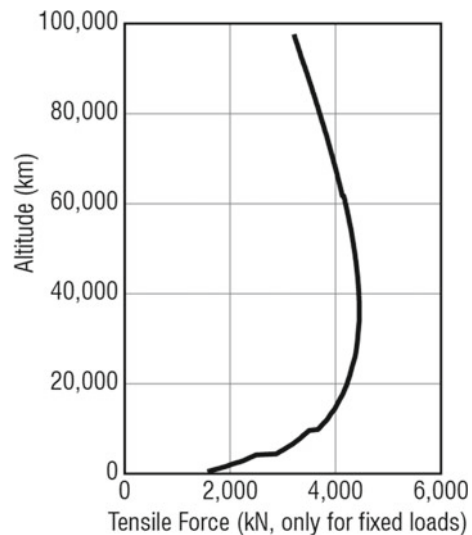


Fig. 4.5 Internal tensile stress of the cable

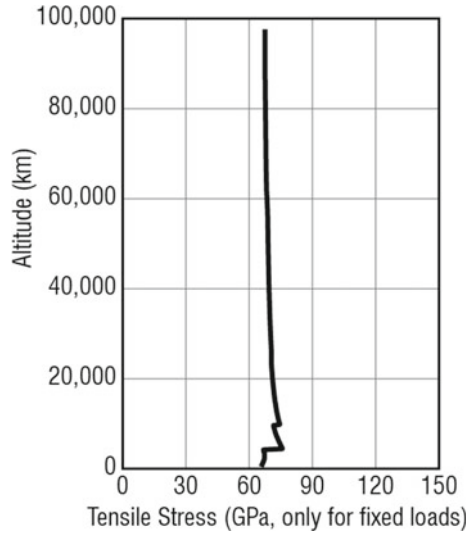
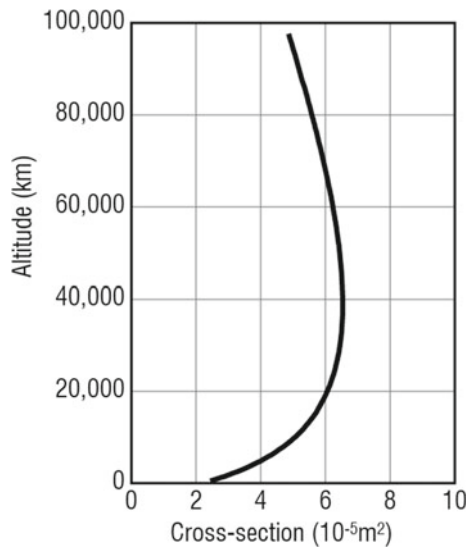


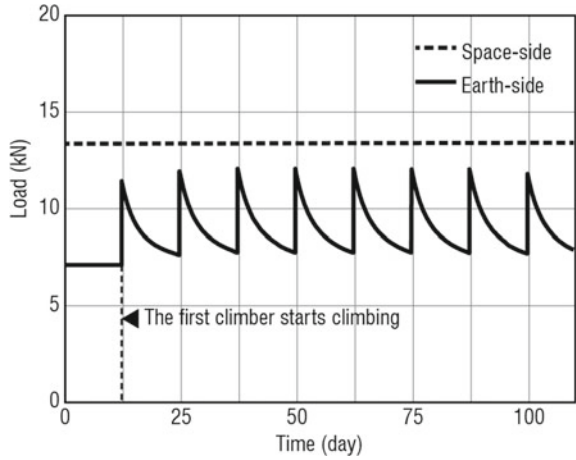
Fig. 4.6 Cross-sectional area distribution of the cable



balance because the Earth-side load does not exceed the load of the space side just as intended. As the climber rises, the load decreases as a result of the decrease in gravitational force, and when it rises to 12,000 km, the next climber can start. This process is repeated every 12.5 days in the construction process, and thus the balance will never be lost.

In addition, since the geostationary orbit altitude is approximately 36,000 km and the next climber can start when the previous one reaches an altitude of 12,000 km,

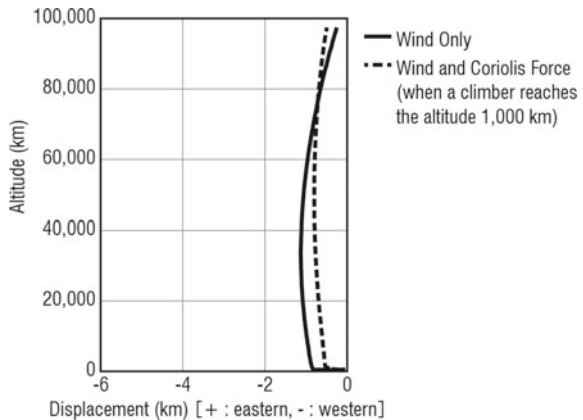
Fig. 4.7 Load balance at geostationary orbit during a construction process (with climbers starting every 12.5 days)



the cable length is set to 96,000 km, which is a multiple of 12,000, so eight climbers may be attached to the cable at the same time. This length was also chosen because it allows spacecraft to fly to most planets in the solar system and also does not resonate with the Moon’s or the Sun’s gravitational fluctuation period.

In this concept, the shape of the cable is assumed to be that of a belt, with a thickness of 1.38 mm and widths of 18 mm on the ground, 48 mm at geostationary orbit, and 36 mm at the tip. A tensile strength of 150 GPa was assumed for carbon nanotubes, which is more than ten times that of steel. Other forces act on the cable as well, namely the moving load of the climber (which varies with altitude as the climber moves), the Coriolis force due to the climber’s movement, the gravitation forces of the Moon and the Sun, the elastic force of expansion and contraction, the air resistance due to wind, and so on. For example, Fig. 4.8 shows the results of the numerical calculation for cable displacement due to wind and Coriolis force.

Fig. 4.8 Cable displacement due to wind and Coriolis force



Typhoons rarely visit the equatorial area, but easterly winds predominate. Furthermore, when the climber rises, a Coriolis force that pushes the cable westward arises, and when the climber descends, it pushes the cable eastward. Figure 4.8 shows that the greatest displacement of the cable due to the wind force and Coriolis force created by an upward climber occurs in the immediate vicinity of the Earth, which ends up displacing the whole cable.

4.2.2.4 Construction Method

The most important process in the construction of a space elevator is the laying of cables. This concept adopts the method of extending cables vertically up and down from geostationary orbit, as described above. However, there is no means to carry cables of 7,000 tons (which is assumed as the final mass in the Obayashi concept) from the Earth's surface to geosynchronous orbit. Thus, in this concept, the following two steps are employed.

In the first step, a cable of 20 tons (the maximum mass that can be carried by a single rocket) is transported to geosynchronous orbit and then extended vertically upward and downward. The lower end, once it reaches the Earth's surface, is fastened to the ground and the upper end is kept at an altitude of 96,000 km.

In the second step, using this cable, 510 climbers climb one after another and attach additional cables to the previous one to make the cable gradually thicker. To maintain the balance of the cables, the mass of the cable that can be reinforced at any one time is only 1.15% of the previous cable mass, and this process must be repeated 510 times using the 510 climbers. Each reinforcing climber joins the counterweight one by one at the top of the cable; however, their mass must be limited so that the mass ratio of the cable to the counterweight is always 1:0.920 in this case. After the cable has been completed, climbers carry the materials to construct stations via the cable.

4.2.2.5 Details of Each Component

Earth Port

The earth port is a facility for connecting cables to the Earth. There has been discussion about whether it should be built on the ground or on the sea, and both have advantages and disadvantages. This concept, however, proposes to use both. On the land, facilities such as airports and hotels that require a large area but a lower security level are built. On an offshore floating structure, major facilities such as a departure/arrival site and quarantine facilities are constructed. Floating structures are located 10 km from the coast, and underwater tunnels are used instead of bridges to travel to and from land because the tunnel can be moved if necessary.

The offshore earth port (hereinafter referred to as the offshore facility) is shown in Fig. 4.9. The earth port consists of an upper building and a concrete floating body.

The upper building has a circular plane with a diameter of 400 m and is equipped with buildings such as a departure/arrival site. The total floor area is set to be 270,000 m² and the number of workers is 5,000. The total weight is approximately 4 million tons, which is several times larger than the largest tanker in the world. The concrete floating body is a hollow structure and this concept adopts a semi-submersible type that is suitable for open ocean environments with long periods and high waves. A box-shaped floating body called a hull is submerged below the water level and a hollow pillar and a deck are installed on it. The hull is moored by legs connecting to suction anchors fixed on the seabed.

A cylindrical space is provided in the center of the earth port and space elevator cables are tied at the bottom. This space is designed to be large enough to house the climber, which is assumed to have a length of 144 m, as described later.

The cable is pre-tensioned as described above. The tension applied to the cable changes as the climber rises and the tension is designed to be actively controlled. A tension control device utilizes the weight of seawater by pumping seawater in and out of a ballast tank. Since the maximum tension during operation is estimated to be approximately 400 tons, there is no need for a large ballast tank.

Geostationary Orbit Station

Since a weightless environment is created at geostationary orbit, the station installed here does not need to be moored to a cable and, as a result, it has no mass limit. The geostationary orbit station is shown in Fig. 4.10. It has a long vertical shape that is stable in space and it can act like a train station platform when a six-car climber arrives. The facility is a combination of 66 hexagonal column-shaped modules of the same shape and size. This modular structure has some advantages: expandability such that it can work efficiently even with one or two modules and the number of modules can be increased even after completion, and redundancy so that the whole structure can be maintained even if one is damaged. The overall shape is designed based on a triple helix, allowing the crew to freely move between modules.

A large-size station such as this is critical for the construction and operation of solar-power satellites, zero-gravity manufacturing, and receiving and transferring minerals and other resources from the Moon and other planets. A space-elevator station such as this is straightforward and affordable and allows full use of space resources. This is central to all discussions on the future of space resources.

Climber

The design of vehicles like the climbers does not come within a construction company's field, so they have not been designed in detail. Nevertheless, specifications are given: a mass of 100 tons, six-car configuration, a full length of 144 m, payload (cargo) of 70 tons, and crew of 35.

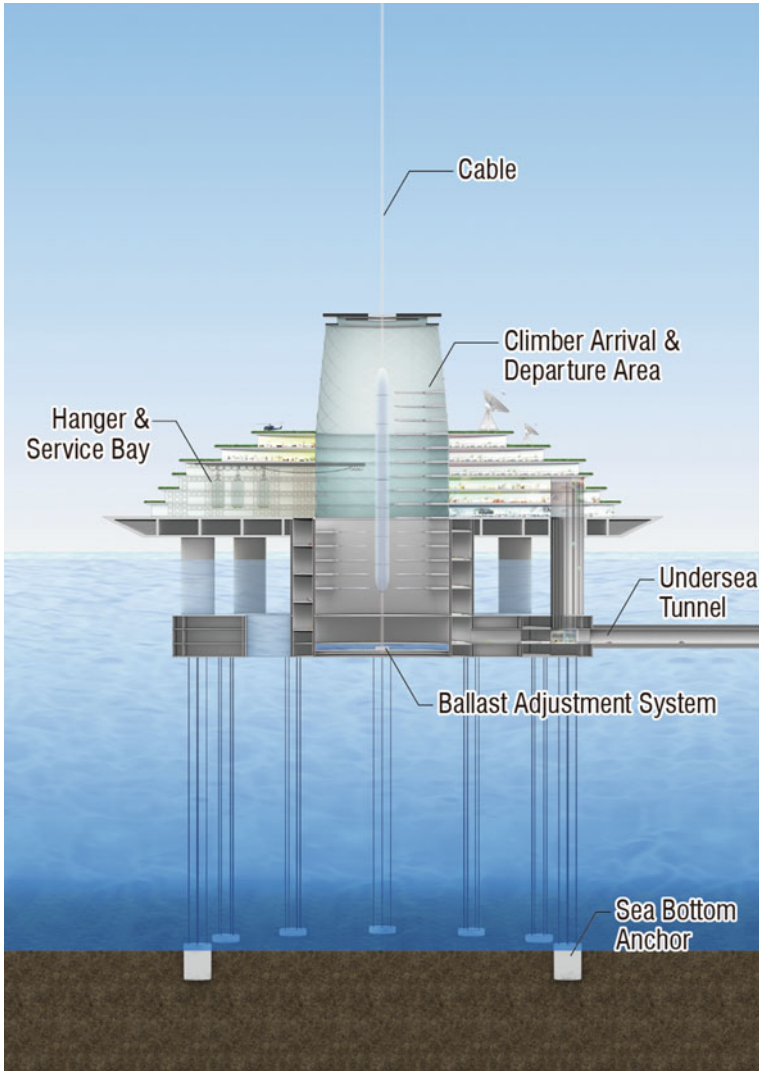
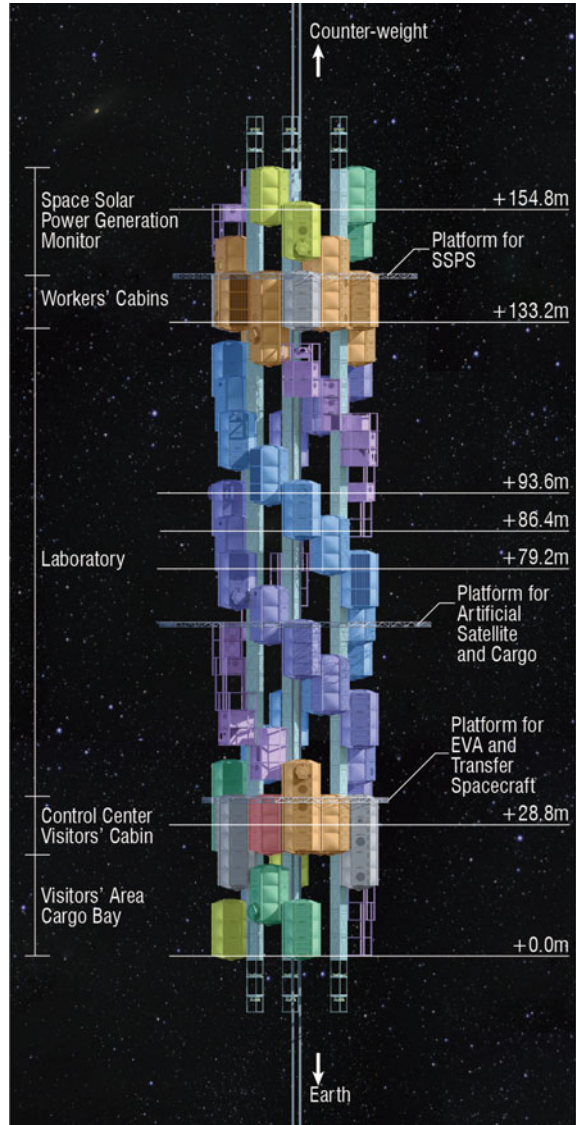


Fig. 4.9 Earth port offshore facility (upper building diameter 400 m)

4.2.2.6 Construction Schedule and Construction Costs

If the construction of the cable started in 2030, then construction of the entire space-elevator system with stations could be completed in 2050, following which the operation might start. The initial construction cost is estimated to be approximately USD 100 billion.

Fig. 4.10 Geostationary orbit station



4.3 Lunar Space Elevator

In general, a space elevator has its center of balance at a geosynchronous orbit of the planet, and from that point a cable is extended upward and downward and becomes the basic component of the space elevator. The lower part of the cable below geosynchronous orbit is dominated by the gravitational force of the planet and the upper part is dominated by the centrifugal force generated by the rotation

of the planet, and these forces create tension in the cable, causing it to stay upright. By raising a payload along the cable and releasing it at a specific altitude with the planet's rotational speed, it can be transported from the surface of the planet to other planets, and vice versa. On the Earth and Mars, space elevators will take this form.

In the case of the Moon, however, the Lagrange points are selected as the center of balance of a lunar space elevator (Pearson 2005), namely the L1 and L2 points among all five Lagrange points. In other words, a cable centered on L1 is basically a cable that extends from the equator of the Moon toward the Earth, while one centered on L2 is a cable that extends from the equator of the Moon to the opposite side of the Earth. The cables of the lunar space elevators must be several times longer than the one on Earth. However, the tension on the cables is not as strong, so special materials such as carbon nanotubes are not necessary and common commercially available materials such as carbon fiber, polybenzoxazole fiber, and extended-chain polyethylene fiber are sufficient. However, since the lunar space elevator enables mass transportation but its transportation speed is slow, it is considered to be more like a pipeline. Therefore, it is not suitable for human transportation to the Moon, which may still require conventional rockets.

Water is an important resource on the Moon and exists at higher latitudes and in polar regions. Therefore, to secure the path connecting such areas and the L1 point, it is proposed that, besides the cable installed between the equator and L1 to maintain tension, another cable be hung diagonally from L1 to a high-latitude or polar region. This makes it possible to utilize resources not only in low-latitude areas but also in high-latitude and polar regions.

Lunar space elevators will allow lunar resources to be used in orbits around the Earth. The lunar regolith can be used as a shielding material or construction material and lunar water can be used as rocket fuel to travel from there to Mars and the asteroids. Lunar space elevators can also be used as a means of supplying materials from the Earth for the construction of lunar bases.

4.4 Martian Space Elevator

On Mars, resources can be found on the surface, underground, and on the moons. Those on the surface and underground include various ones necessary for supporting human life, such as carbon dioxide and water, so it is considered that they will first be used for the inhabitants of Mars. In addition, transporting these resources to the Moon and using them there may be economically advantageous compared to transporting them from the Earth. The moons of Mars, Phobos and Deimos, are also full of useful resources. These resources can be transported to and used on the Martian and lunar surfaces. Mars, the third planet in the solar system, revolves around the Sun just outside the Earth's orbit and can serve as a base camp for mining of resource-rich asteroids. Therefore, from Mars and the Martian satellites, both humans and cargo

can travel to the Moon and the asteroid belt. Thus, with a space elevator from the Martian surface to orbit or through the moons of Mars, this would be a useful space transportation system.

For the construction of space elevators on Mars or the use of tethers, it is better to use the moons of Mars. There are two satellites: Phobos at an altitude of approximately 6,000 km and Deimos at an altitude of approximately 20,000 km. For comparison, the altitude of a geostationary orbit of Mars is approximately 13,600 km.

One way to use tethers on Mars is to use the two moons (Penzo 1984). A tether with a length of 940 km is deployed above Phobos and a tether with a length of 1160 km is deployed below the moon. Similarly, a tether with a length of 6100 km will be deployed above Deimos and a tether with a length of 2960 km will be deployed below the moon. Once a space carrier launched from the surface of Mars reaches low Mars orbit (LMO) at an altitude of approximately 400 km, it deploys the payload directly upward, and it is connected to the space carrier with a tether. This means the carrier and the payload will orbit Mars together at the same angular velocity but at different altitudes while being connected by the tether. If the length of the tether is 375 km and the payload is separated from the tether when it is in phase with Phobos, then the payload will orbit while increasing its altitude. By contrast, the space carrier loses momentum equal to the amount of momentum gained by the payload and orbits while lowering its altitude. The payload reaches the lower end of the tether hanging downward from Phobos after raising its altitude. Then, the captured payload is pulled upward along the tether and, when it reaches Phobos, it shifts to the tether extending upward from Phobos and then is pulled further upward. When it reaches the top of this tether, the payload is released. The payload begins to orbit Mars again, rising in altitude, until it reaches the bottom edge of the tether that hangs down from Deimos. The payload then climbs up this tether, passes the body of Deimos, and then reaches the top of the tether. The payload is released from the top and has already gained the speed to escape Mars. It is worth adding that, if the reverse process is followed, then a payload arriving from the Earth or the Moon into Mars orbit can be economically and safely transported onto the surface of Mars with a soft landing. The economic benefits are calculated as follows. The impulse velocity from the Martian surface to LMO is approximately 3.74 km/s with or without a tether. The impulse velocity from Mars to the Earth is approximately 3.0 km/s. However, from release by a space carrier at LMO to the height of the top of the Deimos tether, the escape velocity using the tether is 3.83 km/s less than it is without the tether. That is, by using the tether system, the corresponding amount of fuel can be saved. In this tether system, the material of the tether may be Kevlar, which is strong enough to withstand the tension.

The second method for using a tether is to use only one of the moons of Mars (Weinstein 2003). The system using Phobos is designed as follows. Cables are deployed upward and downward from Phobos, each with a length of approximately 6,000 km. The bottom end of the lower cable reaches the upper limit of the Martian atmosphere, at an altitude of approximately 60 km. Phobos is located at an altitude of 6,028 km and orbits Mars at a speed of 2.15 km/s. The rotation speed at the top of the upper cable is 3.52 km/s and the rotation speed at the bottom end of the lower cable is

0.77 km/s. On the other hand, the surface of Mars is moving at a speed of 0.25 km/s in the same direction as the Phobos space elevator, so the relative speed between the bottom end of the space elevator and the surface of Mars is only 0.52 km/s.

To make a trip or send materials to the Earth, the Moon, or the asteroid belt, a spacecraft departs the surface using rockets and then latches onto the bottom end of the lower cable of the space elevator. Since the space elevator passes over the same point on the surface of Mars twice during the Martian day, there are two opportunities daily to depart from Mars using the space elevator. Crew and cargo carried by the spacecraft are transferred to a capsule at the bottom end of the space elevator that ascends further. The capsule may be self-propelled or lifted by using motors to drive loop lift cables in increments of about 100 km. The crew and cargo are transferred to other capsules that ascend the upper cable that extends upward from Phobos until they reach the top of the tether. Finally, they board an interplanetary transfer vehicle, which is launched into space at a speed of 3.52 km/s. This velocity is more than the 2.6 km/s hyperbolic velocity and is equivalent to the Hohmann elliptical transfer velocity necessary to reach the Earth's Moon. This velocity is also approximately equal to the velocity required to reach the asteroids inside the asteroid belt. Thus, this space elevator enables interplanetary flight to the Earth, Moon, and asteroids without relying on rocket propulsion.

A human presence on the Moon and on Mars is critical for expansion outward into the solar system to use its limitless resources. Thus, complete bases need to be established on both worlds. The resources acquired on other planets can be utilized for the infrastructure of these manned facilities.

4.5 Space Elevators on Asteroids

As discussed in other chapters in this book, asteroids are known to be rich in a variety of resources. From the perspective of space-resource mining, space elevators are the most valuable on asteroids. Mining is autonomous and resources mined by space elevators attached to asteroids can be transported to the Earth, Moon, and Mars. Many asteroids will have important and valuable resources and an appropriate spin rate, and the gravity of asteroids is small, so the material for space elevator cables does not need to be special. Reusable climbers can lift raw and processed resources along cables and release them at the right altitude and at the proper time to send them to the Earth, the Moon, or Mars. It is a very economical and profitable system. For space-solar-power satellites, which are planned to be constructed in Earth's geosynchronous orbit, it is also more efficient to use resources such as metals mined on asteroids rather than to transport materials from the Earth's surface (Edwards 2020).

4.6 Conclusions

When a space elevator is constructed on Earth, transportation from the Earth to the Moon and other planets will be achieved at a lower cost than in conventional ways and with more massive payloads. Obayashi's design includes the transportation of both people and massive payloads with operations to begin in 2050. In addition, there are several other efforts focusing on less capable prototypes to validate the engineering concept with cargo operations sometime prior to 2040. At the same time, transportation from the Moon, Mars, and asteroids to the Earth will be achieved at a lower cost, which will greatly contribute to the promotion of resource mining. Building a space elevator with a total length of approximately 100,000 km will reduce the cost of traveling to and from almost all points in the solar system. In addition, it is intended to construct space elevators not only on the Earth but also on the Moon, Mars, and asteroids for reduced travel costs when arriving at and departing from the surfaces of these worlds. In this case, transportation between the Earth, the Moon, Mars, the moons of Mars, and asteroids will be made possible by using space elevators at each location. While constructing a space elevator will take tremendous effort, when it is completed, it will contribute greatly to the resource mining of the solar system.

References

- Artsutanov, Y. 1960. To the cosmos by electric train. English translation from Russian original.
- Demczyk, B.G., Y.M. Wang, J. Cumings, M. Hetman, W. Han, A. Zettl, and R.O. Ritchie. 2002. Direct mechanical measurement of the tensile strength and elastic modulus of multiwalled carbon nanotubes. *Materials Science and Engineering A* 334: 173–178.
- Edwards, B.C. 2000. Design and deployment of a space elevator. *Acta Astronautica* 476 (10): 735–744.
- Edwards, B.C., and E.A. Westling. 2002. *The space elevator, A revolutionary Earth-to-space transportation system*, Spageo.
- Edwards, B.C. 2002. *The space elevator*. NIAC report.
- Edwards, B.C. 2003. *The space elevator*. NIAC phase II final report.
- Edwards, B.C. 2020. Private communication, November 4.
- Iijima, S. 1991. Helical microtubules of graphitic carbon. *Nature* 354: 56–58.
- Ishikawa, Y., T. Tamura, K. Otsuka, T. Horiike, T. Iwaoka, N. Masui, K. Hamachi, S. Katsuyama, Y. Aoki, Y. Yamagiwa, F. Inoue, and E. Omoto. 2016. Obayashi Corporation's space elevator construction concept. *Journal of the British Interplanetary Society* 69 (6/7): 227–239.
- Ishikawa, Y., T. Tamura, K. Otsuka, T. Horiike, T. Iwaoka, N. Masui, and K. Hamachi. 2012. The space elevator construction concept. *Kikan Obayashi (Obayashi Quarterly)* No. 53 "Tower," Obayashi Corporation (in Japanese).
- Ishikawa, Y., T. Tamura, K. Otsuka, T. Horiike, T. Iwaoka, N. Masui, K. Hamachi, S. Katsuyama, and Y. Aoki. 2013. The space elevator construction concept. IAC-13-D4.3.6, *64th international astronomical congress (IAC)*, Beijing, China.
- Nixon, A. 2020. Graphene: the last piece of the space elevator puzzle. ISEC Webinar www.isec.org, May 29, 2020, Swan, P., C. Swan, M. Fitzgerald, M. Peet, J. Torla, and V. Hall.

- Pearson, J. 1975. The orbital tower: a spacecraft launcher using the earth's rotational energy. *Acta Astronautica* 2: 785–799.
- Pearson, J., E. Levin, J. Oldson, and H. Wykes. 2005. Lunar space elevator for cis-lunar transportation. *International conference, moon base: a challenge for humanity, Venice workshop, 26–27 May 2005, Venice, Italy.*
- Penzo, P. 1984. Tethers for Mars XE “Mars” space operations. *The Case for Mars II, AAS, Science and Technology Series* 62: 445–465.
- Swan P.A., D.I. Rait, C.W. Swan, R.E. Penny, and J.M. Knapman (ed.). 2013. *Space elevators: An assessment of the technological feasibility and the way forward*, International Academy of Astronautics.
- Swan P.A., D.L. Raitt, J.M. Knapman, A. Tsuchida, M.A. Fitzgerald, and Y. Ishikawa (ed.). 2019. *Road to the space elevator era*. International Academy of Astronautics.
- Swan, P.A., C. Swan, M. Fitzgerald, M. Peet, J. Torla, and V. Hall. 2020. Space elevators are the transportation story of the 21st century. ISEC study report, www.lulu.com.
- Weinstein, L.M. 2003. Space colonization using space-elevators from phobos. *AIP Conference Proceedings* 654: 1227.

Chapter 5

Orbital Hub: Providing an LEO Infrastructure for Multi-disciplinary Science and Commercial Use Cases



Volker Maiwald, Dominik Quantius, Claudia Philpot, and Vincent Vrakking

Abstract Skylab, Saljut, MIR, the International Space Station, and Tiangong have been space stations in low-Earth orbit (LEO), allowing access to a microgravity environment for scientific or technology demonstration experiments. Future applications planned for using resources in LEO are commercial (e.g., Axiom Space Station, Bigelow Commercial Space Station) or scientific (e.g., the Chinese Space Station). To analyze all possible needs of potential users, the authors have surveyed needs from a commercial and a scientific perspective, based on which a design has been elaborated allowing for a versatile, flexible and cost-effective platform. The *Orbital Hub* can serve as a core unit for a larger complex or act on its own. It consists of a base platform, permanently crewed, and a crew-tended Free Flyer facilitating experiments in an unperturbed environment. This chapter presents the design of the *Orbital Hub* and its capability to be used in combination with other space station concepts or even parts of the ISS, outlining, for example, its complementarity with the Lunar Orbital Platform Gateway and its advantages over larger platforms, such as the ISS. Implications concerning application, LEO resources, costs and key technologies are discussed, showing how a small platform can be utilized to access those resources efficiently.

V. Maiwald (✉) · D. Quantius · C. Philpot · V. Vrakking
DLR Institut für Raumfahrtssysteme, Robert-Hooke-Str. 7, 28359 Bremen, Germany
e-mail: volker.maiwald@dlr.de

D. Quantius
e-mail: dominik.quantius@dlr.de

C. Philpot
e-mail: claudia.philpot@dlr.de

V. Vrakking
e-mail: vincent.vrakking@dlr.de

5.1 Introduction

Operating space stations in low Earth orbit (LEO) began in the aftermath of the lunar landings during the early 1970s with *Saljut* and *Skylab*. Later the Soviet space station *Mir* was the first place for international cooperation on a space station, when NASA sent astronauts to *Mir* for the Shuttle–*Mir* program during the 1990s (see Fig. 5.1). The *International Space Station* (ISS) replaced *Mir* in the late 1990s and has been in operation since, granting humanity access to the “resources”, mostly functional, of low Earth orbit for long-duration missions. In addition, the Chinese space stations *Tiangong 1* and 2 have been testbeds for the human spaceflight program of the People’s Republic and the pathfinder for a larger space station.

The ISS’s end of life is approaching, and although the exact date is still open to debate among the partners, it is expected to occur in the 2020s. The ISS and its predecessors have granted humans the possibility of Earth and Sun observation and provided a micro-gravity environment for scientific experiments and technology demonstration.

While currently the focus of future human spaceflight is directed mostly at the lunar environment, it is unlikely that the use of infrastructure in low Earth orbit will come to an end. Current plans for using resources in LEO are aimed at commercial applications (e.g., *Axiom International Commercial Space Station*) or scientific ones (*Chinese Space Station*).

The use of LEO for commercial, scientific and technical applications requires understanding of the available resources, as well as the needs of users and end users, such as companies or scientists on Earth. Therefore, an initial survey of those needs has to precede any design of a suitable infrastructure. This chapter first presents these customer voices, then describes the *Orbital Hub* which has been designed to meet those needs and has the capability to be used in combination with other space station concepts or even parts of the ISS.



Fig. 5.1 Russian space station *Mir* during approach of space shuttle *Atlantis* during STS-71 (NASA, public domain)

Orbital Hub consists of a crewed main platform and a crew-tended *Free Flyer* facilitating experiments in an undisturbed environment. It is a versatile, flexible and cost-effective platform, which can serve as nucleus for a larger complex or act on its own. It can also complement, for example, *Lunar Orbital Platform Gateway*. The *Orbital Hub*, including the *Free Flyer*, can be launched with a total of four launches of the currently available launchers (e.g., ARIANE 5 or 6, Falcon Heavy) and would house a crew of three. It has development costs of €2.2 billion and annual operation costs of €1.5 billion, assuming a life of at least 10 years. It is assumed to be deployable within four to eight years' time (Jahnke et al. 2018).

Details of the design and design process can be found in Jahnke et al. (2018). This chapter focuses on its utilization and how the design can be incorporated into mission scenarios of the future.

5.2 Low Earth Orbit Resources

Low Earth orbit reaches from about 300 km to about 2000 km. Currently, the ISS is operating at about 450 km orbital altitude. Although no resources appear to exist in this orbital region (no minerals can be harvested there for production or return to Earth), something in the LEO has been a sufficient incentive for space-station operations there for the past five decades, with almost continuous operation of stations for more than three. This is in addition to the more than six decades of operation of satellites with various applications.

One resource readily available in LEO, except during eclipses, is solar illumination, which can be exploited for power generation. While plans for orbital power plants have existed for some time (Seboldt 2004), they have not yet become a reality, due to various still unresolved technological obstacles, such as transfer of energy from orbit to Earth (Seboldt 2004; Shen et al. 2019). So solar illumination is a resource currently used to sustain spacecraft operation in LEO, but not a justification for it.

However, if under the heading of resources we include anything beneficial, including conditions rather than just objects or materials, then LEO does have resources available. The following sections explain, first, general resources, and then those related only to human spaceflight.

5.2.1 *Micro-gravity Environment*

One such resource is the micro-gravity environment in LEO. Naturally, this resource is abundant in any orbit around a main body, but in LEO micro-gravity is combined with sufficiently easy access in terms of launch costs and communication. It is simply less efficient to conduct micro-gravity experimentation in a Mars orbit.

The micro-gravity environment can be used for physiological experimentation (Oei and Mirra 2008), biological experiments (Kwok et al. 2020) or convection-free

physics (Balter 2019), as well as many other scientific fields (Warren 2020). These environmental conditions can also be used for additive manufacturing, either for self-maintenance (Fateri et al. 2018; O'Hara et al. 2018) or maintenance of satellites in general (Koryanov 2021).

5.2.2 *Space Radiation and Space Observation*

While space radiation may not immediately spring to mind as an experimental resource, such as biological (Sgambati 2020)—usually it is considered an obstacle to be overcome or protected against – it can also be considered a resource unavailable naturally on Earth due to its protective atmosphere. Similar to micro-gravity, space radiation is more easily accessible in LEO. However, it also differs between LEO and deep space, because of Earth's magnetic field, which shields Earth from some radiation.

Just as the Earth's atmosphere protects it from radiation, it is also an obstacle for space observation by perturbing or absorbing electromagnetic radiation relevant for measurements. An LEO mission can circumvent that obstacle by placing the respective instrument outside the atmosphere. Solar observation has been part of *Skylab* and ISS (e.g., with the SOLAR experiment). General space observations have been part of LEO activities, for example, the famous *Hubble Space Telescope*.

5.2.3 *Earth Observation*

LEO is also particularly useful for observation not just of space but Earth. Earth observation from space has advantages over non-spaceborne methods, including:

- accessibility of regions all over Earth with one instrument and with little delay, not only reducing effort of measurements, but also ensuring comparability of measurements;
- measurements without restriction due to politics or logistics, e.g., measuring CO₂ emissions;
- global view of Earth for measurements, e.g., increases in sea level.

Numerous satellites are used for Earth observation—commercial, scientific and institutional. Examples of experiments on ISS include experiments for atmospheric measurements, such as Atmosphere–Space Interactions Monitor (Nature 2021) or for gravity waves (Magalhães et al. 2021).

5.2.4 *Human Crew*

Most, if not all of the previously mentioned resources can be used with autonomous spacecraft not requiring a crew. The only location in space where human crew is currently available is LEO. While there are plans for human spaceflight in the lunar environment, which shares some of the resources, actual implementation is currently only present for LEO. Even if human spaceflight expands to other regions, the LEO remains the least expensive.

The human crew can have two main functions:

- (1) workforce, or
- (2) experimental subjects.

A human crew can be used to keep experiments technically simple—scientists can focus on their actual experiment during design, not on its automation. In particular, experiments where manipulation is required, such as for manufacturing, become easier when a human crew is actually operating them rather than remote control and automation. The ability to maintain an experiment and the carrying spacecraft also enables longer durations of experimentation if required. A human crew can even allow a space station to become a facility for maintaining or disposing of spacecraft (Koryanov 2021).

The advantage of a human crew over automation and remote control can be estimated using a simple comparison. During the *Apollo* missions, the human crews used the lunar rover to traverse about 20 to 40 km on the lunar surface in a matter of days. For a similar distance the automated rover *Opportunity* required ten years on the surface of Mars, decades later with more advanced technology.

The human crew is usually also part of experiments linked to human physiology, ranging from general health issues to future human spaceflight missions (Crucian et al. 2020).

5.2.5 *Exclusiveness of Location: Tourism*

The LEO is a very exclusive location—and other regions of space even more so. In total, depending on the exact definition of space, a maximum of 579 people have been to space.¹ This compares with the 5,788 people who have climbed Mount Everest.² By this measure, space is at best about a tenth as accessible as the tallest mountain on Earth. The exclusivity of this location, accompanied by an incredible view of Earth and space, makes space attractive for elite tourism. Dennis Tito was the first space tourist, paying for his own seven-day stay on *Mir* in 2001; others followed.

In summary, we can say that LEO offers various non-material resources that can be functional or environmental.

¹ <https://www.worldspaceflight.com/bios/stats.php>, accessed 4 May 2021.

² <https://haexpeditions.com/advice/list-of-mount-everest-climbers/>, accessed 4 May 2021.

5.3 Current Plans for Stations

This section summarizes plans for other space stations to provide a comparison with the *Orbital Hub* design. The resources or function(s) to be covered by the stations are also described.

Currently, the ISS is the only space station in the LEO, operated by an international consortium of partners. While four ISS partners (NASA, ESA, JAXA, CSA) are planning a crewed outpost in a lunar environment, other entities also have plans for LEO stations. However, the details of these plans are often scarce, as business models are vague, and national or international plans for space exploration are often subject to changes in the political landscape. For instance, the USA's Constellation program, initiated by President Bush, was targeted at the Moon and ran from 2005 to 2010, before being cancelled by President Obama. In 2017 President Trump initiated the Artemis program, once more targeted at a lunar landing.

5.3.1 Chinese Space Station

The *Chinese Space Station* (CSS), launched in April 2021, recently began operations (Wall 2021). The finished station will consist of three modules (Jones 2021): the launched core module *Tianhe* (Wall 2021) and two experiment modules (Xinhuanet 2020). Two predecessors, *Tiangong-1* and *Tiangong-2* are the basis for the design of the modules. Assembly is expected to be completed at the end of 2022 (Wall 2021). The launch vehicle for all launches will be the Long March 5B (Jones 2021) and the station will have a mass of about 60,000 kg and support a crew of up to three astronauts for continuous operation (David 2020).

The science onboard will consist of observation, biology and micro-gravity experimentation. An overview of the space station's core module is given in Fig. 5.2. China has not published any plans for space tourism on CSS.

5.3.2 Axiom International Commercial Space Station

Axiom Space is a company which is targeting several fields of LEO operations including space tourism and manufacturing. As an initial step for establishing a space station in LEO, the company plans to attach a commercial segment to ISS, expected to begin service in 2024 (Foust 2020). An independent space station, based on these early components, is planned for 2024–2028 (Wall 2020).

The main purpose of the later station will be manufacturing products intended for terrestrial use but also use on orbit, reducing costs of future missions by providing components which do not have to be launched into space. The station will consist only of rigid modules, manufactured by Thales Alenia Space (Wall 2020a). No further details are currently available.

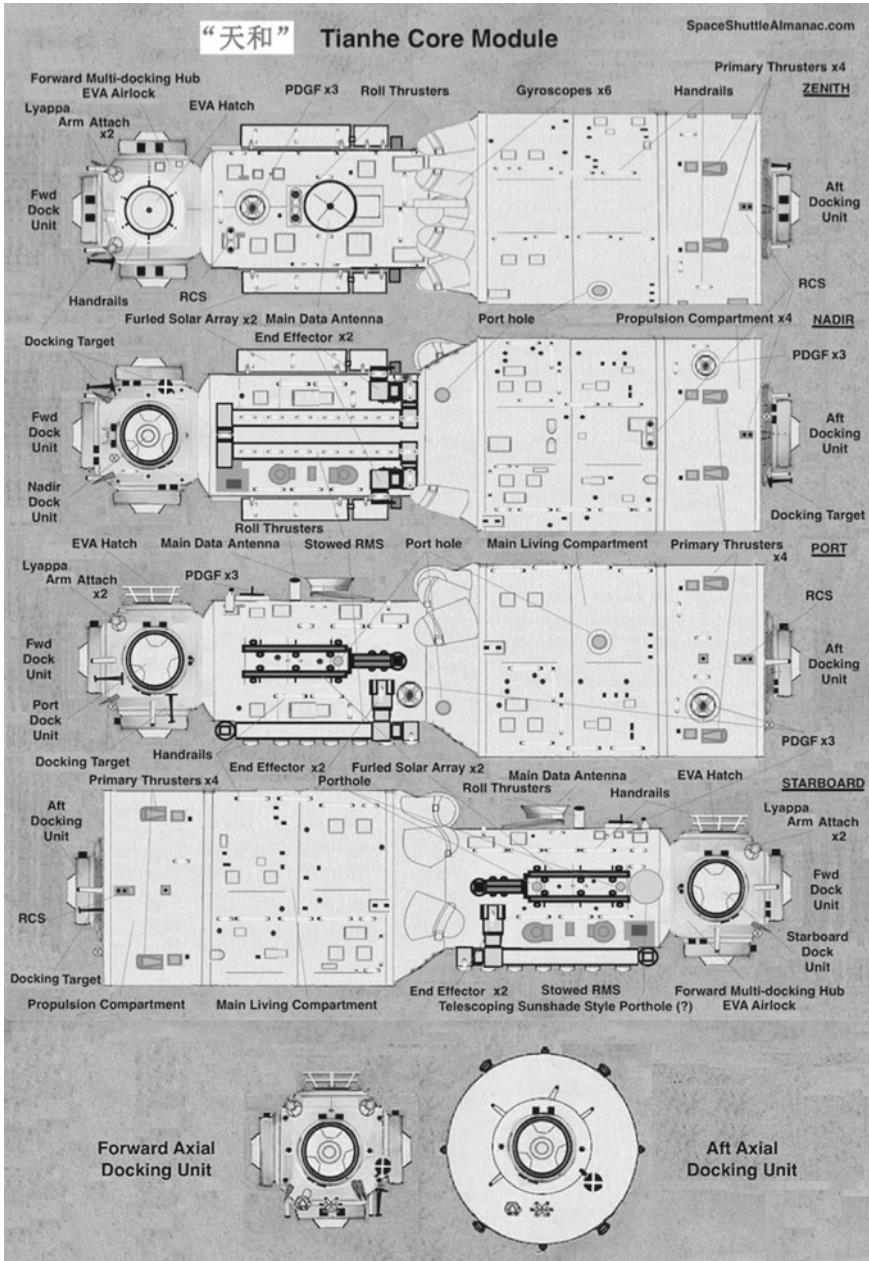


Fig. 5.2 The core module of the planned Chinese space station with docking ports for additional modules and spacecraft (Brandon-Cremer 2020)

5.3.3 *Bigelow Next-Generation Commercial Space Station*

The Bigelow Next-Generation Commercial Space Station has been a changing concept for at least a decade. Plans as to the exact layout of the system are as yet unclear. However, it will be based on technology for expandable modules originally developed by NASA. The Bigelow module BA-330 is designed to operate independently and thus form a space station, but can also be attached, for example to ISS (Wall 2020b).

The station's technology has been tested with the *BEAM*-module at ISS. It could also be used as habitat for human exploration beyond LEO, but in LEO it could serve for tourism and science alike (Wall 2020b). Launch and assembly of the station has been postponed several times and currently no definitive plan has been established. The most recent plan targets using the Vulcan launch vehicle by United Launch Alliance (ULA 2017), currently still in development.

5.3.4 *Gateway Foundation*

The Gateway Foundation is planning to create space stations in LEO, with the ultimate goal of establishing the first true spaceport acting as research station, hotel and transfer node for missions beyond LEO, called simply *Gateway* (Williams 2020).

The *Gateway*'s predecessor, currently expected to be a smaller space station, was previously named *Von Braun Space Station*, after the prominent figure of spaceflight, and is now called the *Voyager Class Space Station* (Gateway Foundation 2020). This station will have a wheel-like configuration with a central node for docking and a ring structure consisting of several modules containing the actual habitat volume, able to accommodate up to 450 people (Gohd 2020).

The station will support scientific undertakings but also act as a hotel accommodating space tourists. The wheel configuration allows artificial gravity to be established with the help of rotation, with micro-gravity environments in non-rotating parts. This can benefit science and tourism alike (Gohd 2020).

The duration of stay is intended to range between days and weeks and the target group are people with an average and above (from a US perspective) income. Gateway expects the station to be completed by 2027 (Gohd 2020). Again, details of the launch plan, construction and assembly have not yet been published.

The next step will be the *Gateway*, which will have a capacity of 1,400 people aboard and a diameter of almost 500 m (Gateway Foundation 2020). It will consist of several rings, supplying artificial gravity by rotation, but also central areas without artificial gravity. The two inner rings will rotate at a speed allowing gravitation akin to that on the lunar surface; the outer ring will provide Mars-level gravitation.

The ring modules will have a diameter of 12 m and a length of 18 m. The modular nature will allow versatility with different purposes for each module (Williams 2020). The designs of the Gateway Foundation resemble the more ambitious projects currently envisioned for human spaceflight in Earth orbit and beyond.

5.3.5 Summary

As can be seen, a majority of the plans for future space stations include experimentation in their portfolio, that is, they will be exploiting the micro-gravity environment and observation opportunities. As far as commercial application is concerned, manufacturing in space and space tourism are also planned, but not for all stations.

The only station which is already ahead of the planning or testing phases is CSS, with its first module already deployed. However, none of the published information for any station is detailed enough to ascertain exactly which resources will be utilized on them.

5.4 User Needs for LEO Outpost

Orbital Hub was designed using the concurrent engineering method, considering user needs as well as lessons learned from previous LEO stations. Since only scientific outposts have existed in LEO to date, the user needs identified also focus on this application. However, they are not specific to *Orbital Hub*.

5.4.1 Scientific Point of View

A survey within the German scientific community with experience in space experiments (physics of soft condensed matter, material science, Earth observation, atmospheric physics, space medicine and physiology, radiation dosimetry, astrobiology, radiation biology, gravitation biology, astrophysics, robotic and exploration, technology demonstration) identifies the following typical future needs from an operational point of view (Jahnke et al. 2018):

- (1) Crew flexibility/higher degree of self-direction
- (2) Safety/clean zones within the station
- (3) Direct communication between crew, station experiments and scientists on ground
- (4) Crew selection and scheduling based on scheduled experiments
- (5) Emphasis on internal station logistics

More details on how the survey was conducted can be found in Jahnke et al. (2018). For brevity, only those survey results that form the basis for the subsequent station design are given here.

The need for an LEO station which is more flexible and more specifically tailored to scientific and commercial activities implies a requirement for stakeholder input, particularly from potential users (e.g., scientists), in the early design phases of a project.

Crew Flexibility and Higher Degree of Self-direction

At the moment, crew activities onboard the ISS are meticulously predefined and approved by ground support. A tight schedule of work, physical training and leisure time is set for ISS expeditions, greatly limiting the astronauts' freedom and affecting their stress levels. This determinism, compared to a regular laboratory on ground, can prohibit creative solutions on board. A higher degree of self-direction for the crew is envisioned whilst ensuring safety standards are met with specific "flight rules". Relaxing this tight schedule would potentially increase crew well-being and productivity.

Safety and Clean Zones

Currently, all station areas have similar standards of safety and cleanliness. High levels of cleanliness and safety result in financial costs and time costs for the crew. Various zones within the station could have different cleanliness and safety standards, reducing cost and time effort. Experimentation areas should be divided into zones depending on the cleanliness requirements, and the crew space should be removed from these areas to allow a more relaxed, comfortable and less sterile living environment for the astronauts, with mental health benefits for the crew. Separation of these areas could be achieved with the use of air locks. However, the design should also be cost-effective, aiming to use as little space as possible and cleverly distributing station systems to achieve zoning but not extending the station's volume and thus costs.

Direct Communication Between Crew, Station Experiments and Scientists on the Ground

Currently, Capcom/Eurocom communication protocols prevent direct discussions between the crew and scientists. This renders support by scientists for the crew less effective, and similarly, reporting of results is cumbersome. This increases the amount of time needed, affecting crew time and mission costs. Capcom/Eurocom communication protocols between the station crew and ground should be relaxed. A direct link between the scientists involved and the crew would facilitate exchange of information on experiments and results, making experiments faster and more effective.

Even without crew involvement, direct communication between automated experiments and scientists would improve the handling of experiments. Increased involvement by scientists in communication concerning experiments would also reduce workload at the operation centers controlling the station. Safety measures within these communications, e.g., when non-routine activities occur, need to ensure that the station is not at risk. This could be achieved by moderation and supervision of communication.

Crew Selection and Scheduling Based on Scheduled Experiments

Currently, experiments and crew qualification/experience are usually not aligned. This affects stress levels (the astronauts need to work on familiarizing themselves

with the topic and equipment) and quality of the work. To counteract this, crew selection and experiment selection could be better aligned. The crew selection process for a given mission and its scheduling during the mission should be adapted more to the experiments planned, enabling a specialist crew to support experiments more fully. This would, however, imply a higher frequency of crew exchanges, increasing costs. More direct communication as previously described, however, could enable suitable selection of crew according to the experiments, without incurring similar increase in costs.

Emphasis on Internal Station Logistics

Astronaut feedback suggests that a major obstacle to effective work is the convoluted internal logistics onboard the station. For instance, one crew might unload a cargo transport with equipment intended or needed (e.g., spare parts) for another crew member. Storing and locating equipment is done by different crews, adding to crew maintenance time, which could be reduced if tools, equipment and parts—for station and experiment maintenance, installation and operation—were easier to locate during a mission. Implementing passive and active markers (e.g., via radio-frequency identification) could allow a clear inventory to be established and items positioned aboard accordingly. This could reduce crew time spent on locating items and manually maintaining inventories.

5.4.2 *Non-scientific Point of View*

Currently, there is no experience concerning space stations outside the realm of scientific application. However, the lessons learned all affect efficiency, and if there is any difference at all between the commercial and scientific points of view, then efficiency is an even more pressing issue for commercial stations.

For instance, if crew flexibility has an effect on productivity, then this should be even more relevant to a commercially operating station. The same is true for crew selection and improved station logistics. Differentiating between work and non-work areas is certainly an asset for stations targeting tourism to improve the quality of the experience for any guests.

Direct communication will be even more necessary for commercial endeavors. Direct communication reduces costs and at the same time confidentiality is ensured (assuming proper protection of the communication channels). A model for a commercially available service in space allowing direct communication is the *Bartolomeo* external platform of the ISS *Columbus* module, which was launched in January 2021. The platform allows direct communication between its experiments and the respective user, even via smartphone, significantly reducing the effort of communication (Clark 2021).

In tourist-oriented space stations in particular there will be an emphasis on comfort as well as safety, because tourists will have far less training than astronauts. User needs to be investigated should include acceptable price ranges and market analysis

of expected facilities. However, as space tourism still is a marginal field of application and currently no tourist-related stations have exceeded the concept phase, this will not be further analyzed in this chapter. Other marginal, but still existing applications include using space as filming locations, as has been announced by US and Russian film productions (Shoard 2021).

For a more industrialized LEO infrastructure, operations are expected to be automated as far as possible to reduce crew and ground segment workload and operation costs, so general system support functions will have to become more autonomous. A more efficient, globally centralized and service-oriented infrastructure for operations and training also has to be implemented. Besides fast and regular access, an industry-oriented legal framework, new partnership models and the protection of intellectual property and production-related data is mandatory. The key for the commercialization of space is a much broader awareness within the “non-space” community of the potentials of the given recourses.

5.5 *Orbital Hub* Main Platform

The ESA *Moon Village* concept is the idea of establishing a common undertaking of human activity on the lunar surface, based on the cooperation of multiple partners with different interests on the Moon (Athanasopoulos 2019). Similarly, *Orbital Hub* is supposed to be a cost-effective nucleus for future LEO activities, an example of what is required for continued operation in LEO and utilizing its resources.

Orbital Hub is separated into two parts, the main platform and the *Free Flyer*, as depicted in Fig. 5.3. The main platform is continuously crewed, with nominally three persons and up to six persons during an exchange week, and thus allows for human physiology experiments as well as other activities requiring regular human interaction. It is intended to be a cost-effective platform, able to continue the human presence in LEO, without unnecessary drains on the budget for human exploration beyond LEO. It can serve as a nucleus for larger endeavors or as a template for stations of other institutions with similar purposes.

5.5.1 *Main Platform Design*

The main platform comprises three basic modules covering all relevant functions while allowing for extension of the platform as needed in the future.

The docking node provides the means for vehicles to dock with the platform. It also contains a cupola for visual activities, supporting docking operations and crew leisure. The unobstructed position of the docking node makes it a reasonable position for communication equipment. A life-support system sustains the crew and the available space is used for exercise equipment.

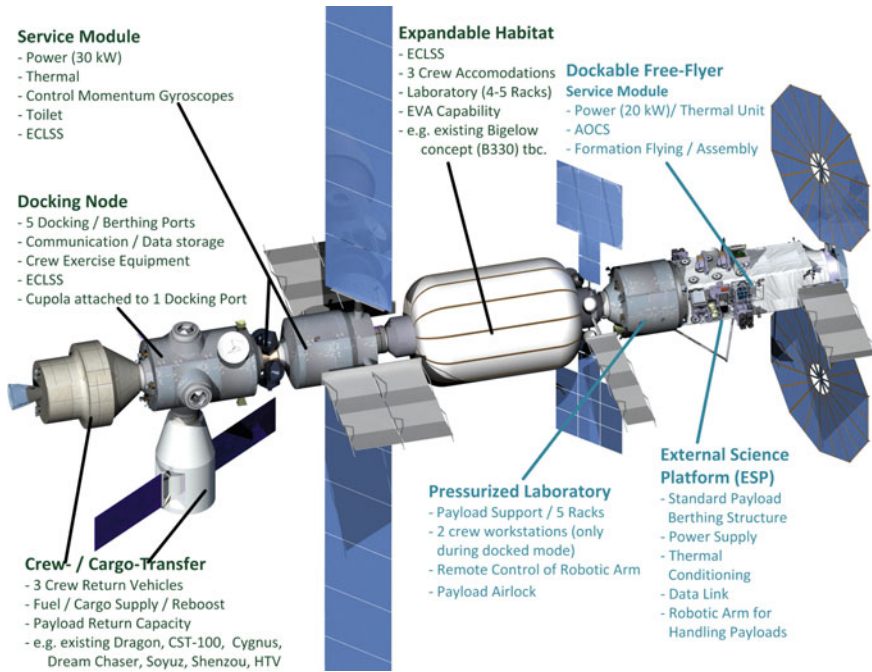


Fig. 5.3 *Orbital Hub*'s configuration, consisting of the main platform (left) comprising the Expandable Habitat, the Service Module and the Docking Node (and some spacecraft for visualization, not part of the platform) and the *Free Flyer* (right). Power values do not include losses, but indicate net power generation

The service module is pressurized and contains the platform's power system: control units, batteries and the solar arrays. It is also equipped with the thermal control system, especially the radiators. Crew waste management (including toilet) and attitude and orbital control system (including control momentum gyroscopes, CMGs) are also located here. A second, redundant life-support system is also installed in the service module to ensure crew safety. Data handling and storage also occurs here.

The main crew accommodation can be found in the habitat, which in the current design is an expandable module, based on Bigelow Aerospace's BA-330 design (as mentioned above). Individual cabins are reserved for the crew quarters and there is a separate meal-preparation area. The center axis of the habitat module has room for experimentation racks, but also for subsystems. They are accommodated in a truss with four lines of racks and a center space, providing room for access and maintenance. While most systems are removed from the crew living area, space and mass limitations do not allow a separate module. However, especially disturbing subsystems, such as the toilet, have been removed from the habitat.

Combined with the docking ring for the *Free Flyer* at the end of the habitat, there is an expandable emergency air lock with the equipment necessary to conduct emergency extra-vehicular activities (EVAs)—in nominal operation no EVAs are anticipated.

5.5.2 Mass and Power Budgets

The overall mass of the main platform is about 65 metric tons, distributed over three modules, which have launch masses of between approx. 18 and 25 metric tons. The mass breakdown per module is given in Table 5.1.

The power budget depends on the actual mode of operation and is summarized in Table 5.2. For the design three major modes have been defined:

- Standard mode: nominal operations, i.e., scientific experiments, crew aboard and orbit maintenance, where necessary.
- Crew exchange: two crews are present on the platform, possibly exchanging experiments, before one crew leaves for Earth.
- Survival mode: main priority is attitude control, survival of crew and experiments and otherwise minimization of power usage, to accommodate a possible loss of solar cell illumination.

Table 5.1 Mass budget (in kg) for the main platform per domain and module

Domain	Docking node	Service module	Habitat module
Attitude and orbit control	11	1229	46
Communication	80	41	0
Crew equipment	440	180	827
Data handling	0	1,008	173
Environmental control and life support	1,232	1,232	1,232
Extra-vehicular activities	0	0	818
Payload	0	0	3,301
Power generation	0	2,774	1,460
Propulsion	0	489	376
Robotics	152	0	0
Structure, shielding and mechanisms	12,097	6,106	11,226
Thermal control	768	4,140	864
Harness	26	26	26
System margin (20%)	2,961	3,445	4,069
Propellant		938	814
Launch adapter	125	125	125
Total mass	17,893	21,734	25,358

Table 5.2 The power budget for the main platform by domain and operation mode. Power values are given in watts including a component margin based on ESA (2017) and system margin of 20%. The power demand given for the power generation domain represents the losses of the power system

Domain	Standard mode	Crew exchange	Survival mode
Attitude and orbit control	287	287	256
Communication	781	781	749
Crew equipment	320	320	179
Data handling	4,442	4,442	2,810
Environmental control and life support	7,093	9,220	3,476
Payload	7,975	7,995	762
Power generation (losses)	13,023	13,023	13,023
Propulsion	1,365	1,365	1,365
Robotics	27	27	0
Structure and mechanisms	0	0	0
Thermal control	2,801	2,801	2,801
Total power demand	38,114	40,241	25,453

Nominal operations require approx. 40 kW of power generation on average including margins on component level compliant with ESA (2017), which is only slightly exceeded during crew exchange. These numbers include losses of about 13 kW for power generation, conversion and distribution. Adding a 20% system margin to the demand excluding these losses, the main platform is supplied with 32 kW of electrical power by the power subsystem. Significantly less power is needed during survival mode, where the losses accounted for in the power generation domain are the largest part of the budget. This is because losses have been calculated for maximum power demand, providing margin in all other cases.

5.5.3 *Subsystems*

Attitude and Orbit Control

Due to the module masses (see Table 5.1), it is assumed that each module will be launched separately during the initial phase of operations and docking will occur on orbit. This means each module requires actuators and sensors for attitude and orbit control.

Control moment gyroscopes are used for attitude control between the docking node and service module (see Fig. 5.3). The service module will also contain thrusters to act as active elements during assembly of the platform and after assembly is complete will be used for attitude control.

Communication

The largest impact on the communication design is from the payload, ensuring all data is downloaded on time to the scientists responsible. The capabilities are similar to those of the ISS (25 Mbps up, 300 Mbps down).

Depending on the exact function, three systems are foreseen. Telemetry and audio will be covered via S-band. Video signals and all payload data will be handled via K-band and for emergency EVAs and docking procedures UHF is used.

To ensure unobstructed communication a relay system is assumed to be used, similar to ISS. Examples would be the TDRS or the EDRS systems. While these systems may have been phased out by the time *Orbital Hub* is operational, it is assumed that continuing LEO human spaceflight will either lead to replacements or commercial activities with similar purposes.

Crew

For nominal operations a crew of three is assumed to be permanently onboard the station—although the station will be autonomously operational. During crew exchange, including replacement of experiments and equipment, the design allows for a total of six crew members to be present and covered by the platform's life-support system.

The crew facilities include hygiene stations, toilets, exercise equipment, cooking and dining areas. These are distributed over the modules.

Data Handling

The data-handling requirements are similar to those of the ISS, therefore a similar data-handling system consisting of multiplexers and demultiplexers, as well as laptop computers, has been designed.

Environmental Control and Life-Support System

This subsystem manages safety, climate, atmosphere and water recovery. Each module is equipped with identical subsystem components to create redundancy for crew safety. Only one of these system sets will be operating at 100% during nominal operations and the other at 50%.

Crew exchange requires all three systems to be operating, but extra redundancy is provided by the visiting vehicles' life-support system. This also allows crew evacuation.

Payload

The payload has been designed as a black box, containing examples from existing or planned experiments for ISS. From these, power values and sizes were derived to model realistic demands for the strawman payloads. Accommodation is assumed to occur in International Standard Payload Racks (ISPR). The habitat module mostly contains experiments for human physiology, biology, radiation and manufacturing processes, and those which require regular maintenance by the crew.

Power Generation

The power generation system has been designed to supply on average about 30 kW of power (not including losses of the power system of about 13 kW, see Table 5.2), which includes a 20% system margin. The eclipse time, determined as 36 min, has to be covered by battery power, and overall the batteries are sized to supply power to the station in survival mode for two full orbits, even after a previous eclipse period (i.e., not fully charged batteries). The primary bus is two-failure redundant and overall the design is similar to ISS.

Propulsion

The propulsion system handles the rendezvous maneuvers during platform assembly. In addition, it is used for orbit maintenance and debris avoidance maneuvers, six of which are assumed per year with Δv s of about 1 m/s each. Approximately 90 m/s are budgeted for orbit raising. Furthermore, the propulsion system will be responsible for desaturation of the CMGs.

Nominally, the visiting vehicle or even the docked *Free Flyer* are assumed to conduct the maneuvers. The docking node is equipped with thrusters for contingencies and autonomous operation.

Robotics

Robotics contain robotic assistants for the crew, present on the main platform. This could be similar to NASA's Robonaut or DLR's Justin. The idea is to ease operations and have maintenance tasks delegated to the robotic assistant, increasing crew time available for experiments.

Thermal Control

The thermal control system is active. Coolant loops connect heat sources with the radiators. The basic architecture is similar to ISS: an internal cooling loop using water as coolant, and an external cooling loop, which uses ammonia. The rotatable radiators (see Fig. 5.3) can reject up to 30 kW of heat. Their area is 90 m².

5.5.4 Design Options

The major design option has been the configuration of the platform. In the configuration shown in Fig. 5.3, the docking node has a half-sphere of room for maneuvering vehicles. Positioning it behind the service module would obstruct the approach with the radiators and solar array.

The habitat could also be switched with the service module, providing even more clearance for maneuvering vehicles. However, this would require the robotic arm of the *Free Flyer* to reach past solar arrays in order to reach the airlock for large external payloads.

The present configuration has been chosen to allow easy access to the airlock and still have enough room for maneuvering.

To reduce costs for the construction of modules, it would be possible to re-use some modules of the ISS for the purpose of the *Orbital Hub*. These could include the Japanese *Kibo* module acting as laboratory or a docking node of the ISS. However, testing for safety would be required, ensuring that the systems are still safe for several years of operation. Furthermore, this would limit the onboard possibilities of using the modern technology that is particularly required for autonomy.

5.6 Free Flyer

The *Free Flyer* is a versatile vehicle, which can dock and undock from the platform and supplies an experimentation space with even less acceleration perturbations, e.g., through lack of continuous operation of life-support systems or crew movement. More details can be found in Jahnke et al. (2018).

5.6.1 Free Flyer Design

The *Free Flyer* consists of three parts: the pressurized laboratory, external platform and service module. An illustration is given in Fig. 5.4. The pressurized platform has a docking adapter, with which it can be connected to the platform or another space station, such as the ISS, and the service module contains thrusters for maneuvering the *Free Flyer* away from the platform. Both the pressurized laboratory and the external platform can carry payloads.



Fig. 5.4 The *Free Flyer* in un-docked operation with the pressurized part in front

Due to crew accessibility, the pressurized lab can be maintained and new experiments can be installed. A small air lock, similar to that in the Japanese laboratory on ISS, allows payloads to be placed outside, using a robotic arm. The external platform is critical because it allows versatile experiments in the space environment or concerning observation and astronomy.

The service module is responsible for power generation, propulsion and thermal control. It is not pressurized and maintenance is not foreseen. Assuming an ARIANE 6-4 as a launcher, the *Free Flyer* can be placed on orbit already assembled and has been designed to withstand the respective loads.

5.6.2 *Free Flyer Mass and Power Budgets*

Table 5.3 shows the mass budget for the *Free Flyer* in two configurations: the launch configuration and the actual on-orbit configuration during operation. The difference is the payload mass, which will be added once the spacecraft is on orbit. This separation allows the launch mass to be kept below 20 metric tons and thus in the range of ARIANE 6 or comparable launchers.

Table 5.3 The mass budget for the *Free Flyer* in launch configuration (reduced payload mass) and operational on-orbit configuration. Masses are given in kg

domain	FF launch configuration	FF on-orbit configuration
Attitude and orbit control	874	874
Communication	360	360
Crew equipment	36	36
Data handing	153	153
Environmental control and life support	171	171
Payload	1,356	8,340
Power generation	2,490	2,490
Propulsion	396	396
Robotics	304	304
Structure (incl. mechanisms, debris and radiation shielding)	6,100	6,100
Thermal control	1,873	1,873
Venting system	47	47
Harness	404	404
System margin (20%)	2,913	4,310
Propellant	1,100	1,100
Launch adapter	125	125
Total Mass	18,702	27,083

Table 5.4 The (average) power budget for the *Free Flyer* per domain and operation mode. Power values are given in watts including a component margin based on ESA (2017), including a 20% system margin. The power demand given for the power-generation domain represents the losses of the power system

Domain	Standard Mode	Survival Mode
Attitude and orbit control	424	21
Communication	791	330
Crew equipment	0	0
Data handing	575	486
Environmental control and life support	351	0
Payload	10,069	0
Power generation (losses)	10,530	4,473
Propulsion	2,395	3,877
Robotics	60	0
Structure and mechanisms	13	13
Thermal control	673	1,794
Total power demand	25,882	10,994

Besides the payload mass, the structure and power system are the biggest contributors. The structure also includes debris and radiation shielding, which is considerate for human-rated spacecraft. A more detailed description of the subsystems is given in Sect. 5.6.3.

The power budget is given in Table 5.4. Losses within the system and the payload form the largest part, amounting in sum to about 70%, followed by the electrical propulsion system. This is similarly true for the survival mode of the system, which can sustain the spacecraft for 1.5 orbits without any power generation capability and assuming propulsion required for orbit maneuvers.

5.6.3 Subsystems

Attitude and Orbit Control

Similar to the main platform, the *Free Flyer* uses four CMGs and thrusters for attitude control. The CMGs are smaller versions of those found on ISS and are used during nominal operations. Thrusters are for power-saving survival mode.

For nominal attitude determination, satellite navigation systems are used. During survival mode star trackers take over. For less precise measurements, sun and horizon sensors are applied.

Communication

The links are used for communication—one directly to the ground, one using the main platform. The communication components are therefore similar to those of the main platform.

Data Handling

A two-failure tolerant system has been designed for the *Free Flyer*, with a total of three hot redundant onboard computers installed. Data exchange is realized via a triple-LAN gigabit network. Data from external payloads can also be transferred using WiFi.

Environmental Control and Life-Support System

The life-support capabilities of onboard the *Free Flyer* are limited, as a crew is only present when the vehicle is docked to the main platform (or, e.g., ISS), and it is assumed that the majority of tasks is handled by the life support of the respective “harbor”. The limited capabilities are safety relevant and include fans.

Payload

The two payload areas of the *Free Flyer* provide different capabilities. The pressurized area can house up to 5 ISPRs containing experiments. The external platform can house up to ten payloads similar to JEM EFU experiments and provides a further 8 m² for smaller payloads. These positions are distributed over the platform and can allow nadir or zenith pointing when necessary (e.g., for Earth observation).

Power Generation

Adding a 20% systems margin to the power budget (not including losses of the power generation system of about 10 kW), the power generation is designed to supply on average about 20 kW of power during nominal operation and to allow survival in case of a total loss of power generation capability for one orbit.

The solar array size is thus 167 m², based on MegaFlex solar panels (Neubauer et al. 2017). Lithium-ion batteries, with better performance than Ni-based batteries, are used, and can even provide power for two orbits in case of a total loss of power generation.

Propulsion

The service module of the *Free Flyer* houses the propulsion system, containing both low-thrust elements (for long-duration perturbations, such as atmospheric drag) and chemical thrusters (mainly for rendezvous and docking, reaction control and debris avoidance).

There are four 400 N thrusters at the back of the service module and 24 thrusters with 220 N for reaction control. The tanks are designed to be refilled and can hold up to 760 kg of propellant, sufficient for completing a free-flying phase, including debris avoidance maneuvers and other contingencies, as well as rendezvous and docking.

The electrical thrusters are the RIT-10 EVO and RIT-20 thrusters; four of each type are installed, each providing up to 130 mN of thrust. Based on orbit simulations concerning drag, a total of 660 kg of Xenon propellant is assumed to be sufficient for a 15-year operation. The tanks have room for half this amount, requiring one refill or exchange of tanks.

Thermal Control

The active thermal control system is similar to the main platform's system, including two coolant loops. A 12 m² deployable radiator and 25 m² of body-mounted radiators are used for heat transfer into space.

Venting System

For collecting waste gases evacuated, for example, from interfacing payloads, the *Free Flyer* contains a dedicated internal loop system and valves. It can also be used for emergency evacuation in case of fire.

5.7 Discussion

The design philosophy for the *Orbital Hub* has been its modularity and its nature as a nucleus for further additions from possible different partners. The core presented here can support other components from different partners, including private entities, such as Bigelow Aerospace or Axiom (cf. Sect. 5.1 for their plans). These modules can use their own systems for, for example, life support to extend the crew capabilities or system capabilities beyond the original design.

The *Orbital Hub* and *Free Flyer* are designed based on ISS heritage. The *Free Flyer* is derived from Europe's *Automated Transfer Vehicle* (ATV), *Columbus* module and also the ISS Truss and *Kibo* in respect of the external platform. Its launch mass is comparable to ATV, and its mass with payload is comparable to *Kibo* including the external platform.

The *Orbital Hub* has less than 20% of the mass of ISS, similar in size to the ISS when it was first crewed in 2000 during Expedition 1. Its power and mass budgets are similar as well, which supports their plausibility (Jahnke 2018).

5.7.1 Launch and Operation

Orbital Hub is designed for the same orbit as the ISS uses, i.e., around 400 km altitude and 51.6° inclination. On the one hand that allows for a continuous use of the existing launch and ground infrastructures; on the other hand it gives the opportunity for a smooth transition between ISS and post-ISS eras. Since at the present time there is no spacecraft like the US Space Shuttle available for module delivery and assembly, the *Free Flyer* plays a key role as an active docking partner for the installation of *Orbital Hub*. A launch sequence could look like this (Romberg et al. 2017):

- (1) Launch of the *Free Flyer* (e.g., on Ariane 6–4, Proton, Atlas V, Falcon 9, H-II);
- (2) Launch of the habitat module (e.g., on Delta IV, Proton, Falcon Heavy), followed by an autonomous docking by the *Free Flyer*;

- (3) Launch of the service module (e.g., on Ariane 6–4, Proton, Atlas V), followed by an autonomous docking by the Free Flyer plus habitat module stack;
- (4) Launch of the docking node (e.g., on Ariane 6–4, H-II, Atlas V), followed by an autonomous docking by the Free Flyer plus habitat module plus service module stack;
- (5) With the fifth launch the first crew could fly to the *Orbital Hub* (e.g., on Dragon, Soyuz, Dream Chaser, Shenzhou) and inaugurate the new station.

To keep both risk and mission costs down, *Orbital Hub* is designed for heavily automated assembly and operation without the need of EVAs, although for contingencies a small crew hatch is foreseen at the habitat module. Crew or cargo docking ports are available on the habitat module and the *Free Flyer* during the envisioned free-flying campaigns lasting up to three months, or at one of the four docking ports of the docking node, whilst one will most likely be occupied by a cupola. This way crew exchanges and cargo flights can be combined with docked phases of the *Free Flyer* for a servicing period of approximately two weeks.

Docking of the *Free Flyer* at the main platform will occur similarly to ATV docking at ISS. The vehicle size is similar, as is the technology involved—the design reused the ATV components. Thus, even the docking, occurring about four times a year, has heritage and occurs automatically.

In nominal operation, *Orbital Hub*'s docking node will be in forwards flight direction. To avoid re-fueling, orbit raising is intended to be done primarily by crew or cargo vehicles attached to the docking node, which requires that for those periods the habitat module is in the direction of flight. If no visiting vehicle is docked, the docking node thrusters would be used for orbit adjustments.

5.7.2 *The Orbital Hub and Other Platforms*

The *Free Flyer* is a versatile concept in itself and can be used with other stations in LEO or the lunar environment, such as the *International Space Station* or the *Lunar Orbital Platform Gateway*, using either as its base of operation for exchange of goods. The design allows docking with either and the same operational concept as for the *Orbital Hub* would also apply for ISS. It can operate independently and also in conjunction with private platforms such as Bigelow's or Axiom's concept.

Similarly, *Orbital Hub* can benefit from hardware already installed on ISS. Modules which have not expired after ISS ceases its operation can be transferred for use with the *Orbital Hub*, e.g., the Japanese laboratory module *Kibo* or the US lab *Destiny*. Refitting them would allow parts of their racks to be exchanged for crew compartments while retaining the remainder for laboratory work. This would significantly reduce the launch costs, and development costs and time for the *Orbital Hub*. Obviously, the repurposed modules need to be reviewed for their capability to do so and partnerships have to be agreed on.

Utilizing *Kibo* for *Orbital Hub* would not only add its laboratory capability, but also its platform, including such external payload capabilities as are already foreseen for *Orbital Hub*, and *Free Flyer*. It has a compatible size and a mass of approximately 24 metric tons.

Destiny has a mass of about 14 metric tons and could also be attached to *Orbital Hub*. Using the *Free Flyer* as active component, the ISS's robotic arm could detach either module and berth it with the *Free Flyer*, which could then "anchor" the module, providing it with orbit and attitude control until *Orbital Hub*'s service module and docking node can be attached, completing the platform.

This scenario depends on the timeline of ISS operation and *Orbital Hub* implementation. Currently, ISS's end of operation date has not been finalized, but the United States has decided to operate it until 2030 (Babin 2018). This leaves sufficient time to develop and launch the *Free Flyer* for such a scenario.

The *Free Flyer* is an autonomous vehicle, which can also be used as an active part for docking and berthing. It thus could accomplish station assembly for other platforms, such as Axiom, or support in the decommissioning of ISS. Similarly, it could support operation and science for LOP-G in a similar fashion to ISS or any other station. As it is equipped with a propulsion system, the *Free Flyer* could transfer to the lunar environment on its own, if supplied with fuel (e.g., with an external tank, replacing payload mass).

Operation of the *Free Flyer* and *Orbital Hub* can be organized as for *Bartolomeo*, i.e., with a direct link between scientists and experiment. This should be kept in mind when designing the actual processes behind it.

5.7.3 User Applications

The major resource that is utilized by *Orbital Hub*, and even more so by the *Free Flyer*, is the micro-gravity environment, which is undisturbed and needs no crew. Observation capabilities are also used, both for Earth and space observations, and radiation can be used as well as the human crew.

The following needs were identified from a survey of Germany's leading scientists in LEO applications, and have been accomplished by *Orbital Hub*'s design (Romberg et al. 2017):

- Possibility of observing processes in real time (e.g., materials) including on-orbit analysis opportunities to significantly reduce the return of samples,
- Low vibration levels (avoiding astronauts or moving structures for the *Free Flyer*),
- High and flexible modularity (easy access and exchange of samples or instruments),
- High data transmission and storage possibilities,
- Storage room for instruments, spare parts, new hardware or samples,
- Minimum utilization time of ten years,

- Robotic exchange of samples and instrument components,
- Maintenance possibilities and work bench for ad-hoc repair.

Furthermore, from a science perspective a crew exchange after approximately 20 days is favored, resulting in higher sample rates for human physiology and implying no need for extensive exercise devices. However, for cost reasons lower crew launch cadences, such as every six months, will be more likely.

In addition to traditional micro-gravity research, an extended focus was placed on Earth observation, atmospheric physics and technology demonstrations for human-rated platforms. The modular design combining a flexible interior and exterior and a crewed and manned part also opens up future, as yet unthought-of, commercial applications.

Space tourism has not been a major driver for the design of *Orbital Hub*, which is linked to the fact that, until now, no large-scale tourism has been conducted. If providers of such activities become more prominent, the design could be adapted, for example by adding another habitat module comprising equipment for tourists, including accommodation and leisure areas, such as a viewing port.

5.7.4 *Key Technology Availability*

The design of *Orbital Hub* and the *Free Flyer* has been focused on using available technologies. For instance, docking technology and AOCS equipment is similar to that of ATV. Most other components have been flown or are derived from existing technology, e.g., the payload airlock of the *Free Flyer* is of similar design to that on *Kibo*.

A major open point is currently the Habitat module of the main platform, which will be a new module design. However, a prototype called *BEAM* is currently berthed at ISS, proving the readiness of the technology. This prototype needs to be expanded in size to become the BA-330. In case this is not possible, a different habitat module can be used. This will change the design in detail, yet would not make it unfeasible.

5.8 Conclusion

Low Earth orbit is still a relevant environment for human spaceflight and can be utilized for a multitude of operations and scenarios, including tourism and science. Multiple scenarios exist for its utilization with different platforms. The *Orbital Hub* presented in this chapter serves as one flexible example of utilizing the micro-gravity environment and acting as nucleus for larger LEO structures.

It has been shown that it can be launched and operated with current launch vehicles and, overall, is a feasible design with heritage from ISS.

References

- Athanasopoulos, Köpping, Harald. 2019. The moon village and space 4.0: the 'open concept' as a new way of doing space? *Space Policy*, 49.
- Babin, Brian. 2018. H.R.6910—leading human spaceflight act. In *115th congress of the United States*.
- Balter, Michael, Neumann, Christian, Bräuer, Dirk, Dreißigacker, Christoph, and Steinbach, Sonja. 2019. ARTEC—a furnace module for directional solidification and quenching experiments in microgravity. *Review of Scientific Instruments* 90 (12). American Institute of Physics (AIP). <https://doi.org/10.1063/1.5124822>. ISSN 0034–6748.
- Brandon-Cremer, Lee. 2020. SpaceShuttleAlmanac.Com, License: CC BY-SA 4.0. <https://creativecommons.org/licenses/by-sa/4.0/deed.en>, [Accessed 31st March 2020].
- Clark, Stephen. 2021. Space station to receive new outdoor deck for science experiments. <https://spaceflightnow.com/2020/03/06/space-station-to-receive-new-outdoor-deck-for-science-experiments/>, [Accessed 10 April 2021].
- Crucian, Brian E., Makedonas, George, Sams, Clarence F., Pierson, Duane L., Simpson, Richard, Stowe, Raymond P., Smith, Scott M., Zwart, Sara R., Krieger, Stephanie S., Rooney, Bridgette, Douglas, Grace, Downs, Meghan, Nelman-Gonzalez, Mayra, Williams, Thomas J., Mehta, Satish. 2020. Countermeasures-based improvements in stress, immune system dysregulation and latent herpesvirus reactivation onboard the international space station – relevance for deep space missions and terrestrial medicine. *Neuroscience and Biobehavioral Reviews* 115: 68–76.
- David, Leonard. 2020. China details ambitious space station goals. <https://www.space.com/11048-china-space-station-plans-details.html>, [Accessed 31st March 2020].
- ESA. 2017. *CDF standard margin philosophy*, ESA-TECSYE-RS-006510, issue 1, revision 2.
- Fateri, Miranda, Kaouk, Ali, Cowley, Aidan, Siarov, Stefan, Palou, Manel Vera, Gobartt González, Fernando, Marchant, Romain, Cristoforetti, Samantha, Sperl, Matthias. 2018. Feasibility study on additive manufacturing of recyclable objects for space applications. *Additive Manufacturing* 24 : 400–404.
- Foust, Jeff. 2020. NASA selects Axiom Space to build commercial space station module. <https://space.com/nasa-selects-axiom-space-to-build-commercial-space-station-module/>, [Accessed 31st March 2020].
- Gateway Foundation. 2020. The gateway. <https://gateway.spaceport.com/the-gateway/>, [Accessed 15 June 2020].
- Gohd, Chelsea. 2020. Yes, the 'Von Braun' space hotel idea is wild. could we build it by 2025?. <https://www.space.com/gateway-foundation-von-braun-space-station.html>, [Accessed 10 April 2020].
- Jahnke, Stephan Siegfried, Volker Maiwald, Claudia Philpot, Dominik Quantius, Oliver Romberg, Wolfgang Seboldt, Vincent Vrakking, and Conrad Zeidler. 2018. Orbital Hub: a concept for human spaceflight beyond ISS operations. *CEAS Space Journal* 10: 355–379.
- Jones, Andrew. 2021. This is China's new spacecraft to take astronauts to the moon. <https://www.space.com/china-new-spacecraft-crewed-moon-missions.html>, [Accessed 23rd March 2021].
- Koryanov, Vsevolod, Toporkov, Alexey, Pozdnyakov, Anton. 2021. The concept of a long-term service station to increase the life duration of some satellites or to remove space debris. *Journal of Space Safety Engineering* 8 (1): 23–28.
- Kwok, Andy, Samuel Rosas, Ted A. Bateman, Eric Livingston, Thomas L. Smith, Joseph Moore, David C. Zawieja, Tom Hampton, Xiao W. Mao, Michael D. Delp, and Jeffrey S. Willey. 2020. Altered rodent gait characteristics after ~35 days in orbit aboard the international space station. *Life Sciences in Space Research* 24: 9–17.
- Magalhães, Tiago E.C., Silva, Diogo E.C.G., Silva, Carlos E.C.G., Dinis, Afonso A., Magalhães, José P.M., Ribeiro, and Tânia M. 2021. Observation of atmospheric gravity waves using a Raspberry Pi camera module on board the International Space Station. *Acta Astronautica* 182.
- Nature. 2019. Space station re-entry, Moon tapes and Ebola emergency. *Nature* 571: 452–453.
- Neubauer, J., et al. 2017. Progress on the space qualification of the high energy ABSL 18650NL Cell. https://batteryworkshop.msfc.nasa.gov/presentations/Progress_Space_Qual_High_Ener_18650NL_Cell_JNeubauer.pdf, [Accessed 1st March 2017].

- O'Hara, William J., Jason M. Kish, and Mary J. Werkheiser. 2018. Turn-key use of an onboard 3D printer for international space station operations. *Additive Manufacturing* 24: 560–565.
- Oei, Ing, Mirra, Carlo. 2008. Europe's human research experiments integration on the international space station. *Acta Astronautica* 63 (7–10): 1126–1136.
- Romberg, Oliver, Quantius, Dominik, Dittus, Hansjörg, Baerwalde, Sven, Seboldt, Wolfgang, Philpot, Claudia, Jahnke, Stephan Siegfried, Schlegel, Hans, Gold, Mike, Zamka, George, Retat, Ingo, Wohlgemuth, Rainer, Lange, Max. 2017. *Orbital-hub DLR vision 2025*, 2nd edn. DLR-Forschungsbericht. DLR GB-Post-ISS_08/2017.
- Seboldt, Wolfgang. 2004. Space- and earth-based solar power for the growing energy needs of future generations. *Acta Astronautica* 55 (3–9): 389–399.
- Sgambati, A., M. Deiml, A. Stettner, J. Kahrs, P. Brozek, P. Kapoun, V. Latini, M. Mariani, E. Rabbow, P. Manieri, R. Demets, and A. Elsaesser. 2020. SPECTROModule: A modular in-situ spectroscopy platform for exobiology and space sciences. *Acta Astronautica* 166: 377–390.
- Shen, Guojun, Liu, Yanheng, Sun, Geng, Zheng, Tingting, Zhou, Xu, Wang, Aimin. 2019. Suppressing sidelobe level of the planar antenna array in wireless power transmission. *IEEE Access* 7 : 6958–6970.
- Shoard, Catherine. 2021 Space race 2: Russian actor bound for ISS in same month as Tom Cruise. <https://www.theguardian.com/science/2021/may/13/russia-send-actor-director-iss-shoot-first-movie-space>, 2021 [Accessed: 13th May 2021].
- ULA. 2017. Bigelow aerospace and united launch alliance announce agreement to place a B330 Habitat in low lunar orbit. <https://www.ulalaunch.com/about/news/2017/10/17/bigelow-aerospacel-and-united-launch-alliance-announce-agreement-to-place-a-b330-habitat-in-low-lunar-orbit>, [Accessed: 25th April 2021]
- Wall, Mike. 2020a. 1st private space station will become an off-earth manufacturing hub. <https://www.space.com/37079-axiom-commercial-space-station-manufacturing.html>, [Accessed 1st April 2020a]
- Wall, Mike. 2020b. Private space habitat to launch in 2020b. Under commercial spaceflight deal. <https://www.space.com/32541-private-space-habitat-launching-2020b.html>, [Accessed 1st April 2020b].
- Wall, Mile. 2021. China launches core module of new space station to orbit. <https://www.space.com/china-launches-core-module-tianhe-space-station>, [Accessed 28th April 2021].
- Warren, Lara Elisabeth. 2020. International space station open-source data. *Patterns* 1 (9).
- Williams, Matt. 2020. Gateway foundation shows off their plans for an enormous rotating space station. <https://www.universetoday.com/141523/gateway-foundation-shows-off-their-plans-for-an-enormous-rotating-space-station/>, [Accessed 10th May, 2020]
- Xinhuanet. 2020. China readying for space station era: Yang Liwei. http://www.xinhuanet.com/english/2018-07/08/c_137310103.htm, [Accessed 1st April 2020]

Chapter 6

Instrumentation for Planetary Exploration



Emily F. Klonicki-Ference, Michael J. Malaska, Mark P. Panning,
Sarah E. Waller, and Patrick J. Gasda

6.1 Introduction

NASA's "big questions" that currently motivate planetary exploration include how the universe works, how our solar system evolved, what characteristics lead to life, and whether we are alone. To address these civilization-scale science investigations, interplanetary spacecraft transport scientific payload, which may include in-situ and/or remote-sensing instrumentation, to designated bodies to conduct their experiments and make observations. The data is then returned back to Earth for scientists to analyze. Planetary instrumentation is designed and built by institutions throughout the world to meet the mission science objectives at one or more selected target(s).

Traditionally, during mission development, a solicitation for instrument proposals is made that traces to the science goals and/or a science traceability matrix (STM). An STM is a planning spreadsheet that begins with science objectives, then tracks

The original version of this chapter was revised: The small instrument parameters have been updated. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-97913-3_35.

E. F. Klonicki-Ference (✉) · M. J. Malaska · M. P. Panning · S. E. Waller
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., M/S 183-301,
Pasadena, CA 91109, USA
e-mail: Emily.f.klonicki@jpl.nasa.gov

M. J. Malaska
e-mail: Michael.J.Malaska@jpl.nasa.gov

M. P. Panning
e-mail: Mark.P.Panning@jpl.nasa.gov

S. E. Waller
e-mail: sarah.waller@jpl.nasa.gov

P. J. Gasda
Los Alamos National Laboratory, PO Box 1663, Los Alamos, NM 87545, USA
e-mail: gasda@lanl.gov

to required measurement values and precisions, then to requirements for instrument performance. The proposal process then downselects to a set of payloads capable of executing the investigations outlined in the mission objectives (Bayer et al. 2017). However, while objectives are set by the mission proposal, they may be approved to be expanded and or enhanced, enabling scientists to collect data beyond the original scope. This notion is best exemplified by the primary and extended missions of Voyagers 1 and 2. Launched in 1977, each spacecraft carried 10 instruments which conducted 11 investigations including radio science, with the original goal of exploring Jupiter and Saturn. Together, the Voyager spacecraft made multiple major discoveries, including the active volcanoes on Jupiter's moon Io and the identification of Saturn's G-ring (Stone 2017). Voyager 2, the sister ship of Voyager 1, is the only spacecraft to gather data from the outer planets of Uranus and Neptune (Stone 2017). The Voyager Interstellar Mission (VIM), which extended the mission lifetime for both craft, enabled the continued collection of important data beyond the outermost edge of the Sun's realm, decades after launch (Stone 2017).

The sustained successes of Voyagers 1 and 2 are two of the many accomplished mission demonstrations that highlight the importance of planetary instrumentation. The following chapter explores general planetary instrument specifications including size, robustness, and cleanliness. In addition, the chapter highlights the history, basic principles, and future developments of specific instrument classes. Not all instrument specifications or existing planetary instrumentation has been defined in this chapter, as many have been developed throughout the history of space exploration. Furthermore, many continue to be developed for future mission concepts that push the boundary, increasing their capabilities and complexity. The goal of this chapter is to provide an overview of selected instrument classes (synthetic aperture radar, spectrometers, seismic instrumentation, and nano-and microtechnologies), as well as their requirements, so that the reader may understand the operations, observations, and requirements that may be made for a specific technology.

6.2 General Instrument Considerations

From Voyagers 1 and 2 to Mars Perseverance, selected missions and mission concepts have varied in their complexity and challenges for planetary instrumentation. Factors such as the science objectives and the target environment are considered during payload selection and design. As an example, instrument operations and requirements such as instrument form factor and cleanliness significantly differ between a subsurface investigation of Europa and a flyby of Mars.

Technologies that reduce instrument mass, power, and volume allocations without decreasing scientific capability are of significant importance, as spacecraft continue to make breakthroughs in miniaturization in both satellites and landed platforms (NASA 2010). Locations on the spacecraft where instruments may be accommodated can be predetermined to instrument selection, or competing instruments must

demonstrate their ability to be accommodated on the spacecraft in the proposal document. This may potentially reduce the number of payload and/or capabilities that may be chosen. In addition, from the project standpoint, instrument mass and volume generally correlate with cost (Freaner et al. 2010). A review of the instruments that have flown on past and current missions, demonstrates the ability of engineers to design components with a reduced form factor (Roman et al. 2008; Mielczarek et al. 2019; Miles et al. 2016). For example, engineers have developed micro- and nano-electrochemical systems for a miniature time-of-flight planetary mass spectrometer (Roman et al. 2008).

Major advances in instrument miniaturization include—but are not limited to—seismometers, microfluidic devices, thermal probes, magnetometers, particle detectors, optical instruments, and penetrometers (Castillo-Rogez et al. 2017). In platforms that are power limited, such as battery-operated spacecraft, data transmission and number of sampling events are allocated prior to launch based on science operations priorities and power availability. Advances in technologies such as in high-performance multi-core processors, large focal plane arrays, and thermal management systems have reduced the power requirement for instrumentation (NASA 2010). Innovations in materials, fabrication techniques, and electronics have produced higher-performing payload in terms of their reduction in mass, power, and volume requirements (NASA 2010).

Aside from reducing instrument form factor and power requirements, instrumentation must be robust against, or have protective mechanisms to withstand the dynamic environments during prelaunch, tour, and deployment operations. These environments range from shipping and handling pre-launch to the intense vibrations during launch and the vacuum of space. In the mission development phase, engineers and scientists work closely to identify conditions that may affect the instrument lifetime including magnetism, radiation, temperature, pressure, and sample compositions that the instrument may be exposed to (Meyer et al. 1996). The instruments are then tested against specific conditions by placement in thermal vacuum chambers and on vibration machines before ever flying.

In addition, instrumentation must maintain specific cleanliness requirements during hardware build and launch operations. Planetary protection (PP) is an interdisciplinary field with the goal of preventing forward contamination during planetary exploration by terrestrial organisms and organic materials being transmitted by spacecraft. PP also includes protecting Earth against extraterrestrial life or bioactive materials in returned samples (backward contamination) (NASA 2011). For instrumentation, these policies specifically require that guidelines established by the designated PP engineer for a mission are carefully followed and assist in the construction of low biological burden or sterile hardware based on these requirements. Similar to PP, contamination control (CC) establishes cleanliness protocols to prevent particulates such as dust and debris from interfering with sensitive instrumentation. In both practices, insufficient hardware cleanliness may affect the reliability of collected data and preservation of extraterrestrial environments (Dworkin et al. 2018).

6.3 Instrumentation Categories

The following sections in this chapter provide an overview of selected instrument classes: synthetic aperture radar, spectrometers, seismic instrumentation, and nano- and microtechnologies. While there are a significant number of technologies not covered in this chapter, the few highlighted instruments span planetary in-situ and remote-sensing applications and aim to contribute information in regard to instrument principles and planetary history, requirements, and future developments.

6.3.1 *Synthetic Aperture Radar*

6.3.1.1 Principles of Synthetic Aperture Radar

Synthetic aperture radar (SAR) is a technique in which a series of radio signals emitted from an observing platform (spacecraft, airplane) are then reflected by the surface and received by an antenna to build up an image as viewed at these wavelengths (Henderson and Lewis 1998). The returned signal is referred to as backscatter, and an SAR image shows the varying amounts of backscatter returned from the surface. The reflected amount of radar backscatter is sensitive to multiple factors, including: grain-size roughness of materials at the scale of the radar wavelength, angular faces that reflect radar energy back towards (or away from) the receiver, volume scattering, dielectric constant of the material, and the presence of dielectric constant changes (for example, layering of materials with different dielectric properties). Due to its dependence on surface roughness and physical parameters, radar backscatter can provide complementary information to visible or infrared spectroscopy. Radar wavelengths are also longer than visible and infrared wavelengths (usually radar is on the order of cm to 10 s of cm), and can thus interrogate deeper into the surface than visible wavelengths, on the order of 10 s of wavelengths (Fig. 6.1).

6.3.1.2 Synthetic Aperture Radar in Past Missions

Planetary SAR can reveal the surfaces of worlds that have visibly obscuring atmospheres such as those of Venus and Saturn's moon Titan (Ford et al. 1993; Elachi et al. 2005a, b, 2006). Saturn's moon Titan has both an optically thick haze layer and absorption at many infrared wavelengths due to atmospheric methane. The Cassini spacecraft's RADAR instrument was able to use SAR to interrogate surface morphologies and determine structures such as dunes, dissected plateau, craters, lakes, and channels at high resolution (Lopes et al. 2019). The SAR data was used to define Titan's terrain unit classification system and enabled geomorphological mapping (Malaska et al. 2016; Lopes et al. 2020; Schoenfeld et al. 2021). An example image is shown in Fig. 6.2, which shows a highly dissected plateau and valley terrain

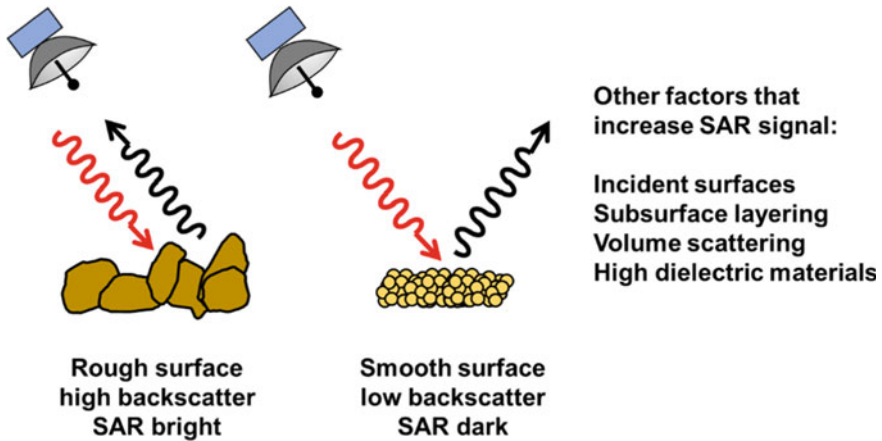


Fig. 6.1 Graphic describing radar backscatter. Large blocky objects have large facets that reflect some of the radar energy back to the radar antenna, while smaller objects reflect radar energy away from the surface. Increased energy detected by the antenna will make the surface appear bright to SAR, while decreased energy return (because it was reflected away) will make the surface appear dark

on Titan. The incident radar beam creates shadowing which can reveal itself as a 3D effect. Most SAR is acquired at oblique angles.

Passive microwave techniques include microwave emissivity, where microwave (radar) energy naturally emitted by the planetary surface is detected by an antenna. This technique can also be performed during SAR acquisition, when active and returned signals are not being received, or at larger distances where SAR is not practical. When coupled with an understanding of the physical temperatures (in order to determine the amount of expected blackbody radiation at these wavelengths), microwave emissivity can provide information on grain size, volume scattering, and other material properties (Janssen et al. 2009, 2016; Le Gall et al. 2016). Microwave emissivity measurements of Titan's surface were used to differentiate between terrain units and classify them as organic terrains and icy terrains (Malaska et al. 2016, 2020).

6.3.1.3 Future Developments in Synthetic Aperture Radar

Advances in radar, electronics, and digital processing technologies have enhanced SAR sensor performance and capabilities in radar remote sensing. These developments have been beneficial for earth scientists to better study our home planet. For example, NASA has been evaluating a compact L-Band synthetic aperture radar for potential use on unmanned aircraft that is capable of measuring changes in the Earth's surface (e.g., volcanoes, faults, landslides and glaciers) (Hensley et al. 2021). A combination of the L-Band SAR and S-band SAR, flying as two separate instruments on two different platforms, were used to develop a tomographic technique that

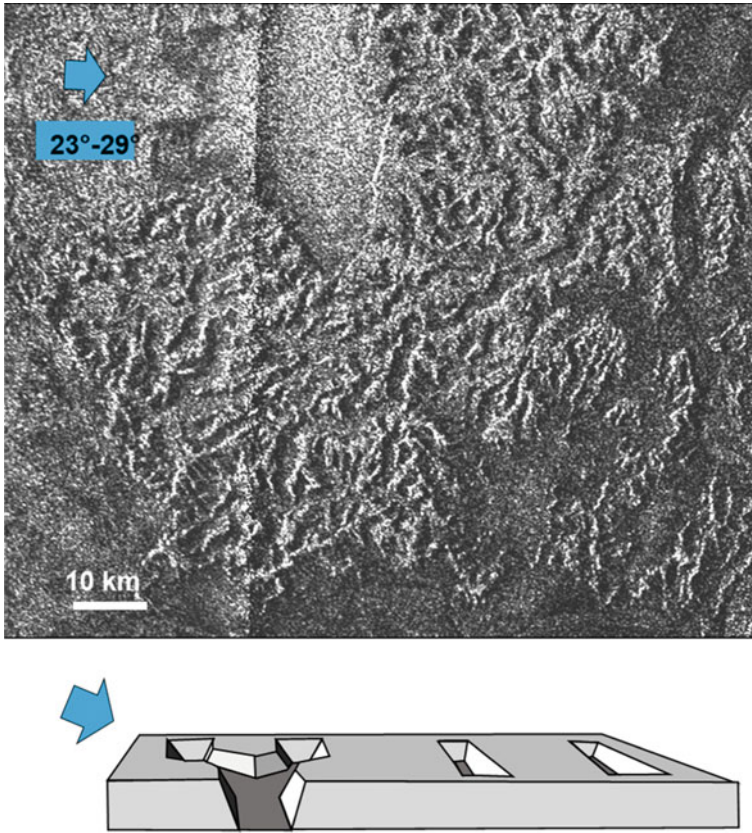


Fig. 6.2 Cassini spacecraft SAR image of the dissected plateau of Sikun Labyrinth, Titan (Malaska et al. 2020). Blue arrow indicates radar emission direction and incidence angle. Dark floored valleys and plains are radar-dark sediments, while the slightly brighter plateau are likely rougher brighter surfaces. The empty basin at center right has a radar bright floor, indicating it is likely a rough, not smooth, surface. Lower figure is a graphic showing how SAR can shade or lighten valley slopes depending on angle with incident radar beam

extracted a 3D structure of a forest to derive biophysical parameters (Hensley et al. 2021). While the Unmanned Aerial Vehicle Synthetic Aperture Radar (UAVSAR) has considerable terrestrial applications, it is used to develop the tools and technologies for future space-based radars. Additionally, DBSAR-2 and EcoSAR are two advanced radar systems recently developed and tested at the NASA Goddard Space Flight Center. These new instruments employ multiple-input multiple-output (MIMO) architectures characterized by “multi-mode operation, software defined waveform generation, digital beamforming, and configurable radar parameters” (Rincon et al. 2016). The beamforming technique allows simultaneous reception and processing which enables retrieval of a complementary rich target dataset including geophysical data, ocean-surface roughness (winds), sea height, soil moisture, and ice

classification. Similar to UAVSAR, these instruments have been matured to support several studies within Earth sciences including ecosystem analysis, topography, soil freeze–thaw characterization, ice sheet composition, glaciology, and surface hydrology (Rincon et al. 2016).

6.3.2 *In-Situ Standoff Instrumentation: Passive and Active*

In-situ standoff instruments can typically record passive reflectance or active spectra from meters away from the target, with no sample preparation. This ability gives standoff techniques advantages over in-situ contact techniques as they allow for larger-scale surveys of the areas surrounding a rover or a lander. The survey ability of the in-situ standoff techniques means that these instruments are used multiple times in each planning cycle, and their measurements provide input to mission science planners on a tactical basis in addition to science.

There are two broad categories of in-situ standoff instruments: passive and active. Passive techniques require only a sensor with ambient radiation sources while active techniques use an illumination source or excite the target prior to sensing. Passive techniques include infrared reflectance and thermal emission spectroscopy, while active techniques include laser-induced breakdown spectroscopy (LIBS), fluorescence spectroscopy, and Raman spectroscopy.

6.3.2.1 *Passive Imagers and Spectrometers*

Passive near infrared reflectance spectroscopy is widely used in orbital remote sensing in planetary exploration and for standoff interrogation on planetary rovers. Passive rover instrument examples include the MER rover Pancam, Curiosity rover Mastcam, the Perseverance rover Mastcam Z, and the Rosalind Franklin rover infrared spectrometer. These instruments record passive reflectance spectral images using color filters. The Curiosity rover ChemCam instrument and the Perseverance rover SuperCam instrument also have passive spectroscopy modes to collect near infrared reflectance spectra. The MER rover miniature thermal emission spectrometer (mini-TES) instrument recorded passive thermal spectra. Passive near infrared spectroscopy and thermal emission spectroscopy are techniques that help determine mineralogy of a sample. In this technique, ambient light or radiation hits the target, and is reflected or emitted towards the sensor; however, some of the photons at key wavelengths are preferentially absorbed by certain minerals and thus can be used to indicate the presence of those minerals. Typically, rover- or lander-mounted cameras use narrow-band pass color filters to collect, in addition to visible color images, passive reflectance infrared images. These images can then be processed to extract the spectra and infer surface mineral composition. The specific narrow-band filters chosen for a lander mission are based on major spectral features observed from orbital remote sensing of the mission target landing site. For example, the filter set

usually chosen for Mars rovers and landers is based on the diagnostic wavelengths for iron-bearing minerals and hydration features of minerals (Table 6.1).

Point spectroscopy is a technique where light from a single unresolved point is passed into a spectrometer to generate a spectrum of that integrated light. For example, the MER mini-TES instrument used a periscope on the rover mast to collect point spectra. When the Curiosity ChemCam instrument and the Perseverance SuperCam instruments do not fire their laser, either to collect dark spectra prior to active LIBS measurements, or in a dedicated measurement, passive spectra are collected. This is similar to how emissivity data is collected in between SAR RADAR active pulses (see Sect. 6.3.1.2). Passive spectra can be collected from rocks (Johnson et al. 2015) or the sky for atmospheric composition measurements (e.g., McConnochie et al. 2018). Due to the pervasive dust cover on Mars (e.g., Levasseur-Regourd et al. 2018), it is usually advantageous to collect passive spectra along with active LIBS measurements, as the LIBS plasma shockwave “cleans” the rock surface (Johnson et al. 2015). Otherwise, passive techniques require the use of tools on the rover arm to clean the surface or finding relatively dust-free targets.

The passive spectra of these imagers and spectrometers have been responsible for numerous discoveries on Mars in addition to overall helping plan traverses etc. The imagers have been integral to discoveries including hydrated features (Wellington et al. 2017), meteorites (Schröder et al. 2008), Mn-bearing materials (Arvidson et al. 2016). Mini-TES was integral in the discovery of opaline silica (Ruff et al. 2011) at the Spirit landing site. The most recent passive spectrometer, SuperCam VISIR on the Mars Perseverance rover, covers a range of 1.3–2.6 μm (Wiens et al. 2021), greatly expanding the range of passive infrared spectroscopy compared to ChemCam.

Table 6.1 Summary of spectral imagers on current and previous NASA Mars missions

Mission	Instrument	Spatial resolution	Color filters	Spectral range (nm)	References
Spirit and Opportunity	PanCam	0.27 mrad/pixel	8	432–1009	Bell et al. (2003)
Spirit and Opportunity	Mini-TES	20 mrad/pixel	n/a	5000–29,000	Christensen et al. (2003)
Curiosity	Mastcam	218 μrad and 74 μrad	12	445–1013	Malin et al. (2017)
Curiosity	ChemCam	150–500 μm per point	n/a	400–840	Johnson et al. (2015)
Perseverance	Mastcam Z	150 μm per pixel	14	400–1000	Bell et al. (2020)
Perseverance	SuperCam	0.685 mrad (VIS); 1.2 mrad (IR)	n/a	VIS: 382.1–467.5; 535–853; IR: 1300–2600	Wiens et al. (2021); Legett et al. (2022)

Note In the table VIS is visible wavelength spectrometer and IR is infrared wavelength spectrometer. In addition, mrad = milliradians and urad = microradians

6.3.2.2 Active Spectrometers

Active spectrometers use an illumination source, typically a laser, that could be continuous or pulsed, to excite a target and produce an emission or adsorption spectrum which is then recorded by a detector. A single point on the target can be excited and detected, in which case the instrument is similar to the point spectroscopy mode previously described. However, if the excitation source is scanned across the target, a spatial grid of spectra can be built up that creates a hyperspectral image cube product. One can think of the spectral cubes as a series of deep spectra pixels that have been all been arranged as an x - y map, or as a series of individual flat images that have been stacked together by wavelength. (So, instead of an RGB image, it is a [wavelength 1 response, wavelength 2 response, wavelength 3 response, wavelength 4 response, etc.].) These hyperspectral products are often in the form of GeoTiffs and can be visualized using proprietary software products such as ENVI (Harris Geospatial), ArcGIS (ESRI), open-source QGIS, or even used as large data tables using Python, MATLAB, or EXCEL (for smaller data products). Since they are three-dimensional, the products have a spatial resolution (pixel size or post spacing), and a spectral resolution resulting from binning of photon wavelengths.

Active spectrometer instruments include: the ChemCam instrument onboard the Curiosity rover (Wiens et al. 2012; Maurice et al. 2012); the Perseverance rover SuperCam (an all-in-one LIBS, Raman, and time-resolved luminescence instrument) (Wiens et al. 2021; Maurice et al. 2020); and the Chinese space agency's Tianwen-1 rover which has a LIBS instrument. Examples of a mapping active spectrometer include SHERLOC for fluorescence and Raman mapping, and PIXL for X-ray fluorescence mapping. These techniques provide elemental composition and distribution, organic detection and distribution, mineral identification, and composition of a selection of rare earth elements, respectively.

One example of an active spectrometry is LIBS, a technique akin to atomic spectroscopy that provides element compositional information. In this technique, a pulsed laser is focused onto a target, exciting the molecules to the point that it atomizes and ionizes a small volume of material to create a small plasma. When the plasma cools, the electrons of those atoms relax back into the atomic orbitals and emit light of the individual elements' characteristic electronic transitions. All elements can be identified with the LIBS technique, though the sensitivity of instruments to the elements vary. ChemCam LIBS collects spectra of targets at distances from 1.6 to 7 m standoff distance using a 1067 nm Nd:KGW pulsed laser (14 mJ/pulse, <8 ns pulse width, 3 Hz) using a mast-mounted telescope (150–500 micron spot size depending on distance). The telescope collects the light generated by the LIBS plasma, which is analyzed with three spectrometers that cover ranges in the UV (240–340 nm), violet (382–469 nm) and VNIR (473–905 nm) with 2.8, 4.0, and 4.0 pixel full-width half-maximum (FWHM) resolution in each spectrometer, respectively (Wiens et al. 2012; Maurice et al. 2012).

Fluorescence spectroscopy is a technique where a short-wavelength photon that excites a target molecule up to a new electronically excited state is followed by a lowering of the upper excited state to a slightly lower excited state (whether from

molecular collision or other phenomena), followed by a drop back down to the near-original state followed by release of a longer wavelength photon. There may be variable time delays due to the different decay processes. The photons emitted tell us about the molecular orbital energy states of the molecule, which are usually due to larger extended aromatic systems and conjugated double bonds. While not diagnostic, it can provide information on the molecular orbital structure of larger structures of a molecule when compared to vibrational spectroscopic techniques such as infrared reflectance and Raman spectroscopy. Biomolecules that fluoresce include aromatic amino acids such as tryptophan, tyrosine, and phenylalanine. The amino acid histidine is also aromatic, but has a very low quantum yield, and the signals in cells are primarily from the more fluorescent tryptophan and phenylalanine (Bhartia et al. 2008). Many other molecules are fluorescent as well, including benzene, naphthalene, and phenanthrene, with larger extended aromatic systems generally having longer wavelength fluorescence (Bhartia et al. 2008). Fluorescence has the advantage of being very sensitive; many organic molecules have quantum yields >10%.

Many microbes and biomolecules will fluoresce under ultraviolet excitation. This technique is referred to as native fluorescence. Imaging spectrometers can reveal bacterial colonies at low concentrations (Fig. 6.3). In a laboratory setting, not yet flown on a planetary mission, chemical derivatizing agents can be added that absorb or covalently bind to biological structures which are then revealed by ultraviolet imaging. This increases the fluorescence response and can be used to image individual cells (for example see Junge et al. 2004).

Time-resolved fluorescence uses specialized detectors along with pulsed lasers that can collect emitted light at discrete time intervals, thus separating out short-lived organic fluorescence from long-lived mineral fluorescence.

Raman spectroscopy is similar to fluorescence in that it involves the transient excitation of a molecule by a photon to create a virtual excited state, but then immediate

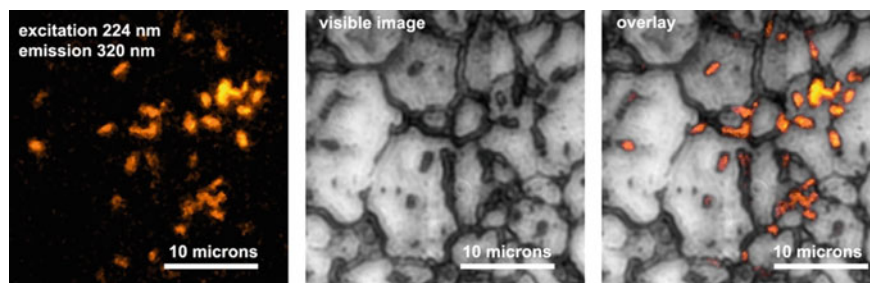


Fig. 6.3 Structure of stainless-steel plate with bacteria. The left panel shows bacteria fluorescence of *Bacillus pumilus* with DUV 225 nm wavelength excitation and 320 nm wavelength emission. Brighter yellow colors show higher signal detected, darker show less. The middle panel shows a visible light image of the stainless-steel plate, darker areas indicating grain boundaries. The right panel shows a combination of the fluorescence and visible image showing the bacteria and agar substrate. The bacteria preferentially inhabit the grain boundaries of the steel plate. Figure adapted from Bhartia et al. (2010)

de-excitation from the virtual state almost back down to the ground state. However, instead of dropping completely back to the original ground state and rereleasing a photon of the same inbound wavelength, the molecule drops down to an ever-so-slightly elevated ground rovibrational state—the molecule almost relaxes, but not quite. The re-emitted photon is just ever-so-slightly longer since the molecule did not relax all the way down to the ground state. The difference in rovibrational state energies is based on functional group atom–atom vibrations, rotations, and stretches. Raman thus gives information about the individual functional groups and atom–atom attachments in a molecule. Raman is also sensitive to the molecule’s overall environment, and it has recently been shown that molecules in a biological system have a different spectral signature compared to the individual component parts (Sapers et al. 2019). A graphical representation of the components by weight in an example bacterial cell is shown on the left side of Fig. 6.4, while the Raman intensity of the cellular components are shown on the right side: in cells, the nucleic acids are the major source of Raman signal.

In general, because most of the molecules decay fully to the ground state, the Raman effect is fairly insensitive. Compared to fluorescence, the Raman quantum yields (the number of returned photons per exciting photon) can be 7 orders of magnitudes lower than for fluorescence. For Raman spectroscopy, the typical technique is excitation-illumination by a laser with an edge filter that allows the re-emitted weak Raman signals near the laser line to be visible. There are several techniques that can be tried to help enhance the Raman signal. One technique is to use excitation wavelengths that are preferentially absorbed by the molecule (formally absorbed to an excited electronic state, not just a virtual excitation state); the technique is referred to as fluorescence-enhanced Raman. Another technique is to use time-resolved techniques for Raman spectroscopy. Time-resolved techniques are especially important for Raman spectroscopy as the Raman laser can stimulate luminescence in samples; since Raman is a weak phenomenon, the weak Raman signal can also at times be drowned out by background fluorescence. For example, in order to detect Raman signals, the SuperCam instrument collects only the first 100 ns of light after the laser pulse to exclude longer-delay mineral fluorescence signals.

Wavelength selection is important for Raman spectroscopy. For green Raman and red Raman, the wavelengths are too long to allow excitation-fluorescence of most organic molecules, but these still allow a weak Raman band to be detected. The use of shorter-wavelength deep UV illumination allows absorption of the UV photons by aromatic organic molecules followed by fluorescence emission and fluorescence-enhanced Raman signals. SuperCam uses two lasers: a 1067 nm laser for LIBS and a 532 nm laser for Raman and time-resolved fluorescence. SuperCam uses an intensifier in its visible spectrometer to both precisely time-gate the collected light based on the laser pulse and multiply the number of photons in order to produce a Raman spectrum (Raman spectral coverage and resolution). SuperCam’s UV and violet spectrometers cover the same range as ChemCam’s with the same resolution, but the intensified spectrometer has three ranges: 535–620 nm, 620–712, and 712–853 nm. The three ranges allow for higher resolution ($<12\text{ cm}^{-1}$) to enable Raman spectroscopy (Wiens et al. 2021). ChemCam is responsible for numerous

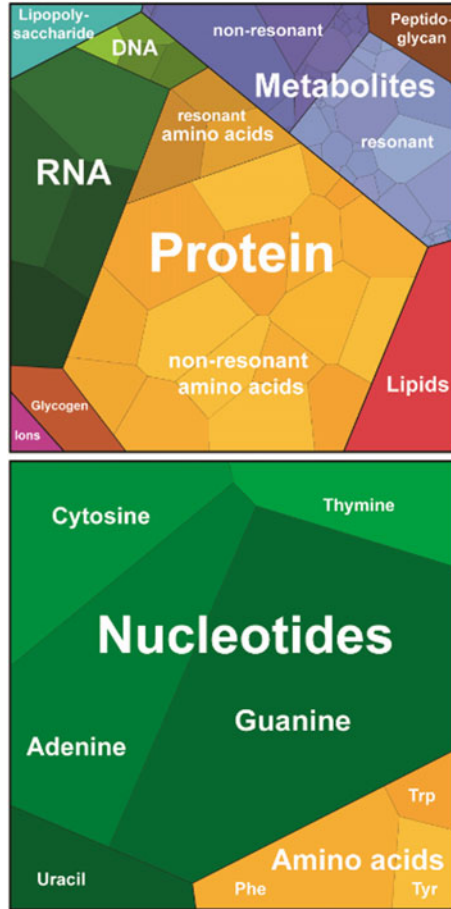


Fig. 6.4 Top of graphic: Proportional spatial area representation of cellular components of *E. coli* by weight, most of the weight is in protein and amino acid components (the darker-shaded yellow components represent the resonant amino acids phenylalanine (phe), tryptophan (trp) and tyrosine (tyr)). Bottom of graphic. Spatial area proportional representation of Raman intensity by molecular type. Most of the Raman signal from cells comes from the nucleic acids, notably guanine (both adenine and guanine are bicyclic heteroaromatics.) Figure adapted from Sapers et al. (2019)

discoveries on Mars including the first in-situ detection of hydrogen (Meslin et al. 2013), the first in-situ detection of boron (Gasda et al. 2017), high manganese (Lanza et al. 2014, 2016), and hydrated sulfates (Rapin et al. 2016), and has contributed to the overall understanding of Martian volcanism (e.g., Sautter et al. 2015) and the chemiostratigraphy of the lacustrine deposits in Gale crater (e.g., Frydenvang et al. 2020).

The Perseverance SHERLOC Deep UV fluorescence and Raman instrument is a combination of both fluorescence and Raman techniques in a mapping spectrometer

(Bhartia et al. 2021). The SHERLOC instrument is able to map a 7×7 cm area with a 50-micron spot laser and acquire both fluorescence and Raman spectra. For each point in the spatial map, the fluorescence spectrum gives sensitivity and overall information about molecular orbitals (Beegle et al. 2015), while the Raman spectrum provides information about individual bonds and functional groups. The two techniques combined can break degeneracies and ultimately provide better chemical identification, and thus quantitation. The spatial distribution can provide details of patterns of microhabitats or biological structures (such as stromatolites). The data products are hyperspectral three-dimensional “cubes” where there is an x - y component corresponding to the spatial location, and the z dimension is the variation of intensity versus wavelength.

6.3.2.3 Future Standoff In-Situ Applications and Development

Future applications of a combination deep UV fluorescence Raman instrument include the exploration of the Ocean Worlds. In an aqueous milieu, a simple point spectral measurement can provide information on the organic matter present in the liquid. For a down-borehole use, such as drilling into an icy surface, a deep UV fluorescence Raman mapping spectrometer known as WATSON has been tested in glacial ice. While SHERLOC is designed for surface interrogation, the transparent nature of ice and evaporite deposits (gypsum, halite) allows the WATSON laser light to penetrate into the interior of the medium. It has been shown that the WATSON instrument can interrogate up to 2 cm into ice if it is clear and relatively bubble-free (bubbles scatter photons; Eshelman et al. 2019.) Fig. 6.5 shows results of down-borehole scanning revealing spectrally uniform discrete localized fluorescent “hotspots” in glacial ice at 93.8 depth at Summit Station, Greenland (Malaska et al. 2020). Thus, borehole scanning in transparent material (such as ice and evaporite) can allow in-situ interrogation into unmodified habitats and regions of icy surfaces and penetration into thick deposits of evaporites on Mars or while drilling into potential Deep Ice habitats in the thick icy crusts of the Ocean Worlds.

6.3.3 Mass Spectrometry

6.3.3.1 Principles and History of Planetary Mass Spectrometry

Mass spectrometers (MS) have been key payload instruments in planetary exploration since the Apollo and Viking missions of the 1970s (Arevalo et al. 2020). The basic principles of MS are the production of ions from inorganic and organic substances, the separation of those ions by mass-to-charge, and their subsequent detection. Using mass-to-charge and relative abundances of ions in a mass spectrum, the molecule’s structure and elemental composition can often be determined, providing insight into a particular planetary environment.

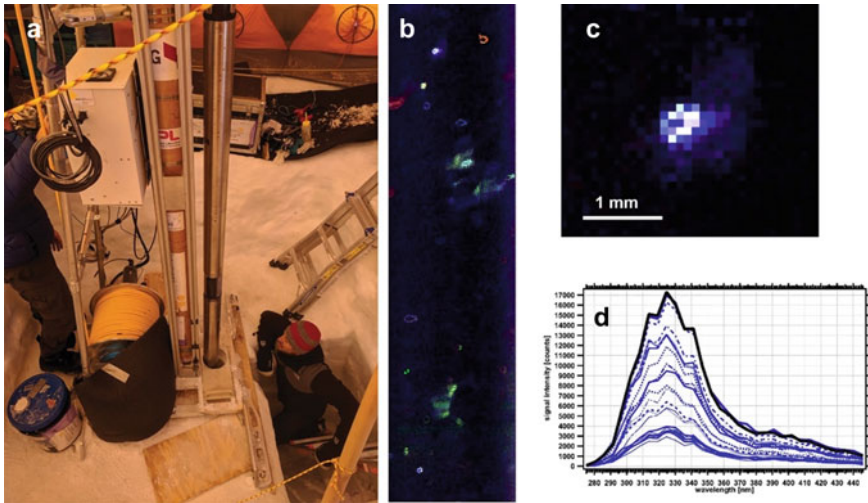


Fig. 6.5 a–d Down-borehole exploration of Greenland ice sheet with a deep UV mapping fluorescence instrument. **a:** inside the drill tent near Summit Station, Greenland. The instrument–drill combination (long silver cylinder) is shown poised over and entering the ice borehole at lower image center. **b:** 1 cm × 4 cm fluorescence map using 248 nm excitation taken at 93.8 m depth in glacial ice. Each spot is a punctate fluorescent hotspot. Different colors indicate different spectral classes. R,G,B indicates intensity responses at 412.9, 385.3, and 313.7 nm, respectively. **c:** detail of a punctate hotspot from the map in C (located at top right). RGB same as in B. Scale as shown. **d:** Spectra extracted from pixels in the hotspot in C. The spectra are spectrally uniform across the spot; varying only in intensity. Figures adapted from Malaska et al. (2020)

Arevalo et al. (2020) and Palmer and Limero (2001) both describe in detail the planetary history of mass spectrometry and each category of mass analyzer that has been utilized in space exploration. These include time of flight (TOF), sector instruments, quadrupole arrays, quadrupole ion traps, and cylindrical ion trap mass spectrometers. Previous successes of flown mass spectrometers spanning from the 1970s to the 1990s include the Cassini, Galileo, Huygens, Venus Pioneer, and Mars Viking Lander missions (Palmer and Limero 2001). Decades of development and research investments have led to more sensitive, miniaturized, and robust MS instrumentation. For example, the Pioneer Venus Orbiter Neutral Mass Spectrometer (ONMS), was a 3.8-kg instrument that successfully measured atomic and molecular ions between 1 and 46 Da (Arevalo et al. 2020). It had a mass resolving power of $m/\Delta m \approx 50$ FWHM or unit mass resolution (Arevalo et al. 2020). The MAss SPectrometer for Planetary EXploration/Europa (MASPEX), which will be onboard the Europa Clipper Spacecraft (tentative launch in 2024), is a next-generation spectrometer with significantly improved robustness and resolution (25,000 $m/\Delta m$ at 10% peak height) compared to previously flown MS instrumentation (Brockwell et al. 2016).

Due to the extensive planetary history and types of analyzers, not all flown instruments and future developments will be covered in this section. For the purposes of

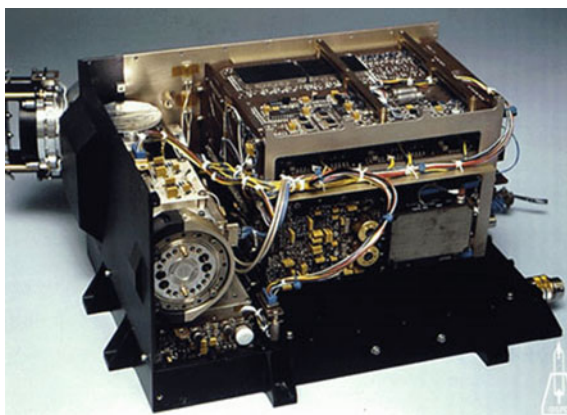
illustrating the unique features of the sample introduction systems and mass analyzers in MS instrumentation, the next section will focus on active plume sampling.

6.3.3.2 Mass Spectrometry for Active Plume Sampling

For sampling plumes at actively venting worlds such as Enceladus or possibly Europa, a plume fly-through mission while at orbital speeds is a potential option. MS using impact-induced ionization is an effective means of detecting inorganic and organic molecules entrained within ice grains, though care must be taken when selecting sampling speeds. Cassini spacecraft instruments (Fig. 6.6) sampled the Enceladus plume gases and grains at hypervelocity (7–17 km/s) during multiple flyby encounters, and detected H₂, NH₃, CH₄, Ar, silica nanograins, salts, and simple and complex organic molecules (Waite et al. 2006, 2009; Postberg et al., 2008, 2009, 2011, 2018; Hsu et al., 2015; Sekine et al. 2015).

Recent developments have focused on developing advanced instrumentation to determine biotic and abiotic distributions (Klenner et al. 2020a, b) and predicting optimal encounter velocities, which are thought to be 3–6 km/s for amino and fatty acids (Klenner et al. 2020a, b). Part of this effort involves understanding the physical processes that occur during hypervelocity sampling through laboratory testing. Cassini instrument salt and organic mass spectral distributions have been reproduced in laboratories with laser-induced acceleration-based experiments (Postberg et al. 2009; Khawaja et al. 2019). Other lab-based simulations include a light gas gun at the University of Kent that fires a sabot of water ice onto a surface (Hibbert et al. 2017; New et al. 2020a, b), and the Laboratory for Atmospheric and Space Physics (LASP) at CU Boulder impacts metal particles onto a cryocooled water–ice surface (Nelson et al. 2016; Ulibarri et al. 2019). While these analog experiments have provided insight into hypervelocity sampling, experiments replicating ice grain impacts of appropriate size and composition followed by MS analysis have not yet been accomplished. However, efforts towards ice grain impact studies are underway.

Fig. 6.6 The ion and neutral mass spectrometer (INMS), a quadrupole mass spectrometer, that was aboard the Cassini orbiter (©NASA)



The Aerosol Impact Spectrometer at UC San Diego can trap, accelerate, impact, and mass analyze a single ice grain at speeds >2 km/s. (Miller et al. 2019), with recent modifications to achieve 4.1 km/s. The Hypervelocity Ice Grain System at the NASA Jet Propulsion Laboratory can generate an ensemble of charged and neutral molecules, water clusters, and ice grains with velocities from 2 to 5 km/s (Waller et al. 2020), and upcoming experiments of impacts followed by mass analysis are planned. Theoretical models are being developed to predict how ice grain impact velocities relate to the survivability of organics which can then be validated by the experimental facilities (Jaramillo-Botero et al. 2021; Srama et al. 2004). Several upcoming missions and mission concepts will be utilizing MS instrumentation to including SURface Dust Analyzer (SUDA) which will be aboard Europa Clipper to analyze dust sputtered from the Moon's surface (Srama et al. 2004). In addition, a return mission to Enceladus to analyze the plume with particular focus on biosignature detection and quantification (Rhe et al. 2016) and a mission to Triton to determine if it hosts a subsurface ocean and to characterize its plumes that are in the concept stages aim to employ MS instrumentation.

6.3.3.3 Future Technology Development

As space exploration looks towards challenging in-situ missions and mission concepts in the coming decades (e.g., Europa Lander and Dragonfly) the requirements for the analyses and robotic platforms have pushed for significant research developments. Innovations in the development of robust, radiation-hardened, and miniaturized analyzers, as well as designing novel ionization sources and advancing chemical separation techniques, will enable scientific discovery in more challenging planetary missions (Arevalo et al. 2020). For example, researchers in ultrahigh resolution sensors aim to increase the effective speed of the analyzers, and to improve the resolving power, detection limit, and dynamic range (Zubarev and Makarov 2013). In addition, next-generation miniaturized laser sources capable of generating femtosecond pulses are currently in development (Tulej et al. 2021). Improvements in this field (reviewed in depth in Tulej et al. 2021) have led to and will continue the significant progress in highly sensitive element and isotope measurements.

6.3.4 Seismic Instrumentation

6.3.4.1 Seismology for Both Shallow and Deep Internal Structure

Similar to Earth applications, planetary seismology relies on either active approaches, which frequently use multiple relatively low-sensitivity instruments recording controlled sources, such as explosions or vibrational sources, at known locations, or passive approaches using more widely distributed high-sensitivity instruments recording natural sources such as tectonic quakes and impacts. Active approaches

are widely used on Earth for resource exploration and environmental characterization. Passive approaches have revealed interfaces dividing the Earth's crust, mantle, outer core, and inner core through the first half of the twentieth century.

Seismic data from the Apollo missions to the Moon, still the largest and most widely used planetary seismic dataset, demonstrate these two approaches. Apollo 12, 14, 15, and 16 all placed sensitive seismometers on the lunar surface (Nunn et al. 2020). These instruments operated continuously for 5 + years until being turned off in 1977. They passively recorded an extensive catalog of moonquakes and impacts (Nakamura et al. 1981), which have been used to model the deep internal structure of the Moon for more than four decades (e.g., Goins et al. 1981; Weber et al. 2011; Garcia et al. 2019). Apollo 17, however, included an array of less sensitive geophones, which were used for active source surveying to constrain the upper 10 km of the lunar crust (Kovach and Watkins 1976).

Instruments for a given mission design should be chosen to achieve acceptable signal-to-noise ratio, but will vary greatly depending on the selected mission goals and objectives. Signal strength in active surveys with stations very near controlled sources can be quite large and not require extreme sensitivity, while signal strength in passive experiments will depend on the expected seismicity of the target body, which will vary greatly. Noise for a particular mission will depend on the background noise of the instrument itself, which can be characterized by pre-mission tests (e.g., Lognonné et al. 2019 for the InSight SEIS instruments), but also by the ambient noise of the target body and the style of instrument deployment, which may be difficult to characterize before mission data is returned (e.g., Panning et al. 2020 comparing noise on deck, on the ground and under a wind shield for the InSight mission).

6.3.4.2 Principles of Seismometry

Most seismic instrumentation relies on inertial measurements of translational motion; the simplest example is a mass on a spring. When the supporting frame of the mass (presumed to be coupled with the ground) moves with ground motion, the inertial mass on a spring remains still. For an instrument like the rugged geophones used in active surveys for resource exploration on Earth (e.g., Rodgers 1994 for typical geophone sensitivity levels), the mass is typically magnetic and surrounded by a coiled wire, allowing for measurement of relative motions through magnetic induction. Micro-electromechanical system (MEMS) accelerometers with sensitivity which can be comparable to geophones are widely used in electronic devices like mobile phones (e.g., D'Allesandro and D'Anna 2013). These frequently rely on using a piezoelectric material to measure motions of masses etched from a silicon die. Both geophones and MEMS accelerometers generally have strongly peaked sensitivity near the resonant frequency which is a function of the mass of the test mass and the spring constant.

For passive seismology, where high sensitivity over a broad frequency range is desired, force feedback systems are generally used. There are a variety of designs of these kind of systems, but in general, they are designed with circuits to keep the

mass centered during ground motion, with the feedback system used to record a signal proportional to ground velocity with a flat response which can be designed for high sensitivity from frequencies ranging between 10^{-3} and 100 Hz. This type of sensor was used, for example, for both the very broad-band (VBB) and short-period (SP) instruments of the InSight SEIS experiment on Mars (Lognonné et al. 2019).

6.3.4.3 Planetary Seismic Instrumentation in Past Missions

Moon missions. Seismometers were proposed for the earliest lunar missions, including Ranger 3 (e.g., Lognonné and Pike 2015), but the vast majority of the data came from the Apollo missions. The Apollo 11 mission included a short-lived seismic experiment, but the Apollo 12, 14, 15, and 16 missions all installed long-lived stations with long-period (LP) and short-period (SP) instruments (Nunn et al. 2020). The SP instrument had peak sensitivity near 2 Hz, detecting motions down to just above 10^{-9} m/s² at 2 Hz, while the LP instrument was mainly used in a mode that had a peak sensitivity of approximately 5×10^{-10} m/s² at a resonant frequency of 0.5 Hz. This high instrument sensitivity was sufficient for recording hundreds of moonquakes and impacts, but the overall noise floor of the recordings was usually driven by the coarse 7-bit digitization (Lognonné and Pike 2015). Because the noise floor was driven by the electronics, it is possible the Moon has a much lower ambient noise floor which may be exploited by more sensitive instruments in future missions.

Mars missions. The Viking 1 and 2 landers both included seismometers (Anderson et al. 1977), although, unfortunately, the Viking 1 seismometer did not uncage and never returned data. The Viking 2 seismometer (Fig. 6.7) did return data, but most of the data appeared to be dominated by wind acting on the lander due to the seismometer's placement on the deck. More than 40 years later, the Mars InSight mission (Banerdt et al. 2020) delivered high-quality seismometers back to the surface of Mars, and importantly placed them on the surface under a wind shield, reducing noise levels down to $\sim 10^{-9}$ m/s² between 0.1 and 1 Hz, permitting the measurement of hundreds of identified seismic events (Clinton et al. 2020) (Fig. 6.8).

Venus missions. The Soviet Venera 13 and 14 landers on Venus both included vertical component geophones, although only about an hour of total recording time was achieved across the missions. Only short periods of raw waveforms were returned, as well as a counter for the number of times it crossed a certain threshold. Lorenz and Panning (2018) argued the counter data was compatible with a background noise comparable to background noise on Earth.

6.3.4.4 Future Seismology Technology Development

In the near future, SP and VBB instruments could be adapted for other planetary applications, but multiple other planetary-capable seismometers are currently under development. For example, an extremely sensitive seismic system based on reducing resonant frequency through electrostatic repulsion has been proposed and is under

Fig. 6.7 A view of the Viking 2 seismometer (highlighted by arrow) on the deck of the lander on the surface of Mars (©NASA)

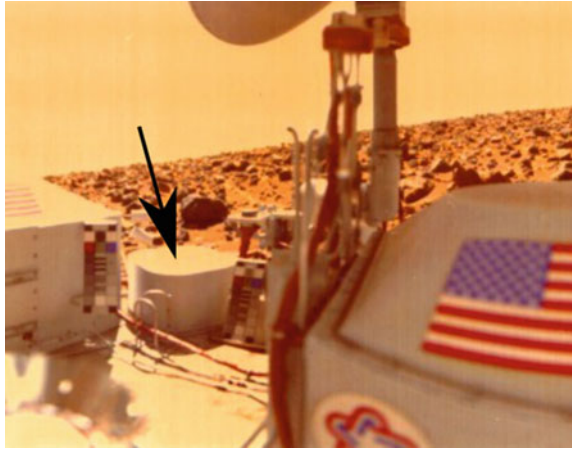
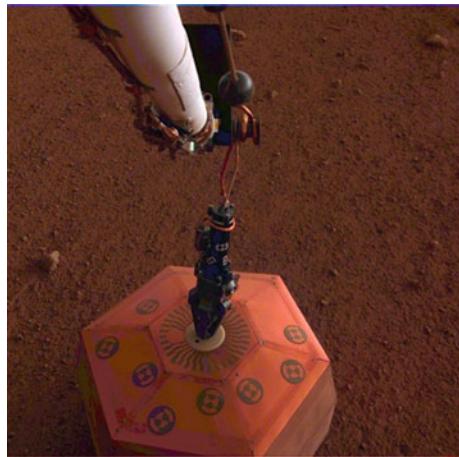


Fig. 6.8 InSight seismometer deployed on the surface of Mars (©NASA)



development as the Planetary Broad Band Seismometer (PBBS), which may produce a system comparable to or exceeding the sensitivity of the VBB instrument (Griggs et al. 2007).

Several smaller, yet still highly capable, seismometers comparable to the SP system are also under development. The SP itself, based on a micromachined silicon die, can be adapted for lower gravity bodies like the Moon or Europa and achieve lower noise levels than achieved on Mars (Nunn et al. 2020). Another system, designed by the commercial firm Silicon Audio, based on rugged geophone development but with increased sensitivity achieved through measuring the proof mass through optical interferometry has been tested in environments analogous to icy moons like Europa (Marusiak et al. 2020). JAXA funded a seismometer which is effectively a very low-noise geophone with a resonant frequency near 1 Hz for use

on a lunar penetrator mission (Yamada et al. 2009). A modified version of the JAXA seismometer is included as a contribution to the upcoming Dragonfly mission to Saturn's moon Titan, currently planned to land in the 2030s (Turtle et al. 2020). The seismometers described above cover a wide range of sensitivity levels and mass and power requirements, allowing future missions to find the appropriate seismometer to reach the individual science requirements.

6.3.5 Nano- and Microtechnology for Habitability and Life-Detection Investigations

6.3.5.1 Principles of Nano- and Microtechnology

Nano- and microtechnology are applications that use materials on the scales of approximately 1–100 nm (1×10^{-9} m) and 1–100 μ m (1×10^{-6} m). Nano- and microstructures and materials are among the smallest objects that can be made and contain unique chemical, physical, and electrical properties compared to similar properties at the macroscale. Planetary science has benefited from significant advances in nano- and microfabrication to develop instrumentation that expands mission capabilities and will enable new mission concepts such as subsurface ocean world exploration.

Specifically, single-photon and single-molecule detection with exceedingly low mass and power consumption have become a reality through the use of nano- and microstructures. Fabrication capabilities are constantly improving (Madou 2011), enabling the creation of novel materials for the platform itself and potential targeting molecules (e.g., nanoparticles, antibodies). While making instrument components as small as possible has the benefit of lowering the size, weight, and power (SWaP) footprint of the payload, the ability to manipulate nano- and microstructures has come along with additional benefits, including higher efficiency sensing, lower shock and vibration sensitivity, the ability to handle extremely small samples, and, in some cases, improved performance (e.g., sensitivity and/or selectivity). While this subsection focuses on the application of instruments in regard to habitability and life detection, nanosensors in planetary exploration covers remote sensing, vehicle performance and monitoring, astrobiological and geochemical research, and human spaceflight (Meyyappan and Dastoor 2004).

6.3.5.2 Current Advances in Nano- and Microtechnology

With the objective of creating miniaturized high-resolution, high-selectivity instrumentation, nano- and microtechnology has relied on significant investment in material science and nanomechanics. This investment has led to the development and testing of miniaturized life-detection instruments both terrestrially and onboard the

International Space Station (ISS) (Maule et al. 2008; Mora et al. 2020). One area that has seen substantial growth is in the field of microfluidic instrumentation. Micron-width channels transport samples or chemicals at nano to picoliter fluid volumes throughout the platform. When paired with other appropriate components and characteristics (e.g., pumps, valves, chemicals, applied electric fields), these lab-on-a-chip technologies complete complicated sample manipulation (e.g., chemical tagging and molecular separations), enabling faster analysis of smaller samples (Daw and Finkelstein 2006) either as direct liquid sampling (e.g., Ocean Worlds) or as solid sample extracts (e.g., Mars). For example, the Lab-on-a-Chip Applications Development (LOCAD) was a set of related lab-on-a-chip advancement activities that included projects that have flown on the ISS and span the fields of environmental control and life systems support (ECLSS), medical systems, and remote exploration (Maule et al. 2008). Many separation and detection techniques are compatible with lab-on-a-chip methods, offering a sensing platform that may be integrated with other payloads. Assessment of the presence and characterization of whole cells, amino acids, carboxylic acids, peptides, and other biomolecules have all been achieved with microfluidic platforms (Maule et al. 2008; Mora et al. 2020).

An example of a microfluidic technology that delivers fully automated microchip electrophoresis separations (sample separation based on solute size and charge) coupled to laser-induced fluorescence is the Chemical Laptop (Mora et al. 2020). Capillary electrophoresis is a convenient way to separate many different molecules of interest, and this technique translates well to microfluidic systems. It is also especially capable of handling water-based samples that may contain salts (Benhabib et al. 2010), whereas other techniques use derivatization reagents that have undesirable side reactions or poor chromatographic separation. The Chemical Laptop is capable of achieving low (less than 8) parts per billion (ppb) detection limits for amino acids (Mora et al. 2020) by coupling microchip electrophoresis with laser-induced fluorescence, thereby offering several orders of magnitude improvement over the current state-of-the-art amino acid detection capabilities on planetary missions (typically, GCMS) (Stalport et al. 2012) (Fig. 6.9).

Additional technologies that may be used in a lab-on-a-chip device or in a standalone instrument are antibody- and aptamer-based sensors. While antibody- and aptamer-based sensors use different targeting molecules (proteins versus short deoxyribonucleic acid (DNA) sequences) and defer in specific characteristics, both molecules are commonly used to identify microorganisms and inorganic and organic compounds at a high selectivity and sensitivity (Byrne et al. 2009). While there are challenges to implementing these forms of sensing in a space exploration mission (e.g., shelf-life and effect of spaceflight conditions), the Life Detector Chip (LDChip) (García-Descalzo et al. 2019) is one example of technology development exploiting antibody-based sensors for life-detection applications. For this instrument, a sample undergoes biochemical extraction and fractionation via sonication in a buffer. The aqueous solution is then introduced to a chip containing an array of antibodies selective for various biomarkers ranging from whole microbial cells to amino acids. If a particular cell or biomarker is present, it will bind to its corresponding antibody. Following analyte–antibody bonding, a second treatment of fluorescently tagged

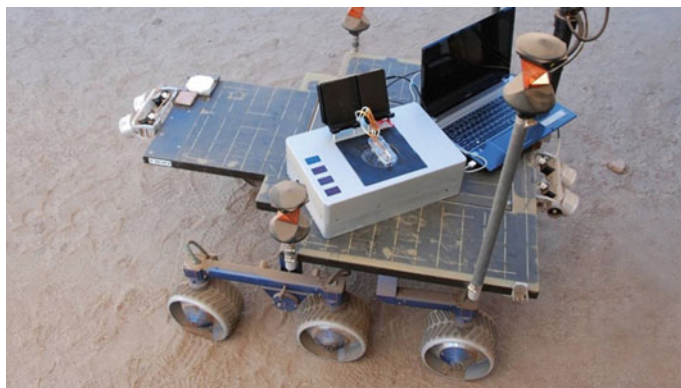


Fig. 6.9 The Chemical Laptop at the Mars Yard at the Jet Propulsion Laboratory (©NASA/JPL-Caltech)

antibodies are added to “sandwich” the bound targeted molecules. Once this process is completed and excess unbound fluorescent antibodies are washed, the colored fluorescence intensity can be measured by region of the chip.

For potential targets such as DNA and ribonucleic acid (RNA), which make up the building blocks of life here on Earth, a technique known as nanopore sequencing is commercially available for terrestrial applications. A complete packaged instrument is obtainable weighing less than 450 g (Oxford Nanopore Technologies). The Search for Extra-Terrestrial Genomes (SETG) instrument seeks to exploit this technology development for life detection on other planetary targets (Carr et al. 2020). This method uses protein nanopores which are extremely small holes that in nature are gateways across membranes. The nanopores are embedded within a membrane that is bathed in an electrophysiological solution. An ionic current is then passed through the nanopores and the movement of the DNA or RNA strands through the pores results in specific electrical signal changes. The change in current is then analyzed to identify the sequence of the strands passing through the nanopore.

Significant developments in other nano- and microstructured sensors for chemical and biomarker sensing including microelectromechanical devices, nanoparticles, and nanotubes have become more common (Jianrong et al. 2004). In addition, improvements in detection techniques such as acoustic, radio frequency, surface-enhanced Raman spectroscopy, electrochemical (Voiculescu and Nordin 2012; Kim et al. 2006; Chao et al. 2016) have been extended into habitability studies, such as nanomaterial-based ion selective electrochemical sensors (Jaramillo and Noell 2020).

6.3.5.3 Future Development in Nano- and Microtechnology

While the principal overall understandings of nano- and micro-technologies continue to advance, researchers are still working to unlock new governing laws at these scales. NASA has identified several challenges that, if explored, may better enable the development and use of nano- and micro-instrumentation. These challenges include material “production and refinement, manipulation and control, lithography, nano-micro-macro integration, toxicology, robust and reliable architectures, self-calibrating networks, and data fusion” (Meyyappan and Dastoor 2004). Beyond the capabilities highlighted in this section, current and future advances in these technologies include lasers, emitters, detectors, optical components, and individual sensing units. With our ability to control the creation of these structures continuously improving and rapidly being integrated into Earth-based systems, it is expected that these nano- and micro-technologies will play an even larger role in flight technologies.

6.4 Summary and Conclusions

Planetary instrumentation has enabled scientists and citizen-scientists to learn more about the universe and beyond. Advances in technology such as miniaturization and early identification instrument requirements (e.g., environmental constraints and planetary protection) have facilitated a long history of successful missions. As reviewed in this chapter, planetary synthetic aperture radar can reveal the surface of worlds that have obscuring atmospheres that make shorter-wavelength imaging not practical. The Cassini spacecraft’s RADAR instrument was able to use SAR to interrogate surface morphologies at high resolution (Lopes et al. 2019). In-situ standoff instruments can record passive reflectance or active spectra from meters away from the target, with no sample preparation. Passive techniques include infrared reflectance and thermal emission spectroscopy (e.g., Mastcam). Active spectrometers use an illumination source to excite a target and produce an emission or adsorption spectrum (e.g., ChemCam onboard the Curiosity rover). Mass spectrometry (MS) using impact-induced ionization is an effective means of detecting inorganic and organic molecules entrained within ice grains, such as previously demonstrated on Cassini. Fluorescence spectroscopy can reveal information about the electronic states of a molecule including the molecular orbital structure of larger structures of a molecule. Raman gives information about the individual functional groups and atom–atom attachments in a molecule, the current state-of-the-art instrument for planetary applications being the Mars Perseverance SHERLOC instrument which acquires both fluorescence and Raman spectra. Planetary seismology relies on either active or passive approaches to record tectonic quakes and impacts, resource exploration and environmental characterization. Nano- and microtechnologies encompass a vast array of growing capabilities and can offer enhanced performance in a package of lower size, weight, and power.

Future complex missions and mission concepts are driving technical instrumentation development to be low powered, robust, compact, and autonomous. As concepts are beginning to develop that aim to explore challenging uncharted bodies (e.g., the surface and subsurface of Ocean Worlds), it is critical to identify risks and dependencies across instrument technology trades and develop baseline system requirements including in instrument integration. Advances in these areas have and will continue to enhance capabilities for science investigations both on Earth and beyond. Through extensive research and innovation spanning government, university, international, and industry partners, planetary instrumentation has delivered major scientific discoveries such as the few that are listed in this chapter.

References

- Anderson, D., W. Miller, G. Latham, Y. Nakamura, and M. Toksöz. 1977. Seismology on Mars. *Journal of Geophysical Research* 82 (28): 4524–4546.
- Arevalo, R., Z. Ni, and R.M. Danell. 2020. Mass spectrometry and planetary exploration: A brief review and future projection. *Journal of Mass Spectrometry* 2020 (55): e4454. <https://doi.org/10.1002/jms.4454>.
- Arvidson, R.E., S.W. Squyres, R.V. Morris, A.H. Knoll, R. Gellert, B.C. Clark, and P. A. de Souza Jr. 2016. High concentrations of manganese and sulfur in deposits on Murray Ridge, Endeavour Crater, Mars. *American Mineralogist*, 101 (6): 1389–1405.
- Banerdt, W.B., S. Smrekar, D. Banfield, D. Giardini, M. Golombek, C. Johnson, and P. Lognonné. 2020. Early results from the InSight mission: Mission overview and global seismic activity. *Nature Geoscience* 13: 183–189.
- Bayer, T., B. Buffington, J.F. Castet, M. Jackson, G. Lee, K. Lewis, and K. Kirby. 2017, March. Europa mission update: Beyond payload selection. In *2017 IEEE Aerospace Conference*, 1–12. IEEE.
- Beegle L, et al. 2015. SHERLOC: Scanning habitable environments with Raman and luminescence for organics and chemicals. In *2015 IEEE Aerospace Conference*, Big Sky, MT, 1–11. <https://doi.org/10.1109/AERO.2015.7119105>
- Bell, J.F., S.W. Squyres, K.E. Herkenhoff, J.N. Maki, H.M. Arneson, D. Brown, et al. 2003. Mars exploration rover athena panoramic camera (Pancam) investigation: MER ATHENA PANORAMIC CAMERA INVESTIGATION. *Journal of Geophysical Research: Planets* 108(E12). <https://doi.org/10.1029/2003JE002070>
- Bell, J.J., J.N. Maki, G.L. Mehall, M.A. Ravine, M.A. Caplinger, Z.J. Bailey, K.M. Kinch, et al. 2020. The Mars 2020 Rover Mast Camera Zoom (Mastcam-Z) Multispectral, Stereo-scopic Imaging Investigation. *Space Science Reviews*.
- Benhabib, Merwan, Thomas N. Chiesi, Amanda M. Stockton, James R. Scherer, and Richard A. Mathies. 2010. Multichannel capillary electrophoresis microdevice and instrumentation for in situ planetary analysis of organic molecules and biomarkers. *Analytical Chemistry* 82 (6): 2372–2379. <https://doi.org/10.1021/ac9025994>.
- Bhartia, R., W.F. Hug, E.C. Salas, R.D. Reid, K.K. Sijapati, A. Tsapin, W. Abbey, K.H. Nealsen, A.L. Lane, and P.G. Conrad. 2008. Classification of organic and biological materials with deep ultraviolet excitation. *Applied Spectroscopy* 62: 1070–1077. <https://doi.org/10.1366/000370208786049123>.
- Bhartia, R., E.C. Salas, W.G. Hug, R.D. Reid, A.L. Lane, K.J. Edwards, and K.H. Nealsen. 2010. Label-free bacterial imaging with Deep UV laser induced native fluorescence. *Applied and Environmental Microbiology* 76: 6231–7237. <https://doi.org/10.1128/AEM.00943-10>.

- Bhartia, R., L. Beegle, L. DeFlores, et al. 2021. Perseverance's scanning habitable environments with Raman and Luminescence for organics and chemicals (SHERLOC) investigation. *Space Science Reviews* 217:58. <https://doi.org/10.1007/s11214-021-00812-z>.
- Brockwell, T.G., K.J. Meech, K. Pickens, et al. 2016. The mass spectrometer for planetary exploration (MASPEX). In *2016 IEEE Aerospace Conference*.
- Byrne, B., E. Stack, N. Gilmartin, and R. O'Kennedy. 2009. Antibody-based sensors: Principles, problems and potential for detection of pathogens and associated toxins. *Sensors (basel, Switzerland)* 9 (6): 4407–4445. <https://doi.org/10.3390/s90604407>.
- Carr, C.E., N.C. Bryan, K.N. Saboda, et al. 2020. Nanopore sequencing at Mars, Europa, and microgravity conditions. *NPJ Microgravity* 6:24. <https://doi.org/10.1038/s41526-020-00113-9>.
- Castillo-Rogez, J.C., S.M. Feldman, J.D. Baker, G. Vane. 2017. Small Instruments for planetary science applications- status and way forward. In *Planetary Science Vision 2050 Workshop 2017 (LPI Contrib. No. 1989)*.
- Chao, J., W. Cao, S. Su, L. Weng, S. Song, C. Fan, and L. Wang. 2016. Nanostructure-based surface-enhanced Raman scattering biosensors for nucleic acids and proteins. *Journal of Materials Chemistry B* 4 (10): 1757–1769. <https://doi.org/10.1039/c5tb02135a>.
- Christensen, P.R., G.L. Mehall, S.H. Silverman, S. Anwar, G. Cannon, N. Gorelick, et al. 2003. Miniature thermal emission spectrometer for the Mars exploration rovers: MARS EXPLORATION ROVER MINI-TES. *Journal of Geophysical Research: Planets* 108(E12). <https://doi.org/10.1029/2003JE002117>.
- Clinton, J.F., S. Ceylan, M. van Driel, D. Giardini, S.C. Stähler, M. Böse, C. Charalambous, N.L. Dahmen. 2020. The marsquake catalog from insight, sols 0–478. *Physics of the Earth and Planetary Interiors*
- D'Alessandro, A., and G. D'Anna. 2013. Suitability of low-cost three-axis MEMS accelerometers in strong-motion seismology: Tests on the LIS331DLH (iPhone) accelerometer. *Bulletin of the Seismological Society of America* 103 (5): 2906–2913.
- Daw, R., and J. Finkelstein. 2006. Lab on a chip. *Nature* 442: 367. <https://doi.org/10.1038/442367a>.
- Dworkin, J.P., L.A. Adelman, T. Ajluni, A.V. Andronikov, J.C. Aponte, A.E. Bartels, E. Beshore, E.B. Bierhaus. 2018. OSIRIS-REx contamination control strategy and implementation. *Space Science Review* 214(1):19. <https://doi.org/10.1007/s11214-017-0439-4>. Epub 2017 Dec 13. PMID: 30713357; PMCID: PMC6350808.
- Elachi, C., M. Allison, Y. Anderson, R. Boehmer, P. Callahan, P. Encrenaz, E. Flameni, and G. Francescetti. 2005a. Cassini RADAR's first view of the surface of Titan. *Science* 13: 970–974.
- Elachi, C., M.D. Allison, L. Borgarelli, P. Encrenaz, E. Im, M.A. Janssen, W.T.K. Johnson, and R.L. Kirk. 2005b. RADAR: The Cassini Titan radar mapper. *Space Science Review* 117: 71–110.
- Elachi, C., S. Wall, M. Janssen, E. Stofan, R. Lopes, R. Kirk, R. Lorenz, and Lunine. 2006. Titan Radar mapper observations from Cassini's T3 fly-by. *Nature* 441: 709–713. <https://doi.org/10.1038/nature04786>.
- Eshelman, E.J., M.J. Malaska, K.S. Manatt, I.J. Doloboff, G. Wanger, M.C. Willis, W.J. Abbey, L.W. Beegle, J.C. Oriscu, and R. Bhartia. 2019. WATSON: In situ organic detection in subsurface ice using deep-UV fluorescence spectroscopy. *Astrobiology* 19: 1–14. <https://doi.org/10.1089/ast.2018.1925>.
- Ford, J.P., et al. 1993. *Guide to magellan image interpretation*: JPL Publ. 93–24, 148.
- Freaner, C., R. Bitten, D. Emmons. 2010. Inherent optimism in early conceptual designs and its effect on cost and schedule growth: an update. In *2010 NASA PM Challenge*, Houston, Texas, Feb 9–10.
- Frydenvang, J., N. Mangold, R.C. Wiens, A.A. Fraeman, L.A. Edgar, C. Fedo, et al. 2020. The chemostratigraphy of the Murray formation and role of diagenesis at Vera Rubin ridge in Gale crater, Mars, as observed by the ChemCam instrument. *Journal of Geophysical Research: Planets* <https://doi.org/10.1029/2019JE006320>
- Garcia, R.F., A. Khan, M. Drilleau, L. Margerin, T. Kawamura, D. Sun, M.A. Wieczorek, A. Rivoldini, C. Nunn, R.C. Weber, A.G. Marusiak, P. Lognonné, Y. Nakamura, and P. Zhu. 2019. Lunar seismology: An update on interior structure models. *Space Science Reviews* 215: 50.

- García-Descalzo, L., V. Parro, M. García-Villadangos, C.S. Cockell, C. Moissl-Eichinger, A. Perras, Rettberg. 2019. Microbial markers profile in anaerobic mars analogue environments using the LDChip (Life Detector Chip) antibody microarray core of the SOLID (Signs of Life Detector) platform. *Microorganisms* 7(9):365. <https://doi.org/10.3390/microorganisms7090365>.
- Gasda, P.J., E.B. Haldeman, R.C. Wiens, W. Rapin, T.F. Bristow, J.C. Bridges, et al. 2017. In situ detection of boron by ChemCam on Mars: First detection of Boron on Mars. *Geophysical Research Letters* 44 (17): 8739–8748. <https://doi.org/10.1002/2017GL074480>.
- Goins, N.R., A.M. Dainty, and M.N. Toksöz. 1981. Lunar seismology: The internal structure of the Moon. *Journal of Geophysical Research* 86 (B6): 5061–5074.
- Griggs, C.E., H.J. Paik, T. Chui, K. Penanen, and J. Young. 2007. Seismometer for strange quark nugget search and for lunar science studies. *Nuclear Physics B-Proceedings Supplements* 166: 209–213.
- Henderson, F.M., A.J. Lewis, (eds.). 1998. *Principles and applications of imaging radar: Manual of remote sensing*, 3rd edn, vol 2. Wiley and Son, New York, NY
- Hensley, S., et al. 2021. A comparison of L-band and S-band interferometry and tomography of the BERMS Borel forest with UAVSAR and F-SAR Datasets, EUSAR 2021. In *13th European Conference on Synthetic Aperture Radar*, 1–4.
- Hibbert, R., M.J. Cole, M.C. Price, et al. 2017. The hypervelocity impact facility at the University of Kent: Recent upgrades and specialized capabilities. *Procedia Engineering* 204: 208–214.
- Hsu, H.-W., F. Postberg, Y. Sekine, et al. 2015. Ongoing hydrothermal activities within Enceladus. *Nature* 519: 207.
- Janssen, M.A., A. Le Gall, R.M. Lopes, R.D. Lorenz, M.J. Malaska, A.G. Hayes, C.D. Neish, A. Solomonidou, K.L. Mitchell, J. Radebaugh, and S.J. Keihm. 2016. Titan's surface at 2.18-cm wavelength imaged by the Cassini RADAR radiometer: Results and interpretations through the first ten years of observation. *Icarus* 270: 443–459.
- Janssen et al., 2009 Janssen, M.A., R.D. Lorenz, R. West, F. Paganelli, R.M. Lopes, R.L. Kirk, C. Elachi, S.D. Wall, W.T.K. Johnson, Y. Anderson, R.A. Boehmer, P. Callahan, Y. Gim, G.A. Hamilton, K.D. Kelleher, L. Roth, B. Stiles, A. Le Gall and the Cassini Radar Team 2009. 2008. Titan's surface at 2.2-cm wavelength imaged by the Cassini RADAR radiometer: Calibration and first results. *Icarus* 200, 222–239. <https://doi.org/10.1016/j.icarus.2008.10.017>
- Jaramillo, E.A., and A.C. Noell. 2020. Development of miniature solid contact ion selective electrodes for in situ instrumentation. *Electroanalysis* 32 (9): 1896–1904. <https://doi.org/10.1002/elan.201900761>.
- Jaramillo-Botero, A., M.L. Cable, A.E. Hofmann, M. Malaska, R. Hodyss, and J. Lunine. 2021. Understanding hypervelocity sampling of biosignatures in space missions. *Astrobiology* 21: 421–442. <https://doi.org/10.1089/ast.2020.2301>.
- Chen Jianrong, Miao Yuqing, He Nongyue, Wu Xiaohua, Li Sijiao. 2004. Nanotechnology and biosensors. *Biotechnology Advances* 22(7). ISSN 505–518 0734 9750. <https://doi.org/10.1016/j.biotechadv.2004.03.004>
- Johnson, J.R., J.F. Bell, S. Bender, D. Blaney, E. Cloutis, L. DeFlores, et al. 2015. ChemCam passive reflectance spectroscopy of surface materials at the Curiosity landing site, Mars. *Icarus* 249: 74–92. <https://doi.org/10.1016/j.icarus.2014.02.028>.
- Junge, K., H. Eicken, and J. Deming. 2004. Bacterial activity at -2 to -20°C in Arctic wintertime sea ice. *Applied and Environmental Microbiology* 70: 550–557. <https://doi.org/10.1128/AEM.70.1.550-557.2004>.
- Khawaja, N., F. Postberg, J. Hillier, et al. 2019. Low-mass nitrogen-, oxygen-bearing, and aromatic compounds in Enceladean ice grains. *Monthly Notices of the Royal Astronomical Society* 489 (4): 5231–5243.
- Kim, Y.I., T.S. Park, J.H. Kang, M.C. Lee, J.T. Kim, J.H. Park, H.K. Baik. 2006. Biosensors for label free detection based on RF and MEMS technology. *Sensors and Actuators B: Chemical* 119(2):592–599. ISSN 0925-4005. <https://doi.org/10.1016/j.snb.2006.01.015>.
- Klenner, F., F. Postber, J. Hillier, et al. 2020a. Discriminating abiotic and biotic fingerprints of amino acids and fatty acids in ice grains relevant to ocean worlds. *Astrobiology* 20 (12): 1–17.

- Klenner, F., F. Postber, J. Hillier, et al. 2020b. Analog experiments for the identification of trace biosignatures in ice grains from extraterrestrial ocean worlds. *Astrobiology* 20 (2): 179–189.
- Kovach, R.L., J.S. Watkins. 1976. Apollo 14 and 16 active seismic experiments, Apollo 17 lunar seismic profiling. Tech. rep., Stanford University. <https://repository.hou.usra.edu/handle/20.500.11753/822>
- Lanza, N.L., W.W. Fischer, R.C. Wiens, J. Grotzinger, A.M. Ollila, A. Cousin, et al. 2014. High manganese concentrations in rocks at Gale crater, Mars. *Geophysical Research Letters* 41 (16): 5755–5763. <https://doi.org/10.1002/2014GL060329>.
- Lanza, N.L., R.C. Wiens, R.E. Arvidson, B.C. Clark, W.W. Fischer, R. Gellert, et al. 2016. Oxidation of manganese in an ancient aquifer, Kimberley formation, Gale crater, Mars: Manganese Fracture Fills in Gale Crater. *Geophysical Research Letters* 43 (14): 7398–7407. <https://doi.org/10.1002/2016GL069109>.
- Le Gall, A., M.J. Malaska, R.D. Lorenz, M.A. Janssen, T. Tokano, A.G. Hayes, M. Mastroguseppe, J.I. Lunine, G. Veyssi re, P. Encrenaz, and O. Karatekin. 2016. Composition, seasonal change, and bathymetry of Ligeia Mare, Titan derived from microwave thermal emission. *Journal of Geophysical Research: Planets* 121: 233–251. <https://doi.org/10.1002/2015JE004920>.
- Legett, C., R. T. Newell, A. L. Reyes-Newell, A. E. Nelson, P. Bernardi, S. C. Bender, et al. 2022. Optical calibration of the SuperCam instrument body unit spectrometers. *Applied Optics* 61 (11), 2967. <https://doi.org/10.1364/AO.447680>.
- Levasseur-Regourd, A.C., J. Agarwal, H. Cottin, C. Engrand, G. Flynn, M. Fulle, and A. Westphal. 2018. Cometary dust. *Space Science Reviews*, 214 (3): 1–56.
- Lognonn , P., W.T. Pike. 2015. Planetary seismometry. In: *Extraterrestrial Seismology*, eds. Tong, V.C.H., R. Garcia, R., 36–48. Cambridge, UK, Cambridge University Press.
- Lognonn , P., W.B. Banerdt, D. Giardini, W.T. Pike, U. Christensen, P. Laudet, S. de Raucourt, S., Zweifel, P., Calcutt, S., 2019. SEIS: The seismic experiment for internal structure of inSight. *Space Science Review* 215.
- Lopes, R.M.C., S.D. Wall, C. Elachi, S.P.D. Birch, P. Corlie, A. Coustenis, A.G. Hayes, and J.D. Hofgartner. 2019. Titan as revealed by the Cassini Radar. *Space Science Reviews* 215: 1–50. <https://doi.org/10.1007/s11214-019-0598-6>.
- Lopes, R.M.L., M.J. Malaska, A.M. Schoenfeld, A. Solomonidou, S.P.D. Birch, M. Florence, A.G. Hayes, D.A. Williams, J. Radebaugh, T. Verlander, E.P. Turtle, A. Le Gall, and S. Wall. 2020. A global geomorphological map of Saturn’s moon Titan. *Nature Astronomy* 4: 228–233.
- Lorenz, R.D., and M.P. Panning. 2018. Empirical recurrence rates for ground motion signals on planetary surfaces. *Icarus* 303: 273–279.
- Madou, M.J. 2011. *Fundamentals of microfabrication and nanotechnology*, vol. 1. Boca Raton, FL: CRC Press.
- Malaska, M.J., R.M.C. Lopes, D.A. Williams, C.D. Neish, A. Solominidou, J.M. Soderblom, A.M. Schoenfeld, S.P.D. Birch, A.G. Hayes, A. Le Gall, M.A. Janssen, T.G. Farr, R.D. Lorenz, J. Radebaugh, and E.P. Turtle. 2016. Geomorphological map of the Afekan Crater region, Titan: Terrain relationships in the equatorial and mid-latitude regions. *Icarus* 270: 130–161. <https://doi.org/10.1016/j.icarus.2016.02.021>.
- Malaska, M.J., R. Bhartia, K.S. Manatt, J.C. Priscu, W.J. Abbey, B. Mellerowicz, J. Palmowski, G.L. Paulsen, K. Zacny, E.J. Eshelman, and J. D’Andrilli. 2020. Subsurface in situ detection of microbes and diverse organic matter hotspots in the Greenland ice sheet. *Astrobiology* 20: 1185–1211. <https://doi.org/10.1089/ast.2020.2241>.
- Malaska et al. 2020 Malaska, M.J., J. Radebaugh, R.M.C. Lopes, K.L. Mitchell, T. Verlander, A.M. Schoenfeld, M.M. Florence, A. Le Gall, A. Solomonidou, A.G. Hayes, S.P.D. Birch, M.A. Janssen, L. Shurmeier, T. Cornet, C. Ahrens, T.G. Farr, Cassini RADAR Team. 2020. Labyrinth Terrain on Titan. *Icarus* 344:113764. <https://doi.org/10.1016/j.icarus.2020.113764>
- Malin, M.C., M.A. Ravine, M.A. Caplinger, F. Tony Ghaemi, J.A. Schaffner, J.N. Maki, et al. 2017. The mars science laboratory (MSL) mast cameras and descent imager: Investigation and instrument descriptions: MSL Mastcam/MARDI descriptions. *Earth and Space Science* 4 (8): 506–539. <https://doi.org/10.1002/2016EA000252>.

- Marusiak, A.G., N.C. Schmerr, D.N. DellaGiustina, E.C. Pettit, P.H. Dahl, B. Avenson, S.H. Bailey, V.J. Bray, N. Wagner, C.G. Carr, and R.C. Weber. 2020. The deployment of the seismometer to investigate ice and ocean structure (SIOS) on Gulkana Glacier, Alaska. *Seismological Research Letters* 91 (3): 1901–1914.
- Maule, J., Wainwright, N., Steele, A., Monaco, L., Morris, H., Gunter, D., Flores, G., Effinger, M., Damon, M., Wells, M., Williams, S. 2008. LOCAD-PTS: Operation of a new system for microbial monitoring aboard the International Space Station (ISS). In *Space 2008 Conference*. <https://doi.org/10.2514/6.2008-7900>.
- Maurice, S., R.C. Wiens, M. Saccoccio, B. Barraclough, O. Gasnault, O. Forni, et al. 2012. The ChemCam instrument suite on the mars science laboratory (MSL) rover: Science objectives and mast unit description. *Space Science Reviews* 170 (1–4): 95–166. <https://doi.org/10.1007/s11214-012-9912-2>.
- Maurice, S., et al. The SuperCam instrument suite on the Mars 2020 rover: Science objectives and mast-unit description. *Space Science Review*.
- Meslin, P.-Y., O. Gasnault, O. Forni, S. Schroder, A. Cousin, G. Berger, et al. 2013. Soil diversity and hydration as observed by ChemCam at Gale Crater, Mars. *Science* 341 (6153): 1238670–1238670. <https://doi.org/10.1126/science.1238670>.
- Meyer, C., A.H. Treiman, and T. Kostiuk, eds. 1996. *Planetary Surface Instruments Workshop. LPI Tech. Rpt.* 95–05. Lunar and Planetary Institute, Houston, 115.
- Meyyappan, M., M. Dastoor. 2004. Nanotechnology in space exploration. In *NSET Report of the National Nanotechnology Initiative Workshop*.
- Mielczarek, P., J. Silberring, and M. Smoluch. 2019. Miniaturization in mass spectrometry. *Mass Spectrometry Reviews* 39 (5–6): 453–470. <https://doi.org/10.1002/mas.21614>.
- Miles, D.M., I.R. Mann, M. Ciurzynski, D. Barona, B.B. Narod, J.R. Bennest, D.K. Milling. 2016. A miniature, low-power scientific fluxgate magnetometer: A stepping-stone to cube-satellite constellation missions. *Journal of Geophysical Research: Space Physics* 121(12). <https://doi.org/10.1002/2016ja023147>
- Miller, M.E.C., R. Hodyss, M.J. Malaska, et al. 2019. Hypervelocity enceladus ice grain analogue production with the aerosol impact spectrometer. In *EPSC-DPS Joint Meeting 2019*, vol 13. Geneva, Switzerland, EPSC-DPS2019–353–1.
- Maria F. Mora, Florian Kehl, Eric Tavares da Costa, Nathan Bramall, and Peter A. Willis. 2020. Fully automated microchip electrophoresis analyzer for potential life detection missions. *Analytical Chemistry* 92(19):12959–12966. <https://doi.org/10.1021/acs.analchem.0c01628>
- McConnochie, T.H., M.D. Smith, M.J. Wolff, S. Bender, M. Lemmon, R.C. Wiens, and J.F. Bell III. 2018. Retrieval of water vapor column abundance and aerosol properties from ChemCam passive sky spectroscopy. *Icarus*, 307: 294–326.
- Nakamura, Y., G.V. Latham, H.J. Dorman, J.E. Harris. 1981. Passive seismic experiment long-period event catalog, final version. Tech. Rep. UTIG Technical Report No. 18. Galveston Geophysics Laboratory of the University of Texas at Austin, Marine Science Institute.
- NASA. 2010. *Draft science instruments, observatories, and sensor system roadmap, technology area 08*.
- NASA. 2011. Planetary protection provisions for robotic extraterrestrial missions, NASA Procedural Requirements, NPR 8020.12D. NASA, Washington, DC.
- Nelson, A.O., R. Dee, M.S. Gudipati, et al. 2016. New experimental capability to investigate the hypervelocity micrometeoroid bombardment of cryogenic surfaces. *Review of Scientific Instruments* 87 (2): 024502.
- New, J.S., R.A. Mathies, M.C. Price, et al. 2020a. Characterizing organic particle impacts on inert metal surfaces: Foundations for capturing organic molecules during hypervelocity transits of Enceladus plumes. *Meteoritics & Planetary Science* 55 (3): 465–479.
- New, J.S., B. Kazemi, M.C. Price, et al. 2020b. Feasibility of Enceladus plume biosignature analysis: Successful capture of organic ice particles in hypervelocity impacts. *Meteoritics & Planetary Science* 1–13.

- Nunn, C., R.F. Garcia, Y. Nakamura, A.G. Marusiak, T. Kawamura, D. Sun, L. Margerin, R. Weber, M. Drilleau, M.A. Wieczorek, A. Khan, A. Rivoldini, P. Lognonné, and P. Zhu. 2020. Lunar seismology: A data and instrumentation review. *Space Science Reviews* 216: 89.
- Palmer, P.T., T.F. Limero. 2001. Mass spectrometry in the U.S. space program: past, present, and future. *Journal of the American Society for Mass Spectrometry* 12(6):656–75. [https://doi.org/10.1016/S1044-0305\(01\)00249-5](https://doi.org/10.1016/S1044-0305(01)00249-5). PMID: 11401157.
- Panning, M.P., W.T. Pike, P. Lognonné, W.B. Banerdt, N. Murdoch, D. Banfield, C. Charalambous, S. Kedar, R.D. Lorenz, A.G. Marusiak, J.B. McClean, C. Nunn, S.C. Stähler, A.E. Stott, T. Warren. 2020. On-deck seismology: Lessons for InSight for future planetary seismology. *Journal of Geophysical Research: Planets* 125(4).
- Postberg, F., S. Kempf, J. Schmidt, et al. 2009. Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus. *Nature* 459: 1098.
- Postberg, F., J. Schmidt, J. Hillier, et al. 2011. A salt-water reservoir as the source of a compositionally stratified plume on Enceladus. *Nature* 474: 620.
- Postberg, F., N. Khawaja, B. Abel, et al. 2018. Macromolecular organic compounds from the depths of Enceladus. *Nature* 558 (7711): 564–568.
- Postberg, F., S. Kempf, J.K. Hillier, et al. 2008. The E-ring in the vicinity of Enceladus: II. Probing the moon's interior—the composition of E-ring particles. *Icarus*, 193(2):438–454.
- Rapin, W., P.-Y. Meslin, S. Maurice, D. Vaniman, M. Nachon, N. Mangold, et al. 2016. Hydration state of calcium sulfates in Gale crater, Mars: Identification of bassanite veins. *Earth and Planetary Science Letters* 452: 197–205. <https://doi.org/10.1016/j.epsl.2016.07.045>.
- Reh, K., L. Spilker, J.L. Lunine, et al. 2016. In Enceladus Life Finder: The search for life in a habitable Moon. In IEEE Aerospace Conference, 1–8, Mar 5–12.
- Rincon, R., et al. 2016. Development of next generation digital beamforming synthetic aperture radar architectures. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2109–2111. <https://doi.org/10.1109/IGARSS.2016.7729544>.
- Rodgers, P.W. 1994. Self-noise spectra for 34 common electromagnetics seismometer/preamplifier pairs. *Bulletin of the Seismological society of America* 84 (1): 222–228.
- Roman, P., W. Brinckerhoff, F. Herrero, R. Hu, H. Jones, D. Kahle, T. King, and Mahaffy, P. 2008. A miniature MEMS and NEMS enabled time-of-flight mass spectrometer for investigations in planetary science—art. no. 69590G. In *Proc SPIE*. 6959. <https://doi.org/10.1117/12.777670>.
- Ruff, S.W., J.D. Farmer, W.M. Calvin, K.E. Herkenhoff, J.R. Johnson, R.V. Morris, and S.W. Squyres. 2011. Characteristics, distribution, origin, and significance of opaline silica observed by the Spirit rover in Gusev crater, Mars. *Journal of Geophysical Research: Planets*, 116 (E7).
- Sapers, H.M., J. Razzel-Hollis, R. Bhartia, L. Beegle, V.J. Orphan, and J. Amend. 2019. The cell and the sum of its parts: Patterns of complexity in biosignatures as revealed by Deep UV Raman spectroscopy. *Frontiers in Microbiology* 10: 679. <https://doi.org/10.3389/fmicb.2019.00679>.
- Sautter, V., M.J. Toplis, R.C. Wiens, A. Cousin, C. Fabre, O. Gasnault, et al. 2015. In situ evidence for continental crust on early Mars. *Nature Geoscience* 8 (8): 605–609. <https://doi.org/10.1038/ngeo2474>.
- Schoenfeld, A.M., R.M.C. Lopes, M.J. Malaska, A. Solomonidou, D.A. Williams, S.P.D. Birch, A.G. Hayes, P. Corlies, A. Le Gall, M.A. Janssen, S. Le Mouélic Le, E. Turtle, M. Florence, T. Verlander. 2021. Geomorphological map of the South Belet Region of Titan. *Icarus* 366:114516. <https://doi.org/10.1016/j.icarus.2021.114516>
- Schröder, C., D.S. Rodionov, T.J. McCoy, B.L. Jolliff, R. Gellert, L.R. Nittler, and T. Economou. 2008. Meteorites on Mars observed with the Mars exploration rovers. *Journal of Geophysical Research: Planets* 113 (E6).
- Sekine, Y., T. Shibuya, F. Postberg, et al. 2015. High-temperature water–rock interactions and hydrothermal environments in the chondrite-like core of Enceladus. *Nature Communications* 6: 8604.
- Srama, R., T.J. Ahrens, N. Altobelli, et al. 2004. The Cassini cosmic dust analyzer. *Space Science Reviews* 114 (1): 465–518.

- Stalport, F., D.P. Glavin, J.L. Eigenbrode, D. Bish, D. Blake, P. Coll, C. Szopa, A. Buch, A. McAdam, J.P. Dworkin, P.R. Mahaffy. 2012. The influence of mineralogy on recovering organic acids from Mars analogue materials using the “one-pot” derivatization experiment on the Sample Analysis at Mars (SAM) instrument suite. *Planetary and Space Science* 67(1):1–13. ISSN 0032-0633. <https://doi.org/10.1016/j.pss.2012.02.010>
- Stone, 2017 Stone, E. 2017. The voyagers. *Nature Astronomy* 1:896. <https://doi.org/10.1038/s41550-017-0339-2>
- Tulej, M., N.F.W. Ligterink, C. de Koning, V. Grimaudo, R. Lukmanov, P. Keresztes Schmidt, A. Riedo, and P. Wurz. 2021. Current progress in femtosecond laser ablation/ionisation time-of-flight mass spectrometry. *Applied Sciences*. 11 (6): 2562. <https://doi.org/10.3390/app11062562>.
- Turtle, E.P., M.G. Trainer, J.W. Barnes, R.D. Lorenz, K.E. Hibbard, D.S. Adams, P. Bedini, W.B. Brinckerhoff. 2020. Dragonfly: In situ exploration of Titan’s organic chemistry and habitability. In: *51st Lunar and Planetary Science Conference. Lunar and Planetary Inst.*, Houston, TX, Abstract #2288.
- Ulibarri, Z., T.L. Munsat, B. Abel, et al. 2019. On the genesis and detectability of organic chemistry in hypervelocity impact ice spectra. In *AGU Fall Meeting 2019*, San Francisco, CA, USA, vol abstract #P43C-3485.
- Voiculescu, Ioana, Anis Nurashikin Nordin. 2012. Acoustic wave based MEMS devices for biosensing applications. *Biosensors and Bioelectronics* 33(1):1–9. ISSN 0956-5663. <https://doi.org/10.1016/j.bios.2011.12.041>.
- Waite, J.H., M.R. Combi, W.-H. Ip, et al. 2006. Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure. *Science* 311 (5766): 1419–1422.
- Waite, J.H., W.S. Lewis, B.A. Magee, et al. 2009. Liquid water on Enceladus from observations of ammonia and 40Ar in the plume. *Nature* 460: 487.
- Waller, S.E., N.R. Tallarida, J.L. Lambert, et al. 2020. In Analyzing Enceladus’ plume: First steps to experimentally simulating hypervelocity impacts ACS Spring. In *2020 National Meeting & Expo, Philadelphia, PA, USA*, Mar 22–26, American Chemical Society, Philadelphia, PA, USA.
- Weber, R., P.-Y. Lin, E. Garnero, Q. Williams, and P. Lognonné. 2011. Seismic detection of the lunar core. *Science* 331 (6015): 309–312.
- Wellington, D.F., J.F. Bell, J.R. Johnson, K.M. Kinch, M.S. Rice, A. Godber, and C. Hardgrove. 2017. Visible to near-infrared MSL/Mascam multispectral imaging: Initial results from select high interest science targets within Gale Crater, Mars. *American Mineralogist*, 102 (6): 1202–1217.
- Wiens, R.C., S. Maurice, B. Barraclough, M. Saccoccio, W.C. Barkley, J.F. Bell, et al. 2012. The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: body unit and combined system tests. *Space Science Reviews* 170 (1–4): 167–227. <https://doi.org/10.1007/s11214-012-9902-4>.
- Wiens, R.C., S. Maurice, S.H. Robinson, A.E. Nelson, P. Cais, P. Bernardi, et al. 2021. The SuperCam instrument suite on the NASA Mars 2020 Rover: Body unit and combined system tests. *Space Science Reviews* 217 (1): 4. <https://doi.org/10.1007/s11214-020-00777-5>.
- Yamada, R., I. Yamada, H. Shiraishi, S. Tanaka, Y. Takagi, N. Kobayashi, N. Takeuchi, Y. Ishihara, H. Murakami, K. Yomogida, K. Koyama, A. Fujimura, and H. Mizutani. 2009. Capability of the penetrator seismometer system for lunar seismic event observation. *Planetary and Space Science* 57 (7): 751–776.
- Zubarev, R.A., A. Makarov. 2013. Orbitrap mass spectrometry. *Analytical Chemistry* 85(11):5288–5396. <https://doi.org/10.1021/ac4001223>. Epub 2013 May 13. PMID: 23590404.

Chapter 7

Space Debris Recycling by Electromagnetic Melting



**Jan Walter Schroeder, Gary Douglas Calnan, Abdoul-Aziz Bogno,
Toby Joseph Daniel Mould, Romain Pecher Pecher, Joseph W. Pawelski,
and Kai Staats**

Abstract For the commercial space ecosystem to achieve industrial scale a complete value chain for space-resources processing must be established, including the capability for refining and processing structural materials, such as metal, on orbit. Through on-orbit salvage and recycling, space debris can be transformed into a space-derived metal resource. A platform for processing metal would be able to not only reduce the amount of space debris, but also create a supply of necessary materials for a variety of in-space manufacturing and satellite servicing missions. This chapter describes the physical principles of processing metal with electromagnetic levitation as well as its implications for future facilities in orbit. The Electromagnetic-Levitor (EML) on the ISS is discussed and possible designs for future EMLs are presented. Further, the possible applications of metal as a potential source of propellant for electric propulsion systems and as feedstock for in-space manufacturing are discussed. In addition,

J. W. Schroeder (✉) · T. J. D. Mould
CisLunar Industries Inc., Feckersbach 12, 66625 Sötern, Germany
e-mail: walter@cislunarindustries.com

T. J. D. Mould
e-mail: toby@cislunarindustries.com

G. D. Calnan
CisLunar Industries Inc., 137 S Ogden St, 80209 Denver, CO, USA
e-mail: gary@cislunarindustries.com

A.-A. Bogno
CisLunar Industries Inc., 309 Cumberland Street, Apt 705, K1N7J1 Ottawa, Ontario, Canada
e-mail: ablaziz@cislunarindustries.com

R. P. Pecher
CisLunar Industries Inc., Avenue des Hannetons 7, 1170 Watermael-Boitsfort, Belgium
e-mail: roman.pecher@cislunarindustries.com

J. W. Pawelski
CisLunar Industries Inc., 704 E. Elizabeth St., 80524 Fort Collins, CO, USA
e-mail: joe@cislunarindustries.com

K. Staats
CisLunar Industries Inc., 5840 North Cascabel Road, 85602 Benson, AZ, USA
e-mail: kai@cislunarindustries.com

policy and law perspectives are explored, looking at the liability and ownership issues that can arise when processing space debris.

7.1 Introduction

7.1.1 Motivations

The cislunar economy is beginning to take shape, and a new frontier is opening up with the pursuit of valuable space resources. Utilization of space resources will be at the core of this new ecosystem, supplying the materials needed to develop a self-sustaining economy in space. A critical component in this ecosystem will be an industrial scale, in-space processing and refining capability for structural materials. Such a capability will help to enable a cislunar economy less dependent upon components delivered from Earth.

This chapter elaborates on the hypothesis that space debris should be treated as a viable near-term, in-space source of materials and components enabled by on-orbit salvage and recycling facilities. Whether free flying or hosted on a shared platform, such facilities will receive recovered derelict space objects, salvage any parts or components that can be reused, and recycle the remaining materials. In doing so, these facilities will consume space debris, aiding in its clean up, while creating a supply of necessary materials for a variety of in-space manufacturing and satellite servicing missions.

7.1.2 Problem of Space Debris

Space debris is a well-documented and growing problem for the ongoing exploration and use of Earth orbits (Aerospace Corporation 2018). It represents a safety risk for all space objects, which will only escalate as the utilization of the space environment accelerates. The increasing amount of micro-debris (less than 1 mm in diameter) in low Earth orbit (LEO), with average impact velocity of 10 km/s, can easily inflict critical damage on a satellite (Corbett 2015). This problem is therefore receiving attention from space agencies, governments, and institutions.

There are over 8,800 tonnes of space debris distributed in various orbits around the Earth (ESA 2020). While this debris poses a growing risk to space activity, it also represents a large cache of recyclable space-grade material. Space debris consists of 80% abandoned spacecraft and rocket bodies (Anselmo and Pardini 2016) containing components and systems that could potentially be reclaimed and directly repurposed. Low Earth orbit (LEO) and geostationary equatorial/synchronous orbit (GEO/GSO) are of particular interest for active debris removal and satellite servicing missions. In LEO, the Kessler syndrome threatens the usability of orbits for future constellations

and could end the human presence in space. In GEO there are more than 1,000 trackable debris objects larger than 1 m that can act as a collision hazard for GEO satellites essential to modern infrastructure like television and communication (Jakhu et al. 2017). Finding an appropriate method of removing debris wherever possible will be essential to the continued and sustainable use of space.

7.1.3 Current State of Space Debris Remediation

Space debris remediation refers to the active removal of existing, non-functional, man-made space objects. In contrast, space debris mitigation refers to the prevention of new space debris through adherence to best practices and the use of passive or active deorbiting technologies on new spacecraft. In this subsection the focus is on remediation and how treating space debris as a resource could enhance remediation efforts.

Studies first conducted and published by NASA in 2010 (Liou et al. 2010) showed that the risks posed by space debris could be significantly reduced by focusing on removing the largest objects in the most populated orbits. Each one of these objects would add thousands of detectable debris fragments in a collision. The study concluded that “if one assumes that the mitigation measures are very well complied to (no fragmentation in orbit, 25-year rule in LEO), then the retrieval of 5 to 10 properly chosen large debris from the most populated orbits would be enough to stabilize the orbital population.” These large objects also make the best candidates for in-space salvage and recycling.

Effective space debris remediation will require the use of active debris removal (ADR) systems which use a variety of techniques to alter the existing orbit of an object, causing it either to deorbit or move to a safer “graveyard” orbit.

Several missions have either flown or are in the final planning stages, demonstrating active and passive debris removal techniques. The RemoveDEBRIS mission, launched in April 2018, was the first satellite on-orbit to demonstrate active (harpoon, net, LIDAR) debris removal technology in space (Forshaw et al. 2016). This mission also demonstrated passive (dragsail) mitigation technology to deorbit the demonstrator at the end of the mission (Viquerat et al. 2015). Other concept demonstrators include CleanSpace-1, intended to use a collapsible net to capture a 1U cubesat, and e.Deorbit, which is currently investigating technological links between debris remediation efforts and on-orbit servicing (Richard et al. 2013). Other proposals include the use of deboosting thrusters, high-power ground lasers, and a hybrid rocket propulsion system onboard a chaser spacecraft that rendezvous with debris and performs active removal (Tonetti et al. 2018).

The End-of-Life Services by Astroscale (ELSA) aims to retrieve spacecraft from satellite operators for a fee. The first demonstration mission, ELSA-d, is planned for 2020 and will demonstrate debris docking and removal (Blackerby et al. 2018).

Development of reusable orbital transfer vehicles (OTVs) is also advancing rapidly, with several companies engaged. Altius Space Machines is developing the

Bulldog space tug, and Effective Space (acquired by Astroscale) developed the SPACE DRONE. Northrop Grumman Innovation Systems have developed a mission extension vehicle (MEV) which executed the first successful satellite servicing mission in GEO in 2020. Space Systems Loral is developing robotic servicing of geosynchronous satellites (RSGS) for GEO and Restore-L for LEO. Airbus Defense and Space are also developing a space tug and including the removal of derelict satellites from important orbital locations as part of their servicing objectives.

While none of these solutions are being developed to treat space debris as a resource, many of them employ technologies that could be repurposed to deliver space debris to a recycling facility rather than disposing of it.

7.1.4 Space Debris and the Industrialization of Cislunar Space

Development of a self-sustaining space economy is a critical path for humankind if we are to become a space-faring species. A cost-effective, readily available supply of refined materials in orbit is foundational to the establishment of a self-sustaining industrialized space economy. There have been numerous proposals to supply materials in orbit from non-terrestrial feedstocks. While feedstocks will eventually be derived from lunar regolith and other natural resource-rich celestial bodies, space debris represents an in-situ space resource that may be more accessible in the near term. Orbital salvage and recycling platforms offer the practical means to process space debris into refined metal and useful salvaged parts to accelerate the development of an industrialized cislunar economy.

7.2 Space Debris Metal Processing

A space debris salvage, recycling, and processing facility lies at the center of a broader logistics ecosystem with specialized spacecraft capturing and delivering derelict satellites and upper stages to platforms and finished materials to customers. After taking control of a space object, a recycling platform will analyze the object to assess salvageable components and material composition. Components or structural elements which can be salvaged and used “as is” will be removed and stored. The remaining material will be broken down further and reprocessed into products aimed at the in-space manufacturing and construction sectors. At the beginning of the 2020s, this in-space industrial economy is just beginning to emerge, and it is expected to fully develop over the coming decades.

The following section will focus on the metals processing phase of this system and more specifically on a containerless electromagnetic levitation method. This technique offers many advantages in microgravity and/or hard vacuum conditions

over other techniques, and has been the object of a long-running experiment on the ISS: the Electro-Magnetic Levitator (EML).

7.2.1 *The Electro-Magnetic Levitator (EML) on the ISS*

The Material Science Laboratory Electromagnetic Levitator (MSL-EML), EML for short, is a levitation furnace that was jointly developed by the European Space Agency (ESA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The unit was built by Airbus DS and installed in the Columbus science laboratory on the ISS in 2014 by ESA astronaut Alexander Gerst during the Blue Dot Mission (Diefenbach et al. 2020). Design of the EML is based on a predecessor called the TEMPUS experiment (*Tiegelfreies Elektromagnetisches Prozessieren unter Schwerelosigkeit*) which was developed by the DLR and served as a technology demonstrator for the EML.¹ It has been used for containerless processing of metals, alloys, and semiconductors on the ISS since April 2015, to study melting and solidification properties of these various materials in microgravity.

Electromagnetic levitation is achieved by using alternating-current coils to generate high-frequency electromagnetic fields surrounding a sample. These alternating fields induce eddy currents in the sample, which heat the sample through Ohmic losses and generate forces on the sample's external surface² directed towards the center of the coil system. Note that any material with sufficiently high electrical conductivity³ can be levitated and heated in this way; the sample need not be magnetic itself, meaning that levitation above a Curie temperature in the molten state is just as easy for ferromagnetic materials as for non-ferromagnetic materials. It is worth noting that the magnetization hysteretic losses are significant for most materials and will aid in the heating until the Curie temperature is reached. All of these allow melting and levitation of a wide range of conductive metals, setting EML apart from other magnetic confinement methods.

This EML effect works best with non-ferromagnetic and highly conductive materials, like copper or aluminum. Indeed, ferromagnetic materials like iron, nickel, cobalt and rare-earth magnetic alloys have the disadvantage of also being strongly attracted to the coils, opposing the EML repulsion effect. They also usually have higher resistivity and consequently less developed eddy currents (which will be discussed in the next section). With high enough EM field frequencies, even ferromagnetic materials can be levitated. Note that most steels, predominantly composed of iron, are ferromagnetic, except for austenitic stainless steels.

¹ While mostly Earth based, versions of the TEMPUS experiment were also tested in parabolic flights, sounding rockets and finally on three Spacelab missions in the 1990s.

² Eddy currents are a skin effect, concentrated at the surface and absent from the inner bulk of the sample.

³ In practice this is limited to good electronic conductors, although it should be possible with high enough frequencies even to levitate relatively good ionic conductors.

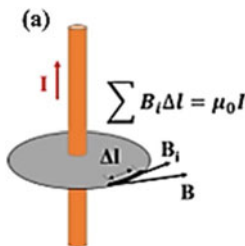
Electromagnetic levitation can be achieved both on Earth and in microgravity, but levitating a sample under normal gravity conditions requires strong electromagnetic fields, inducing convection effects which may create unwanted rapid heating and deformation of the sample's shape. The presence of the gravity force vector only allows for positioning in a metastable equilibrium at best. Notably in microgravity, the sample is unaffected by body forces, takes on an undeformed spherical shape, and only requires very small positioning forces to remain in a stable equilibrium at the center of the coil system. This differentiation allows for much finer control of both positioning and heating rate, and the absence of convection and sedimentation effects in the levitated sample before and during solidification allows for micro-structural homogeneity unmatched by any Earth-based process.

While the MSL-EML itself is a materials research system, the electromagnetic levitation technology behind it is exceptionally well suited to general-purpose metals processing in space. The containerless approach significantly reduces weight requirements for the furnace, while the reduced material degradation of the inner surfaces and absence of moving parts minimize maintenance requirements. The following subsection will examine the physics of EML technology in greater detail and explore how its principles can be transposed to general-purpose metals processing and expanded upon to be used for space debris recycling.

7.2.1.1 Basics of Electromagnetic Levitation

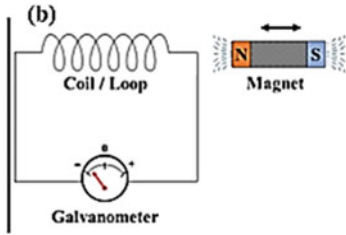
As stated before, electromagnetic levitation is achieved through the application of high-frequency alternating fields on an electrically conducting sample. This process relies on three basic laws governing electromagnetics:

- *Ampère's Law*: relates an integrated magnetic field around a closed loop conductor to the electric current passing through the loop. This is the law that describes how coils generate magnetic fields when a current is passed through them (Fig. 7.1).
- *Faraday's Law*: states that when an electrical conductor is subjected to a time-varying magnetic field, this induces an electromagnetic field (EMF) in that conductor. This induced EMF in turn generates looping currents within the conductor, called eddy currents (Fig. 7.2).



$$\nabla \times B = \mu_0 I_{Enclosed}$$

Fig. 7.1 Schematic representation and differential form of Ampère's Law



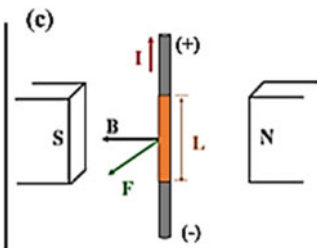
$$emf = -N \frac{\Delta\Phi}{\Delta t}$$

Fig. 7.2 Schematic representation and differential form of Faraday’s Law

- *The Lorentz Force:* describes the force exerted on a charged particle moving through a magnetic field. Eddy currents induced in a conductor through the Faraday principle are in effect just charged particles (generally electrons) looping around in a circle and thus experience the Lorentz force, as a result of the interaction of these eddy currents with the applied high-frequency alternating electromagnetic field. The resulting force exerted on the conductor happens to be directed in the direction pointing away from the coil inducing the eddy currents. As such, the Lorentz force can be used to position an electrical conductor in between two coils (Fig. 7.3).

To synthesize: when an electrically conducting sample is placed in an alternating magnetic field, eddy currents are induced in the sample. These eddy currents interact in turn with the applied electromagnetic field, resulting in a Lorentz force exerted on the sample which can be used to keep the sample in a defined position or control its movement in 3D space.

The eddy currents induced in the sample not only generate the Lorentz force, which can be used for positioning, but generally also lead to significant heating of the sample through Ohmic losses, which in the absence of cooling ultimately leads to the melting of the sample. This process of containerless and contactless melting is known as electromagnetic levitation or levitation melting, and was first suggested by Muck in 1923. It took another 30 years for the first experimental and theoretical work on the subject to be published by Okress et al. (1952). A more in-depth theoretical treatment of the electromagnetic levitation technique was proposed by Rony (1965) in a paper



$$F = q(E + v \times B)$$

Fig. 7.3 Schematic representation and differential form of the Lorentz force equation

that can be considered as the foundation of electromagnetic levitation science. Note that while the term “electromagnetic levitation” is used for historical reasons, because the first attempts all used the induced Lorentz force to counter gravity, effectively levitating the sample, this denomination is not precisely accurate in microgravity, where “electromagnetic positioning” would perhaps be a more appropriate name. In this subsection the common historical name electromagnetic levitation will be used for practical reasons.

Induced eddy currents are not equally distributed throughout the sample bulk. Due to interaction between opposing eddy currents, the higher the frequency of the applied alternating electromagnetic field, the more the eddy currents will concentrate near the surface of the sample. This effect is called the skin effect (eddy currents tend to flow through the outer skin of a conductor); the skin depth is defined as the distance from the conductor’s surface where the induced eddy current density has fallen to $1/e$, or about 37% of the maximum current density at the surface. In the case of electromagnetic levitation, where high frequency is used, this skin effect is so pronounced, and the associated skin depth so small, that the eddy currents can be considered to be flowing exclusively at the sample’s surface. An interesting consequence of this is that when the sample is spherical, as is the case in microgravity,⁴ the magnetic field generated by eddy currents in the sample is equivalent to that produced by either an equatorial current or an alternating magnetic dipole (Okress et al. 1952). Consequently, the eddy current induced in a spherical sample can be treated as approximately equivalent to an equatorial current, a fact which greatly simplifies power absorption calculations (see sect. 7.2.1.2).

7.2.1.2 Electromagnetic Positioning and Heating

A typical Earth-based electromagnetic levitation set-up, where the force that is generated on the sample is used to counter gravity, is depicted in Fig. 7.4. The lower coils provide the bulk of the upward Lorentz force F_L countering the gravity force F_G , while the upper coils, where current flows in the opposite direction, generate a small counter-force on the sample to maintain it in its metastable equilibrium. Without the upper coils, the sample would be in an unstable equilibrium with regard to gravity, and thus fall off the lower coils at the slightest perturbation.

While Rony et al. (1965) provide a suitable model to calculate the levitation force and power absorption in an idealized set-up with a small and perfectly spherical sample such as in Fig. 3, a more detailed and polyvalent model is provided by Brisley and Thornton (1963) and Fromm and Jehn (1965), for the levitation force and power absorption respectively. These will be further explored in the next section.

A system like that depicted in Fig. 7.4 can effectively levitate an electrically conducting sample but at the same time will also heat up the sample considerably,

⁴ For electromagnetic levitation in a gravitational field, where gravity deforms the shape of the sample once it has melted, this is an approximation. Depending on the chosen experimental set-up, this can be a very good approximation (small sample size, additional coils, ...).

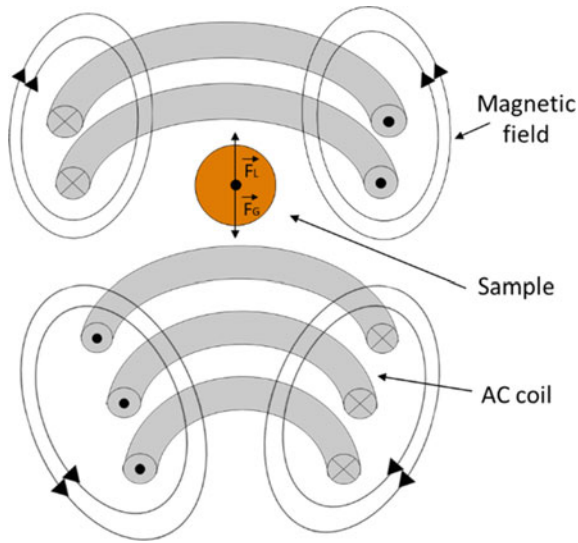


Fig. 7.4 Schematic description of a typical EML set-up, with FL and FG representing the levitation force and the gravitational force respectively (Herlach 1991)

eventually leading to the sample melting. In practical application it is, however, more interesting to be able to decouple positioning from heating, and to be able to control each independently of the other.

Both the ISS-based MSL-EML and its Earth-based TEMPUS predecessor use two separate sets of coils to control positioning on the one hand, and heating on the other hand. Positioning control is achieved by generating a quadrupole field surrounding the sample, while heating is achieved with a superimposed dipole field. In the simplest of set-ups, such fields can be generated with just two sets of series-connected coils, four coils in total. A quadrupole field can be generated between two parallel coils with currents flowing in opposite directions, as can be seen in Fig. 7.5a, while a dipole field can be generated between two parallel coils with currents flowing in the same direction, as illustrated in Fig. 7.5b (Lohöfer 2018).

This set-up allows for independent tuning of the positioning and heating coils. To improve decoupling of positioning and heating, the quadrupole positioning field is usually operated with a strong gradient but low field strength, while the dipole field is kept almost homogeneous (no gradient) but with strong field strength. On Earth, the presence of the gravity force vector does not allow for a complete decoupling of positioning and heating, as the positioning field strength has to be strong enough to counteract gravity, and thus invariably ends up contributing significantly to the heating of the sample. Meanwhile, in microgravity, the sample experiences no net body forces and positioning can be achieved with very low field strength. This reduces the contribution of the positioning field to heating of the sample to insignificant levels and thus allows a complete decoupling of positioning and heating in practice. This simple coil arrangement is well suited to materials research in microgravity but only

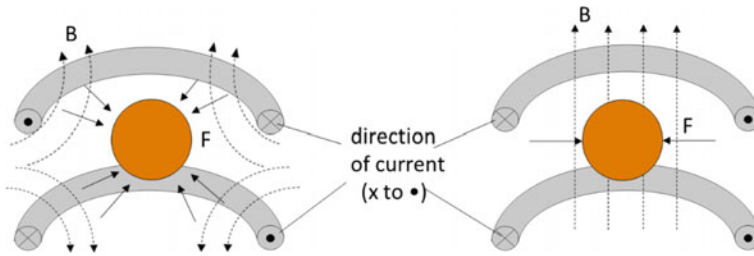


Fig. 7.5 **a** Quadrupole positioning field between two parallel coils with opposite current directions. **b** Dipole heating field generated by two parallel coils with same current directions (Lohöfer 2018)

allows for positioning at the center of the coil system. As such, new coil designs are required to make use of electromagnetic levitation for more complex metals processing in microgravity.

7.2.1.3 Electromagnetic Melting

A consequence of the skin effect described in sect. 7.2.1.1 is that the concentration of induced eddy currents near the sample's surface reduces the effective cross-section through which these currents flow, increasing Ohmic losses. The higher the skin effect, the smaller the effective cross-section becomes, leading to more energy being lost to Ohmic resistivity, and the more the sample heats up. This explains why at identical field frequencies,⁵ positioning coils must operate at low field strengths and heating coils at high field strength. A high field strength induces a stronger skin effect, increasing the eddy current concentration at the sample's equator and thus leading to a higher fraction of the energy from the applied electromagnetic field being converted into heat. A low field strength limits resistive heating and can thus be used primarily for positioning/levitation instead.

Since the skin effect reduces the induced eddy currents to a quasi-equatorial current, heating is also concentrated at the sample's equator. This means that the sample usually starts melting at its equator, from which the melting front then expands to the rest of the sample. It also allows for a relatively easy estimation of power absorption in the sample by approximating the induced eddy currents by an equivalent equatorial current. Substituting the induced eddy currents by an equivalent equatorial current in Brisley's equation for the levitation force (Brisley and Thornton 1963) and Fromm's equation for power absorption (Fromm and Jehn 1965), and using the coordinates shown in Fig. 7.6, results in the following two equations:

Levitation force on the sample, $F(x, z)$:

⁵ Note that both positioning and heating require high-frequency alternating fields, as both rely on the eddy currents induced in the sample. The way to either prioritize the Lorentz force (in case of positioning) or Ohmic losses (in case of heating) is by playing on the field strength.

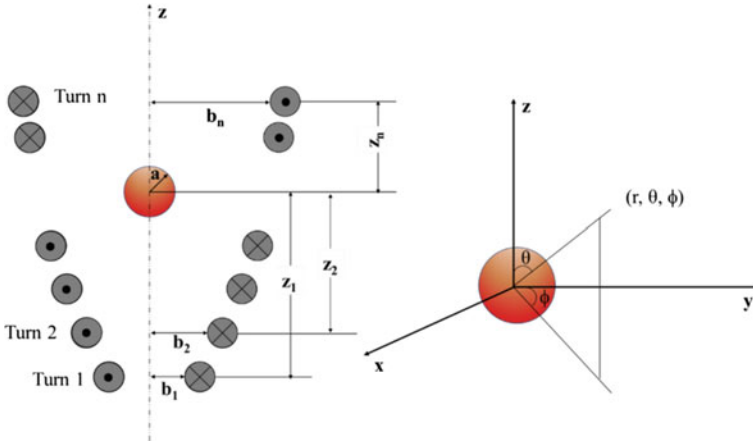


Fig. 7.6 Diagram of the equivalent electrical system with considered x, y and z axes

$$F(x, z) = \frac{3}{2} \pi \mu_0 I^2 a^3 G(x) A(z) \tag{7.1}$$

Power absorption in the sample, P(x, z):

$$P = \frac{3}{4} \pi a \tau I^2 F(x) B(z) \tag{7.2}$$

where:

$$A(z) = \sum_{i=1}^n \frac{\epsilon_i b_i^2}{(b_i^2 + z_i^2)^{\frac{3}{2}}} \sum_{i=1}^n \frac{\epsilon_i b_i^2 z_i}{(b_i^2 + z_i^2)^{\frac{5}{2}}} \tag{7.3}$$

$$B(z) = \left(\sum_{i=1}^n \frac{\epsilon_i b_i^2}{(b_i^2 + z_i^2)^{\frac{3}{2}}} \right)^2 \tag{7.4}$$

$$G(x) = 1 - \frac{3(\sinh \sinh 2x - \sin \sin 2x)}{4x (\sinh^2 x - \sin^2 x)} \tag{7.5}$$

$$F(x) = \frac{x(\sinh 2x - \sin 2x)}{\cosh \cosh 2x - \cos \cos 2x} - 1 \tag{7.6}$$

and with:

I = current in the levitation coil (A)

a = radius of the spherical sample (m)

τ = electrical resistivity of the sample ($\Omega.m$)

$$x = \frac{a}{\delta}; \text{ with } \delta = \text{skin depth (m)} = \sqrt{\frac{\tau}{\pi \mu_0 f_0}};$$

and f_0 = current frequency in the levitation coil (Hz)

b_i = radius of the i th coil (m)

z_i = distance from sample centre to plane occupied by the i th coil (m)

ε_i = a coefficient equal to -1 for the (upper) stabilisation coils and
 $+1$ for the (lower) levitation coils

The total levitation force and power absorption profiles of the levitated sample can be calculated from Eqs. (7.1) and (7.2), respectively. For EML in microgravity, where positioning and heating systems are almost completely decoupled, the required positioning forces are minimal and consequently, positioning power requirements are negligible compared to heating power requirements. Therefore, most of the power requirement of an in-space EML system pertains to melting the levitated sample, and the total energy consumption of a microgravity EML system can be satisfactorily approximated by calculating power absorption in the sample only.

From the above equations, the power absorption varies with the electrical resistivity of the sample material. It can be demonstrated that the larger this electrical resistivity, the more power is absorbed by the sample from the levitation coil's electromagnetic field (Nan et al. 1999).

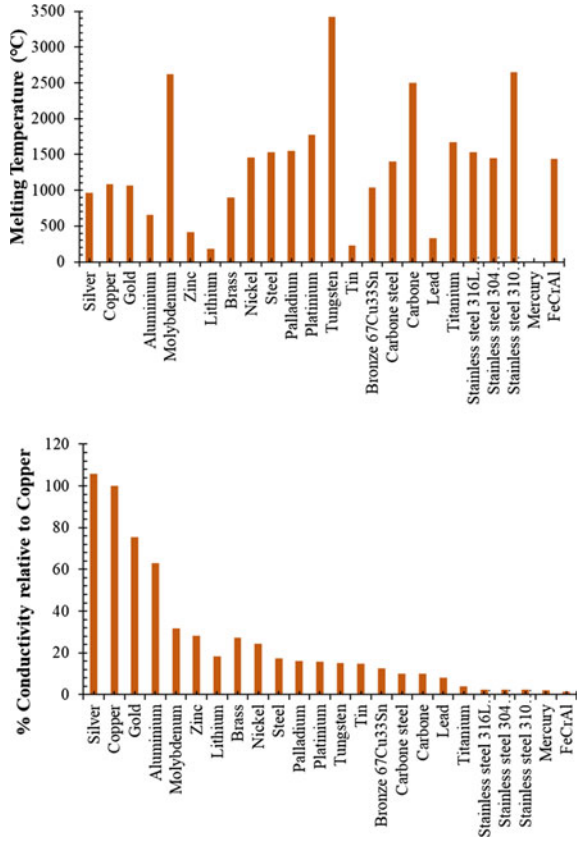
For instance, if platinum (Pt), iron (Fe), and aluminum (Al) are levitated under the same conditions, platinum, [electrical resistivity = 1.06×10^{-7} ($\Omega.m$) at 20°C (Serway 1998)] absorbs more power than either iron [electrical resistivity = 9.7×10^{-8} ($\Omega.m$) at 20°C (Serway 1998)] or aluminum [electrical resistivity = 2.65×10^{-8} ($\Omega.m$) at 20°C (Serway 1998)]. The higher the sample's electrical resistivity, the more efficiently it heats up when submitted to an alternating electromagnetic field and the less power needs to be provided by the heating coils to achieve sample melting.

For purposes of scientific research, a low power-consumption furnace is more desirable because excessive power input can lead to unreliable measurements of various thermophysical properties. A typical example of this is when excessive power input induces instability in the liquid metal meniscus⁶ during melting and/or solidification, leading to distorted surface tension measurements (Pericleous et al. 2006; Wunderlich et al. 2018). In addition, high temperatures inherent to high input power may cause severe evaporation of the alloy (e.g., Al at $2,327^\circ\text{C}$).

Power input should still be sufficiently high to allow for induction melting and/or superheating of the sample. Since the required electrical power input is a function of the electrical conductivity of each metal (inverse of electrical resistivity), a careful choice of the sample composition is necessary. Figure 7.7a and b show the melting points and electrical conductivities of common metals and alloys.

⁶ The interface between the liquid metal and the solid metal in the sample during melting and solidification.

Fig. 7.7 Overview of the melting points **a** and electrical conductivities relative to copper **b** of common metals and alloys



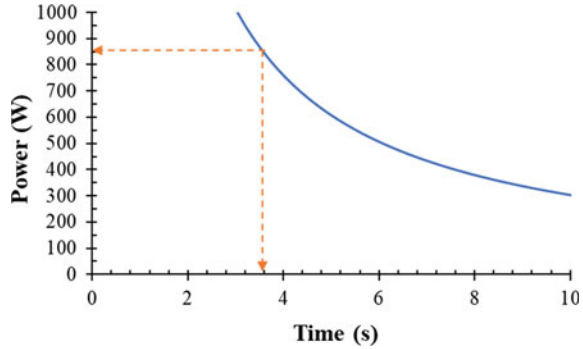
7.2.1.4 Need for Sorting

Due to the range of melting points and conductivity of different metals and alloys, a space debris recycling process using EML technology to melt and reprocess metals, sorting the space debris according to its composition, may be necessary.

7.2.1.5 Time and Power Requirements

A case study of electromagnetic melting in microgravity analyzes the parabolic flights of the TEMPUS experiment in the 1990s. In those experiments, the TEMPUS EML furnace was exposed to a microgravity environment for periods up to 20 s. This proved to be more than enough to conduct heating experiments on small spherical 4.6 g Fe₇₅-Ni₂₅ samples. The TEMPUS furnace managed to heat these samples up to their melting point of 1,580 °C within 3–4 s while using only 860 W (Egry et al. 1992).

Fig. 7.8 Power needed to heat 4.6 g of Fe₇₅-Ni₂₅ to a temperature of 1580 °C, as a function of time. The, etc.



Combining Eq. (7.1) with the following general⁷ expression for the required power P :

$$P \cdot t = m \cdot C_p \cdot \Delta T \quad (7.7)$$

gives us an expression $P_{\Delta T}(t)$ for the power P required to achieve a heating of ΔT in function of the heating time t (s), the sample mass m (kg), the specific heat capacity, C_p (J/kgK), and the desired temperature difference ΔT (K). Figure 7.8. plots this expression $P_{\Delta T}(t)$ for 4.6 g Fe₇₅-Ni₂₅ solid samples heated to a temperature of 1,580 °C, just below the melting point of the alloy.

Using a similar approach, one can plot curves for other sample compositions, sample masses, desired heating range, etc.

To understand power requirements and heating times in a typical EML furnace aimed at space debris recycling, consider an EML system processing three different metals, aluminum, iron⁸ and platinum, in batches of 100 g⁹ and heating them up to a temperature of 2,000 °C. Figure 7.9 plots the $P_{\Delta T}$ curves corresponding to these three samples:

Note that all three sample go through a phase transition in this temperature range.¹⁰ Since our equation for $P_{\Delta T}$ does not account for latent heat requirements during melting, these curves should be considered qualitative rather than quantitative.

Based on this approximation, heating 100 g samples of Al, Fe and Pt to 2,000 °C in an EML furnace, using an electrical power input of 1 kW, is estimated to take ~200 s, ~100 s, and ~30 s respectively. As such, heating power requirements should not be a problem. 1 kW of available power is easily reached with solar panels and a heating process of a couple of minutes allows for high-throughput operation.

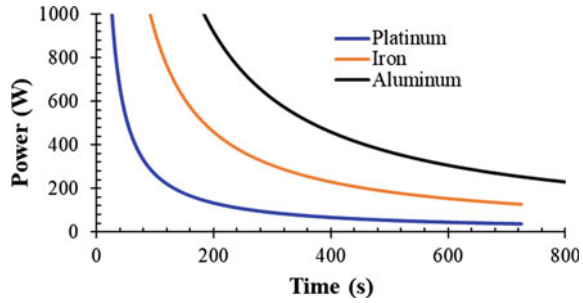
⁷ General as long as the sample does not go through a phase transition during heating.

⁸ Iron is ferromagnetic in its solid state, but not in its liquid state. Levitation melting of a ferromagnetic sample is possible using high enough EM field frequencies, so it is analyzed here as well.

⁹ Space debris is cut up into small pieces before EML processing, their size thus determines the batch size.

¹⁰ Melting points: aluminum = 660.3 °C; iron = 1538 °C; and platinum = 1768 °C.

Fig. 7.9 Approximate power requirements to heat 100 g of **a** Al, **b** Fe and **c** Pt, up to 2,000 °C (2,273 K) as a function of time



7.2.1.6 Radiative Cooling and Solidification

In the previous sections, the absence of gravity was shown to be beneficial to the EML process. It allows for a near-complete decoupling of positioning and heating, reduces positioning power requirements to negligible levels, allows for a stable equilibrium rather than a metastable one, etc. But when it comes to the cooling and solidification of a levitated sample, the picture becomes more mixed.

The absence of convection effects in microgravity EML means that heat transfer from the sample to the environment is driven mainly by radiation. When a molten sample radiates energy to its cooler surroundings, the net radiation heat loss rate can be expressed as follows:

$$q = \epsilon \sigma (T_h^4 - T_c^4) A \tag{7.8}$$

where:

ϵ = Surface emissivity of the material

T_h = Sample body temperature (K)

T_c = Cold surroundings temperature (K)

A = Sample external surface area (m²)

σ = Stefan–Boltzmann constant = $5,6703 \cdot 10^{-8} (W/m^2K)$

Since radiative heat loss in a hard vacuum depends on the temperature of the sample only (since T_c cannot be manipulated), the cooling rate of the sample cannot be easily controlled.¹¹ This poses a problem when one wants to obtain metastable microstructures like martensitic steel, which is typically obtained through fast cooling of an already solidified sample.

The MSL-EML experiment on the ISS solves this problem by using an inert atmosphere in the EML furnace to carry heat away at faster rates than solely radiative

¹¹ One can decrease the radiative cooling rate by irradiating the sample, but there is no way to increase the radiative cooling rate. This difficulty of attaining high cooling rates poses challenges towards obtaining the desired microstructure.

cooling allows for. However, operating a space debris recycling EML furnace in hard vacuum is highly desirable,¹² so other solutions are needed. Research on this topic is still underway, but one possible approach is highlighted here: liquid undercooling and nucleation needles.

7.2.1.7 Liquid Undercooling and Nucleation

When a liquid is cooled below its solidification temperature it does not automatically start its phase transition from liquid to solid. Solidification requires a nucleus, a slight perturbation which jump-starts the phase transition and from which the solidification front expands to the whole sample. Without such a nucleus, a liquid can be cooled well below its solidification temperature without experiencing a phase transition to the solid state. The liquid is then in what is called an undercooled liquid state. On Earth, this state is difficult to achieve due to the destabilizing influence of gravity.¹³ In microgravity EML, this undercooled state is much easier to achieve, and significantly larger undercooling ranges can be attained than on Earth. The more undercooled a liquid is, the faster the phase transition will be when the liquid sample is finally perturbed. The very large undercooling range provided by microgravity EML turns out to be sufficient to induce phase transitions that are fast enough to lock the sample's microstructure in a metastable equilibrium (like the microstructure of martensitic steel for instance) (Perepezko 1994). Nucleation in undercooled liquids can be triggered by touching the sample with a trigger needle inducing nucleation at the point of contact (Diefenbach et al. 2020).

The advantage offered by this method is that it does not require fast cooling rates. One can rely solely on radiative cooling to undercool the liquid metal sample to a desired temperature, and then induce nucleation with a trigger needle. Note that depending on the undercooling temperature, different crystallization pathways can occur, leading to different resulting microstructures in the solidified sample. Finally, since the solidification front starts at the induced nucleus before spreading out to the rest of the sample's bulk, the resulting microstructure is also dependent on where the liquid sample is touched with the nucleation needle. This distortion of the resulting microstructure is, however, quite insignificant for the vast majority of applications and can be minimized if necessary by simultaneous symmetrical application of multiple nucleation needles.

7.2.1.8 Estimating Cooling Times for a Space Debris EML Furnace

Applying the solidification concepts from the previous two sections to the typical space debris EML furnace considered at the end of Sect. 7.2.1.2, cooling times can

¹² Greatly simplifies the design, reduces maintenance requirements, reduces total weight, increases throughput.

¹³ Which is why "naturally occurring" undercooled liquids are typically not encountered on Earth.

be estimated for the 100 g samples of Al, Fe, and Pt. For illustration purposes the starting temperature is the melting point of each sample and the end temperature is the average encountered around the ISS. This exercise is meant to be qualitative rather than quantitative, as one could start cooling from higher temperatures (which will typically be the case) or induce nucleation at a higher temperature than the average surrounding temperature.

A simplified model of the radiative cooling time of a hot spherical metal is given below:

$$t_{cooling} = \frac{Nk}{2\epsilon\sigma A} \left[\frac{1}{T_c^3} - \frac{1}{T_h^3} \right] \tag{7.9}$$

where:

$$k = \text{Boltzmann constant} = 1.38 * 10^{-23} \left(\frac{kg*m^2}{K*s^2} \right)$$

$$N = \text{number of particles contained in the sample} = \frac{m*N_A}{M}$$

m = mass of the sample(kg)

$N_A = 6.022 * 10^{23}$ = Avogadro's number

M = Molar mass of the sample(kg/mol)

ϵ = emissivity coefficient; $0 < \epsilon < 1$

(depends on material and its surface temperature)

The model is approximative as it assumes that the entire metal is at the same temperature, whereas in reality the surface will cool faster. The higher the sample's thermal conductivity, the more accurate this model becomes. A further simplification is to use a constant emissivity coefficient ϵ for each sample, although in reality this factor varies with temperature. Using the following constant values for the emissivity coefficients and using the following parameters one can calculate the required cooling times by using Eq. (7.9) (Tables 7.1, 7.2 and 7.3).

As noted in the results, the estimated radiative cooling rates are relatively slow, but not excessively so. Given that it is preferable for an EML system aimed at space debris recycling to operate in a hard vacuum, extra cooling with an inert atmosphere to increase throughput is not practical. Taking these considerations into account, processed materials could be temporarily stored in a separate chamber for cooling, allowing the EML furnace to process feedstock at a higher rate.

Table 7.1 Constant emissivity coefficients used in the qualitative exercise

Considered material	Considered temperature (°C)	Emissivity coefficient
Aluminum, unoxidized	500	~0.06
Iron, rough ingot	60	~0.03
Platinum, polished plate	60	~0.10

Table 7.2 Parameters used in the qualitative exercise

Parameters	Aluminum	Iron	Platinum
Sample mass (kg)	0.1	0.1	0.1
Average T outside ISS (K)	233	233	233
Temperature of the sample, T_h (K)	933	1,811	2,041
Metal density (kg/m^3)	2,700	7,800	21,460
Specific heat, C_p ($\text{J}/\text{kg}\cdot\text{K}$)	0.9	0.45	0.13
Molar mass (kg/mol)	$27 * 10^{-3}$	$56 * 10^{-3}$	$195 * 10^{-3}$

Table 7.3 Results of the qualitative exercise—cooling time and cooling rate

Samples (100 g)	Cooling time (minutes)	Equivalent cooling rate (K/s)
Aluminum	22.59	0.52
Iron	15.81	1.66
Platinum	6.32	4.76

7.2.1.9 Electromagnetic Casting

Electromagnetic casting is the concept of using electromagnetic fields (EM shaping fields) to force liquid metal into a desired shape before solidification. EM shaping fields are similar to positioning fields, since the aim is to exert force on the sample surface without contributing (significantly) to heating of the sample. But while positioning fields operating in microgravity can keep the sample at the center of the coil system with minimal force, shaping fields must be stronger to overcome surface tension of the melt and deform the liquid sample. In practice, the need for stronger fields will reduce the ability to fully decouple heating from shaping. Research is underway to determine the level of heating associated with a shaping field in microgravity.

Terrestrial EM casting methods

Two studies on the topic of EM casting are of particular interest as they demonstrate the feasibility of the concept on Earth with potentially more difficult conditions than a similar EML process would encounter in space. Spitans et al. (2017) demonstrated the feasibility of the EML concept on Earth for samples up to 500 g, while Yang et al. (2018) demonstrated the feasibility of using electromagnetic dies and molds to shape a levitated molten metal sample.

- *Spitans et al.* The electromagnetic casting technique proved to be an effective containerless metal production technique. Using this technique, Spitans et al. successfully produced large-scale Al and Ti alloys (Spitans et al. 2017). In their work, a furnace with horizontal and orthogonal EM fields of different frequencies was designed to levitate 500 g of molten Al and Ti-6Al-4 V. The levitation experiments showed “that the two-frequency horizontal EM field configuration,

compared to a conventional levitation method, can be used for the levitation melting of aluminum samples with increased mass” (Spitans et al. 2017) (>100 g) compared with the levitation melting in a conventional inductor. “The new Fast-Cast concept that utilizes the novel crucibleless levitation melting method for industrial mass production of single-shot castings” was then introduced (Spitans et al. 2017).

- *Yang et al.* In a counter-gravity configuration, Yang et al. successfully produced TiAl castings. In their work, the TiAl production was carried out in a high-vacuum chamber with inert gas protection. Their technique consisted in lifting the melt, under pressure, through a tube and making it flow into the mold. Similarly, an in-space extrusion could be achieved by using electromagnetic forces to carry the melt through a guiding electromagnetic die.

Based on the electromagnetic casting concept (Evans 1995; Kim et al. 2001; Spitans et al. 2017; Vives and Ricou 1985; Yang et al. 2018), a multi-purpose electromagnetic levitation furnace could potentially be set up for both contactless melting and extrusion, through an electromagnetic die, to recycle space debris into usable shaped structures in space.

7.2.2 In-Space Metal Manufacturing Enabled by Space Debris Recycling

7.2.2.1 Metal propellant

A number of companies are developing viable electric propulsion systems that utilize metal as a propellant. Space tugs, satellites, and other spacecraft fitted with these systems could be refueled with metal recycled from space debris.

A promising example of these systems is being developed by Neumann Space. The Neumann Thruster uses a “patented Centre-Triggered Pulsed Cathodic Arc Thruster (CT-PCAT) technology [that] converts a solid conductive fuel rod into plasma and produces thrust. The system can use a range of conductive fuels.” (NeumannSpace.com 2020).

7.2.2.2 Wire for 3D printing

Additive manufacturing (AM) is an enabling technology for in-orbit manufacturing of replacement parts and tools, which could reduce existing logistics requirements for the ISS and future sustainability of human space missions (Werkheiser 2017). Additive manufacturing has the potential to redefine space architecture. It can “potentially lead to the construction of smaller, more reliable, less massive satellite systems or their key components (including support structure, power distribution system,

solar arrays, instruments, outer protective shell, etc.), which could reduce launch requirements and costs” (National Academy of Sciences 2014). Space debris could be reprocessed into metal wire filament for in-space additive manufacturing.

7.3 Policy and Law

Salvaging and recycling space debris for reprocessing and/or reuse presents unique legal questions. Legal ownership is often transferred between commercial entities for active space assets, while transfer of state liability is rare. There have also been a few precedent-setting examples of active recovery or servicing under a commercial contract. In 1984, with the Space Shuttle, NASA executed the first ever in-space salvage operation and recovered two satellites, Westar VI and Palapa B-2 (Fisher et al. 2013). In 2020, Space Logistics docked its Mission Extension Vehicle-1 to the geostationary satellite Intelsat IS-901 to extend its useful lifespan by supplying additional propulsion and control (Space Logistics 2020).

While satellite servicing is becoming a commercial reality, no well-defined legal framework exists for the active recovery or salvage of uncontrolled derelict space objects (space debris) for repair, removal, or recycling. Nevertheless, this topic has been covered by numerous studies and articles since the early days of space exploration (Hofman et al. 2017; Muñoz-Patchen 2018; Hall 1967; Haley and Hannover 1959). This section provides an overview of the legal and policy issues that must be addressed to utilize space debris as a resource.

7.3.1 Liability

Pursuant to Article VI of the Outer Space Treaty, states bear the burden of international responsibility for their national activities in outer space, including activities conducted by nongovernmental entities. According to Article II of the Liability Convention, the launching state of a spacecraft is absolutely liable if a space object causes damage on the surface of the Earth, or to an aircraft in flight. However, Article III indicates that liability for damage caused anywhere else, e.g., on orbit, is based on negligence. Thus, the launching state will be liable only if the damage is due to its fault, or the fault of the entity for which it is responsible. Importantly, the Convention creates a broad definition of launching state, to include: (1) the state that launches or procures the launch of a space object; and (2) the state from whose territory or facility a space object is launched (Hermida 2004).

Although there have been at least two high-profile cases where a space object has caused damage, the Liability Convention has never been formally invoked. In January 1987, the Soviet satellite Cosmos 954 disintegrated on re-entry, raining radioactive debris on Canadian territory. Canada presented a claim pursuant to the Convention

but accepted a payment outside the Convention of approximately C\$3,000,000 as settlement—far less than actual damage costs. In February 2009, an inactive Russian satellite, *Cosmos-2251*, collided with the active commercial communications satellite, *Iridium 33*, which was operated by a US-based company. This incident, too, was settled by the respective states outside the Convention (Jakhu 2010).

According to the *Cosmos-2251* precedent, the launching state of the space recycling facility could be liable for any damage caused by its launch and operation. Additionally, under current law, a launching state may never revoke its status as a launching state. This means that the launching state and the original launching state of a targeted space object will likely both be liable for any potential damage caused once under the control of the recycling entity. Until a new legal structure is enacted, the launching state will have to negotiate agreements for this shared liability on a bilateral basis with each launching state responsible for objects designated for recycling. Once the original object has been reprocessed, it should be treated the same as if it had deorbited and no longer exists. For new objects that are manufactured in space, it could be argued that the launching state should be that of the customer causing that part to be made, but this is also an untested concept.

To limit the risk to the launching state, the entity intending to recycle space debris could sign an agreement to accept liability for any damage caused by spacecraft developed by the recycling entity due to negligence or any space objects for which the recycling entity has obtained ownership. This agreement would likely need to be underwritten by an insurance company (Wang 2016).

7.3.2 Ownership

The transfer of commercial ownership of space objects follows an established process but also may be partially limited due to the liability of the original launching state. Currently, this is done by entering into an agreement with another private company. However, state liability and launching state status do not change without the state's approval—and even if both states agree, whether a launching state can ever give up its responsibility has not yet been tested. This is something that two states would have to agree to bilaterally. Additionally, even if the object is non-functional, both states and commercial owners may have concerns that any technology used to acquire a space object could be viewed as a surveillance or military activity and could endanger sensitive intellectual property, potentially blocking a transfer of ownership. Absent a new legal regime, a possible approach is to negotiate a bilateral agreement for the transfer of ownership with the owner/operator of the space objects targeted for salvage and recycling, subject to the approval and written confirmation of both the original launching state for the object and the launching state for the company.

7.3.3 Policy

With the rapid expansion of commercial space activities, many have argued that existing space policies have become outdated and need to be modernized. The USA, Luxembourg, UAE, and Japan have passed national legislation to clarify the legal rights and regulations regarding the extraction and utilization of space resources and other countries are working to do the same (Hofmann and Bergamasco 2020, Foust 2021). Supranational organizations like the UN and the Hague Space Resources Governance Working Group are working to develop agreed policies at an international level. However, at the time of writing, no legal entity or jurisdiction has directly addressed the salvage of derelict spacecraft for reuse or recycling.

7.4 Conclusions

With Earth orbits becoming increasingly crowded and more contested, there is broad agreement that more must be done to address the challenges posed by existing and future space debris. Although not yet clearly defined, legal approaches do exist to conduct commercially contracted salvage and recycling operations. By passing enabling legislation at a national or supranational level, policy makers could accelerate the adoption of in-space salvage and recycling as an important component in the effort to maintain space sustainability.

7.5 Summary and Conclusions

An in-space industrial economy is emerging in Earth orbit and cislunar space and the supply of material to support this ecosystem will be a crucial factor. Space debris is a source of risk to the ecosystem but also creates an opportunity as a near-term resource which can be salvaged and recycled in orbit. Current proposed techniques to address space debris focus solely on the safe deorbit into Earth's atmosphere or boosting objects into an unused graveyard orbit. On the other hand, an orbital salvage and recycling platform could process space debris into useful components and feedstock material to support a self-sustaining, industrialized, in-space economy.

The technical process to recycle space debris starts with taking control of the object, then analyzing the object's structure and material composition, salvaging components which can be reused, and finally processing the remaining material by melting and reshaping. The melting of metals in microgravity has been studied on Earth and in space. The Electromagnetic Levitator (EML) on the ISS has enabled containerless processing of metals and alloys and the study of melting and solidification in microgravity since 2015. Containerless processing using EML technology offers a promising basis for a scaled-up in-space metal processing capability central to the concept of space debris recycling.

Electromagnetic levitation is achieved by using alternating-current coils to generate high-frequency electromagnetic fields surrounding a sample. These alternating fields induce eddy currents in the sample, which heat the sample and generate opposing forces on its external surface. In microgravity, electromagnetic fields used for heating can be decoupled from electromagnetic fields used for positioning, allowing for precise control of melt temperature and the achievement of undercooled states not possible on Earth. In this undercooled state a needle can be used to trigger nucleation and rapid solidification of a sample at the desired temperature. These aspects unique to the use of EML systems in persistent microgravity have proven useful to materials science research on the ISS.

Because the EML on the ISS uses a simple coil arrangement, metal samples naturally form a spherical shape. To achieve other desirable shapes, more complex coil arrangements are required. Using electromagnetic fields to shape molten metals, electromagnetic casting, has been studied on Earth and is a promising method to manufacture desired geometric shapes in microgravity. One such shape is a metal cylinder which could be used as propellant in pulsed plasma thrusters, such as the Newman Drive. Other desired shapes are metal wires which could be used for additive manufacturing in space.

In addition to the technological challenges of processing metals in microgravity, the legal aspects must be considered. The Liability Convention does not clearly regulate the transmission of ownership of space debris. Until a clear legal structure is established, launching states will have to negotiate the liability question on a bilateral basis. On the other hand, the transfer of commercial satellite ownership follows an established process and precedents have occurred for satellite recovery and satellite servicing on a contract basis. While a potential legal path exists for space debris salvage and recycling, positive action on the part of policy makers could accelerate its adoption and benefits to the cause of space debris remediation.

In conclusion, by addressing the technical and legal challenges of space debris salvage and recycling, derelict space objects can be removed as a hazard and transformed into a resource to support a self-sustaining industrial economy in orbit.

Acknowledgements This chapter has evolved from the research conducted by the authors as they have worked to make space debris salvage and recycling a reality through their company, CisLunar Industries.

The authors would like to thank Dr. Andrew Petruska, Colorado School of Mines, for sharing his knowledge of the practical application of electromagnetic fields and for reviewing this chapter. Thanks also go to Dr. Gonzalo Garcia-Atance Fatjo, University of Central Lancashire, for his review of this chapter.

For three years of mentorship and generous knowledge sharing on space resources and containerless processing, much gratitude goes to Dr Angel Abbud-Madrid and Dr Chris Dreyer, Colorado School of Mines. The authors also express their gratitude to Michelle Hanlon for her contribution to the authors' understanding of space law in the context of space debris recycling. For advancing the field of pulsed plasma thrusters and promoting the use of space debris as a potential propellant resource, the authors thank Dr Paddy Neumann and Neumann Space. The authors also express their thanks to Gary Martin and Josh Izenberg for their continued mentorship and support from the early days of CisLunar Industries. To the International Space University, thank you for inspiring this endeavor and connecting so many incredible people in the space community.

Most importantly, the authors are extremely grateful to their families, friends and colleagues who have steadfastly supported their efforts to build CisLunar Industries and make space debris recycling a reality. This wouldn't have been possible without you!

References

- Aerospace Corporation. (2018). Space debris and space traffic management. <https://aerospace.org/article/space-debris-and-space-traffic-management>. Accessed 15 Nov 2020.
- Airbus Bartolomeo User Guide. (2018). Airbus Bartolomeo user guide, Issue 1.
- Anselmo, L., and C. Pardini. 2016. Ranking upper stages in low earth orbit for active removal. *Acta Astronautica* 122: 19–27.
- Blackerby, C., Okamoto, A., Fujimoto, K., Okada, N., Forshaw, J.L., Auburn, J. (2018). ELSA-d: an in-orbit end-of-life demonstration mission. In: *69th international astronomical congress*.
- Brisley, W., and B.S. Thornton. 1963. Electromagnetic levitation calculations for axially symmetric systems. *British Journal of Applied Physics* 14: 682–686.
- Corbett, J. (2015). Micrometeoroids and orbital debris (MMOD). Retrieved November 16, 2020, from https://www.nasa.gov/centers/wstf/site_tour/remote_hypervelocity_test_laboratory/micrometeoroid_and_orbital_debris.html. Accessed 16 Nov 2020.
- Diefenbach, A., Schneider, S., Volkman, T. (2020). Experiment preparation and performance for the electromagnetic levitator (EML) onboard the international space station. In: *Preparation of space experiments*. IntechOpen.
- Egry, I., G. Lohofer, P. Neuhaus, and S. Sauerland. 1992. Surface tension measurements of liquid metals using levitation, microgravity and image processing. *International Journal of Thermophysics* 13: 65–74.
- European Space Agency. (2020). Space debris by the numbers. Retrieved November 15, 2020, from https://www.esa.int/Safety_Security/Space_Debris/Space_debris_by_the_numbers. Accessed 15 Nov 2020.
- Evans, J.W. 1995. The use of electromagnetic casting for al alloys and other metals. *JOM Journal of the Minerals Metals and Materials Society* 47 (5): 38–41.
- Fisher, J. (2013). Retrievals of Westar VI, Palapa B-2: An epic adventure from a special commemorative issue of SCG Journal, December 1984. <https://www.hughesscgheritage.com/retrievals-of-westar-vi-palapa-b-2-an-epic-adventure-from-a-special-commemorative-issue-of-scg-journal-december-1984/>. Accessed Nov 17 2020.
- Forshaw, J.L., G.S. Aglietti, N. Navarathinam, H. Kadhem, T. Salmon, et al. 2016. RemoveDEBRIS: An in-orbit active debris removal demonstration mission. *Acta Astronautica* 127: 448–463.
- Foust, J. (2021). Japan passes space resources law. SpaceNews. <https://spacenews.com/japan-passes-space-resources-law/>. Accessed 24 Sep 2021.
- Fromm, E., and H. Jehn. 1965. Electromagnetic forces and power absorption in levitation melting. *British Journal of Applied Physics* 16: 653–663.
- Haley, A.G., and W.H.P. Hannover. 1959. Space age presents immediate legal problems. In *First Colloquium on the law of outer space The Hague 1958 Proceedings*, 5–27. Springer-Verlag Wien.
- Hall, R.C. 1967. Comments on salvage and removal of man-made objects from outer space. *Journal of Air Law and Commerce* 33 (2): 288.
- Herlach, D.M. (1991). Containerless undercooling and solidification of pure metals. *Annual Review of Materials Science* 23–44.
- Hermida, J. 2004. *International space law*. Legal basis for a national space legislation: Springer.
- Hofmann, M., Loukakis, A. (2017). *Ownership of satellites*. Luxembourg Workshop on Space and Satellite Communication Law.
- Hofmann, M., Bergamasco, F. (2020). Space resources activities from the perspective of sustainability: legal aspects. In *Global sustainability*, vol. 4. Cambridge University Press.

- Hofmann, D.C., and S.N. Roberts. 2015. Microgravity metal processing: From undercooled liquid to bulk metallic glasses. *Microgravity* 1: 15003.
- Jakhu, R.S. (2010). Iridium-cosmos collision and its implications for space operations. In: *Yearbook on space policy 2008/2009*, pp. 254–275. Vienna, Springer.
- Jakhu, R., Nyampong, Y.O.M., Sgobba, T. (2017). Regulatory framework and organization for space debris removal and on orbit servicing of satellites. *Journal of Space Safety Engineering*, 129–137.
- Kim, S.-W., H. Hao, U.-J. Lee, K.-D. Woo, and J.-Z. Jin. 2001. Microstructural characteristics and wear resistance of electromagnetic casting aluminum alloys. *Materials Transactions* 42 (9): 1952–1958.
- Liou, J.-C., N.L. Johnson, and N.M. Hill. 2010. Controlling the growth of future LEO debris populations with active debris removal. *Acta Astronautica* 66: 648–653.
- Lohöfer, G. 2018. High-resolution inductive measurement of electrical resistivity and density of electromagnetically levitated liquid metal droplets. *Review of Scientific Instruments* 89 (12): 124709.
- Made In Space. (2018). Additive manufacturing facility. <http://madeinspace.us/projects/amf>. Accessed 31 May 2018.
- Muck, O. (1923). German Patent No. 422004.
- Muñoz-Patchen, C. 2018. Regulating the space commons: Treating space debris as abandoned property in violation of the outer space treaty. *Chicago Journal of International Law* 19: 233.
- Nan, W., X. Wen-Ju, and W. Bing-Bo. 1999. Physical characteristics of electromagnetic levitation processing. *Acta Physica Sinica* 8: 503–513.
- National Academy of Sciences. (2014). Summary. In *3D printing in space*, 2. Washington, DC: National Academies Press.
- NeumannSpace.com. (2020). <https://neumannspace.com/>. Accessed 16 Nov 2020.
- Okress, E.C., D.M. Wroughton, G. Comenetz, P.H. Brace, and J.C.R. Kelly. 1952. Electromagnetic levitation of solid and molten metals. *Journal of Applied Physics* 23: 545–552.
- Perepezko, J.H. 1994. Solidification reactions in undercooled alloys. *Materials Science and Engineering: A* 179: 52–56.
- Pericleous, K., V. Bojarevics, G. Djambazov, R.A. Harding, and M. Wickins. 2006. Experimental and numerical study of the cold crucible melting process. *Applied Mathematical Modelling* 30: 1262–1280.
- Ratke, L., Diefenbach, S. (1995). Liquid immiscible alloys. *Materials Science and Engineering R: Reports* 15(7–8):263–347.
- Richard, M., Kronig, L.G., Belloni, F., Gass, V., Araromi, O.A., Shea, H., Paccolat, C., Thiran, J-P. (2013). Uncooperative rendezvous and docking for MicroSats. In: 6th Int.Conf. on Recent Adv. in Space Tech., RAST.
- Rony, P.R. (1965). The electromagnetic levitation of metals. In *Trans. Vacuum Met. Conference*, ed. M. A. Cocca, 55–135. Boston, MA, Vacuum Society.
- Sacco, E., Moon, S.K. (2019). Additive manufacturing for space: Status and promises. *International Journal of Advanced Manufacturing Technology* 105:4123–4146
- Serway, R.A. (1998). *Principles of physics*, 2nd edn. Fort Worth, TX; London: Saunders College Pub. p. 602. ISBN 978–0–03–020457–9.
- Space Logistics. (2020). SpaceLogistics. <https://www.northropgrumman.com/space/space-logistics-services/>. Accessed 17 Nov 2020.
- Spitans, S., H. Franz, E. Baake, and A. Jakovičs. 2017. Large-scale levitation melting and casting of Titanium alloys. *Magnetohydrodynamics* 53 (4): 633–641.
- Tonetti, S., S. Cornara, M. Faenza, O. Verberne, T. Langener, and G. Vicario de Miguel. 2018. Active debris removal and space debris mitigation using hybrid propulsion solutions. *Astrophysics and Space Science Proceedings Book Series* 52: 163–180.
- Torres, A., Ganley, J., Maji, A., Tucker, D., Starodubov, D. (2013). Enhanced processability of ZrF4-BaF2-LaF3-AlF3-NaF glass in microgravity. In *Proc. SPIE 8704, Infrared Technology and Applications*, XXXIX, 87042C.

- Viquerat, A., Schenk, M., Lappas, V., Sanders, B. (2015). Functional and qualification testing of the inflatesail technology demonstrator. In *2nd AIAA Spacecraft Stru. Conf.*, 1627.
- Vives, C., and R. Ricou. 1985. Experimental study of continuous electromagnetic casting of aluminum alloys. *Metallurgical Transactions B* 16 (2): 377–384.
- Wang, T. 2016. A liability and insurance regime for space debris mitigation. *Science & Global Security* 24 (1): 22–36.
- Werkheiser, N. (2017). NASA additive manufacturing overview. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170001551.pdf>.
- Wunderlich, R.K., U. Hecht, F. Hediger, and H.-J. Fecht. 2018. Surface tension, viscosity, and selected thermophysical properties of Ti48Al48Nb2Cr2, Ti46Al46Nb8, and Ti46Al46Ta8 from microgravity experiments. *Advanced Engineering Materials* 20: 1800346.
- Yang, J., H. Wang, Y. Wu, X. Wang, and R. Hu. 2018. A combined electromagnetic levitation melting, counter-gravity casting, and mold preheating furnace for producing TiAl Alloy. *Advanced Engineering Materials* 20: 1700526.

Part II
Mercury and Venus

Chapter 8

Planetary Exploration of Mercury With BepiColombo and Prospects of Studying Venus During Its Cruise Phase



Johannes Benkhoff and Joe Zender

Abstract Mercury and Venus are key planets for the understanding of the evolution of our solar system and therefore also for the question of how the Earth and life formed. In the case of Mercury, it is mainly because of its position among the planets. It is the planet closest to the Sun and experiences the harsh environment of the Sun the most. Venus is more Earth-like in respect of its size and the existence of an atmosphere which could indicate the possibility of having been habitable in the past. Bepi-Colombo is a planetary mission devoted to the thorough exploration of Mercury and its environment. The mission will be carried out as a joint project between ESA and JAXA. The mission consists of two spacecraft, the Mercury Planetary Orbiter (MPO) and the Mercury Magnetospheric Orbiter (MIO), carrying a comprehensive suite of instruments which will carry out scientific measurements. BepiColombo, launched on 20 October 2018 from the European spaceport in Kourou, French Guiana, is now en route to Mercury. Its route requires a 7.2-year-long cruise phase, with one Earth flyby, two Venus flybys and six Mercury flybys before arriving at Mercury at the end of 2025. After a weak gravity capture maneuver, both spacecraft will be placed into their dedicated polar orbits of 590 x 11,640 km (MIO) and 480 x 1500 km (MPO), respectively. As part of the cruise phase, BepiColombo will also fly past Venus twice. The two flybys at Venus will give scientists good opportunities to study the atmosphere and ionosphere of Venus in more detail.

8.1 Introduction

Mercury and Venus are key planets for the understanding of the history of our solar system and therefore also for the question of how the Earth and life were formed. In the case of Mercury, it is mainly because of its position among the planets. It is the planet closest to the Sun and experiences the harsh environment of the Sun the most. Venus is more Earth-like in respect of its size and the existence of an atmosphere which could indicate the possibility of having been habitable in the past.

J. Benkhoff (✉) · J. Zender
ESA-ESTEC, Science and Operations Department, Noordwijk, The Netherlands
e-mail: Johannes.Benkhoff@esa.int; jj@benkhoff.eu

© Springer Nature Switzerland AG 2023
V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_8

337

The view of Mercury has changed over recent decades. More than 30 years ago Mercury was seen as not very spectacular, and there was no case for speculations about life or active surface processes such as volcanism. That view changed dramatically after NASA's highly successful MESSENGER mission (McNutt et al. 2004, 2006; Solomon and Anderson 2018). MESSENGER orbited Mercury for about four years and some of the results were rather unexpected (Solomon et al. 2018), causing Mercury to become a planet of mysteries for many scientists—and to remain so today.

Some of MESSENGER's results did not fit in with our current thinking of how Mercury had evolved, or what we might expect to find on a planet close to the Sun. Mercury's Earth-like magnetic field had already been discovered in 1974–1975 when Mariner 10, another NASA mission, performed several flybys (Storm and Sprague 2003). This unexpected discovery was eventually confirmed by MESSENGER. Although the dipole field of Mercury is 100 times weaker than that of Earth, it requires a fluid outer core layer inside the planet and much hotter temperatures in the interior than was initially assumed for such a small planet. Scientists have sought explanations by, for example, adding elements like sulfur or silicon to the material the core was assumed to be made of. It was thought that these elements could help to keep the material inside the outer core liquid at lower temperatures or a small planet like Mercury after 4.5 billion years of existence in our solar system (Genova et al. 2021). But MESSENGER also found that the dipole field of Mercury seems to be shifted about 400 km to the north by instead of being centered along the planet's rotational axis. Another surprising result is that measurements proved that Mercury contains more volatiles than expected on its surface, indicating lower formation temperatures. Are our existing models of the solar system's formation still valid, or was Mercury formed somewhere else in the solar system? What are the implications of these findings for planetary formation in general, considering also the many exoplanets found in similar orbits?

The diurnal temperature variations on Mercury's surface are huge. Since there is no atmosphere, the temperature can be as cold as minus 170 °C but also as high as 450 °C. It seemed unlikely that water or life could exist under these conditions. However, some water exists on Mercury's surface, albeit hidden in permanently shadowed near-polar craters, since Mercury's rotation axis is almost not tilted (Harmon and Slade 1992; Harmon et al. 1994; Chabot et al. 2016, 2018). Images from the surface obtained from MESSENGER revealed a structure called hollows (Blewett et al. 2011, 2016) not seen on any other planet before. Hollows are bright deposits on the surface found in and around some craters. These features appear young and seem to have been formed by escaping gas. A big question is whether these processes stopped a few thousands or millions of years ago or are still active today (Rothery et al. 2020). Will future missions see changes in these structures compared to what was observed by MESSENGER?

Not only do many unanswered questions remain about the planet, but processes in the environment are also very different from what scientists expected. Due to the proximity to the Sun, the magnetosphere, and the lack of atmosphere, the solar wind interacts with Mercury differently compared with the Earth. Reconnection

processes, asymmetries, particles reaching the surface, a magnetosphere that is highly compressed when the Sun is active, as well as manifold other processes that are not fully understood, have been observed and require further investigation (Milillo et al. 2020).

To resolve all or, more realistically, some of these mysteries, further exploration of Mercury and its environment is needed. Here scientists are in a fortunate situation. Although not originally planned for in this way, the upcoming BepiColombo mission (Benkhoff et al. 2021), which will reach Mercury in 2025, is particularly well suited to solving Mercury's mysteries and providing another view of the planet about 10 years after MESSENGER, including potential changes in its surface and its magnetic field configuration.

8.2 The BepiColombo Mission to Mercury

BepiColombo is a planetary mission of the European Space Agency's (ESA) Cosmic Vision Programme. The mission is devoted to the thorough exploration of Mercury and its environment with the aim of understanding the process of planetary formation and evolution in the hottest part of the proto-planetary nebula, as well as to understanding similarities and differences between the magnetospheres of Mercury and Earth. The mission is a joint project between ESA and JAXA (Japanese Aerospace Exploration Agency) (Benkhoff et al. 2021).

The BepiColombo mission consists of two spacecraft: the Mercury Planetary Orbiter (MPO) and the Mercury Magnetospheric Orbiter (MMO, nicknamed Mio). The orbiters were launched together on an Ariane V rocket from ESA's spaceport in Kourou, French Guiana on 20 October 2018 (Fig. 8.1). ESA is responsible for MPO and JAXA is responsible for the Mio spacecraft. ESA also provided the transfer spacecraft (Mercury Transfer Module, MTM), which is equipped with a solar electric propulsion engine for the transport of the two spacecraft to Mercury. Upon arrival at Mercury the MTM will be jettisoned shortly before the insertion of the spacecraft into their dedicated orbits in late 2025. During cruise the spacecraft are connected in a stacked configuration. The lowest module is the MTM. The MPO is mounted on top of the MTM and Mio is hidden behind a sunshield needed to protect the spacecraft from overheating during the cruise phase to Mercury (see Fig. 8.2).

8.2.1 Science Goals

The scientific objectives behind BepiColombo can be summarized in eleven questions:

Does Mercury hold any clues about the composition of the early solar nebula and the formation of the planetary system?



Fig. 8.1 Launch of BepiColombo in October 2018 with an Ariane V rocket



Fig. 8.2 BepiColombo spacecraft in cruise configuration. The Mercury Transfer Module with its two large solar wings carries the Mercury Planetary Orbiter (middle with one solar wing) and the Mercury Magnetospheric Orbiter, nicknamed Mio, hidden inside the solar shield

Why is Mercury’s density significantly higher than that of all other terrestrial planets, the Moon included?

Is Mercury’s core liquid or solid?

Is Mercury tectonically active today?

Why does such a small planet possess such a strong magnetic field, while Venus, Mars and the Moon do not have any?

Why are we unable to detect any presence of iron through spectroscopic observations, while this element is supposedly the major constituent of Mercury?

Is there any water or sulfur ice in the permanently shadowed craters of Mercury's polar regions?

What are the production mechanisms of the exosphere?

In the absence of any ionosphere, how does the magnetic field interact with the solar wind?

Is Mercury's magnetized environment characterized by features reminiscent of the aurorae, radiation belts and magnetospheric sub-storms observed on Earth?

Since the advance of Mercury's perihelion was explained in terms of space-time curvature, can we take advantage of the proximity of the Sun to test general relativity with improved accuracy?

In order to answer these questions, BepiColombo has a comprehensive suite of instruments on both spacecraft which will perform science measurements. At the end of the mission it is hoped that the following will have been obtained: a very detailed topographical, elemental and mineralogical map of the planet, a good characterization on the interior of the planet and its magnetic field, detailed knowledge on processes going on in the thin atmosphere called exosphere around the planet, understanding of the interaction of the solar wind with Mercury's magnetic field and Mercury itself, more details on the ices buried in permanently shadowed craters near the poles, more details on the hollows discovered by MESSENGER, the surface age and history of volcanism, and much more. All of this will hopefully lead to a better understanding of the evolution of Mercury and our solar system.

8.2.2 *The Spacecraft Modules*

BepiColombo is dedicated to the thorough exploration of Mercury and its environment. The mission design is driven essentially by the scientific payload requirements, the launch mass constraints, and the harsh thermal and radiation environment at Mercury. Key technologies required for the implementation of this challenging mission include high-temperature thermal control materials, a special radiator design for high-infrared environment, high-temperature and high-intensity solar cell development, steerable high-gain and medium-gain antennas and novel payload technologies.

The total mass of all the spacecraft modules is about 4,100 kg. Despite traveling towards the Sun, the transfer module (MTM) requires two large solar arrays, each about 14 m long to provide the necessary power for the Solar Electric Propulsion System (SEP). The MPO solar panel has a length of about 7.5 m. The size of the MPO radiator is about 8 square meters to radiate the excess heat produced inside the spacecraft into space. The thermal design of the spacecraft was a major design driver for the BepiColombo mission (Ferrero et al. 2016).

The Mercury Planetary Orbiter (MPO, Figs. 8.3 and 8.4) is a three-axis stabilized spacecraft, which accommodates eleven scientific instruments and has a box-like shape with a size of $3.9 \times 2.2 \times 1.7$ m, and a dry mass of about 1080 kg. The

tremendous heat load at Mercury imposes strong requirements on the spacecraft design, requiring high-temperature multi-layer insulation and solar array technology. As the radiator has to face away from the Sun the MPO needs to flip by 180° at perihelion and aphelion in its polar orbit in order to avoid direct sunlight. Most of the instruments are mounted on the planet-facing (nadir) side. The altitude range is expected to be 480–1500 km, with the latitude of the closest distance to Mercury (periherm) varying between 16° north and 16° south over the course of the one Earth year-long nominal science phase. This latitude range was chosen to obtain high-resolution spatial coverage at global scales and ensure resolution symmetry at both hemispheres.

Four redundant 22 N thrusters are mounted on the nadir face and are used for orbital maneuvers until the final orbit has been reached. The control of the attitude is provided by a set of four reaction wheels and four 10 N thrusters for momentum wheel desaturation. These thrusters are mounted on the radiator. Three star trackers



Fig. 8.3 Artist's impression of Mercury Planetary Orbiter (MPO) in orbit around Mercury

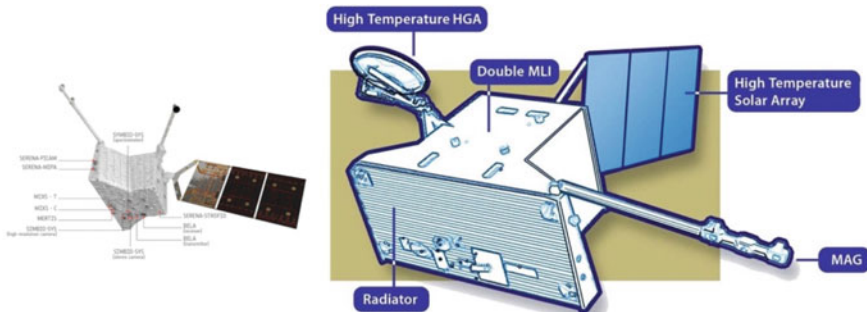


Fig. 8.4 Sketch of the Mercury Planetary Orbiter (MPO) indicating the position of some of the instruments onboard the spacecraft (left) and the location of radiator, high-gain antenna, solar array, and magnetometer boom

are also mounted on the radiator side to provide the precise attitude control required by several of the instruments. The radiator contains highly reflective fins, mounted at an appropriate angle to minimize absorption of heat radiated from Mercury and at the same time to allow radiation towards deep space. Because of the intense heat, the three-panel solar array is a 70–30% mixture of solar cells and optical surface reflectors (OSR, i.e., mirrors) to keep its temperature below 200 °C. During eclipse periods the operations of the instruments and temperature inside the spacecraft are maintained. The power is provided by an internal battery.

The science payload for the Mercury Planetary Orbiter (MPO) contains the following eleven instruments:

- (1) BELA (BepiColombo Laser Altimeter) to characterize the topography and surface morphology of Mercury. BELA will aim at measuring Mercury's tidal deformation and will provide constraints on Mercury's rotation state. It is led by the university of Bern, Switzerland and DLR Berlin, Germany (Thomas et al. 2021).
- (2) ISA (Italian Spring Accelerometer) to study Mercury's interior structure and to test Einstein's theory of relativity. The instrument is led by Institute for Astrophysics and Space Planetology, Italy (Santoli et al. 2020).
- (3) MPO-MAG (Magnetic Field Investigation), an instrument to describe Mercury's magnetic field and its source, and led by Technical University of Brunswick, Germany (Heyner et al. 2021).
- (4) MERTIS (Mercury Radiometer and Thermal Imaging Spectrometer) to study Mercury's mineralogical composition and provide global temperature maps, and led by University of Münster, Germany (Hiesinger et al. 2020).
- (5) MGNS (Mercury Gamma-Ray and Neutron Spectrometer) to determine the elemental compositions of distinguishable regions over the entire surface of Mercury, and led by Russian Academy of Sciences, Russia (Mitrofanov et al. 2021).
- (6) MIXS (Mercury Imaging X-ray Spectrometer), an instrument that uses X-ray fluorescence analysis to provide a global map of the surface atomic composition, to produce high spatial resolution elemental maps of dedicated Mercury surface regions, and led by the University of Leicester, United Kingdom (Bunce et al. 2020).
- (7) MORE (Mercury Orbiter Radio science Experiment), two-way multi-frequency radio science measurements to determine Mercury's gravity field as well as the size and physical state of its core, and led by University of Rome "La Sapienza", Italy (Iess et al. 2021).
- (8) PHEBUS (Probing of Hermean Exosphere by Ultraviolet Spectroscopy), an UV spectrometer to characterize the composition and dynamics of Mercury's exosphere, and led by LATMOS-IPSL, France (Quemerais et al. 2020).
- (9) SERENA (Search for Exosphere Refilling and Emitted Neutral Abundances), a suite of four sensor units to study the interactions between the surface, exosphere, magnetosphere and the solar wind, and led by the Institute for Astrophysics and Space Planetology, Italy (Orsini et al. 2020).

- (10) SIMBIO-SYS (Spectrometers and Imagers for MPO BepiColombo Integrated Observatory), a camera and spectrometer system to provide global, high-resolution, and IR imaging of the surface and generating a digital terrain model of the entire planet. The hyperspectral imaging push-broom spectrometer will provide the global mineralogical composition of the surface. The consortium is led by the Astronomical Observatory of Padua, Italy (Cremonese et al. 2020).
- (11) SIXS (Solar Intensity X-ray and Particle Spectrometer) to perform measurements of solar X-rays and particles at high-time resolution, and led by the University of Helsinki, Finland (Huovelin et al. 2020).

The spacecraft has also a radiation monitor, BERM (BepiColombo Radiation Monitor), to measure particles and plasma during cruise and nominal mission (Benkhoff et al. 2021).

The Mercury Magnetospheric Orbiter (MMO also nicknamed Mio, Fig. 8.5) is a spin axis-stabilized spacecraft with a spin period of 4 s. The spin axis is nearly aligned with Mercury's. Altitude range is currently expected to be from 590 to 11,639 km, meaning that the most distant point is nearly 6 planetary radii from the planet's center. The main body of the spacecraft is octagonal and would fit inside a circle of 1.8 m diameter. The height of the side panel is 0.9 m. The upper portion of the spacecraft is covered by about 50% solar cells and 50% optical solar reflectors (OSRs). The spacecraft attitude will be determined by a pair of sun sensors on the side panel and a star scanner attached to the bottom surface. The attitude is controlled by the propulsion system with a cold gas jet. Mio accommodates five scientific instruments or scientific instrument suites (Murakami et al. 2020). During cruise, the JAXA-provided Mio spacecraft is a passive passenger, not involved in the control of the composite. This is done within the MPO, while the MTM provides propulsion. During cruise the "spinning" Mio spacecraft is in a fixed position and has to be shielded from sunlight.

The science payload for Mio was selected by the JAXA and contains the following instruments:

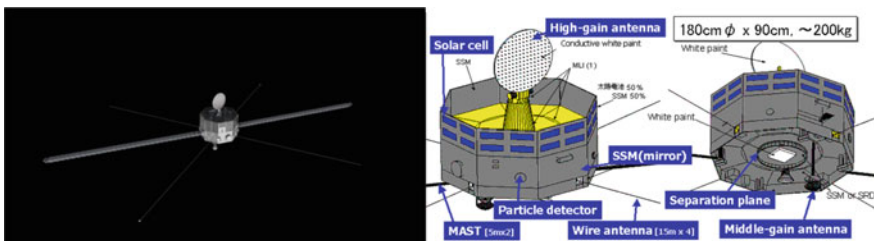


Fig. 8.5 Artist's impression of Mio spacecraft in orbit, and sketch of the spacecraft. Some instrument sensors are mounted at the end of four booms and the mast, respectively

- (1) MGF (Magnetic Field Investigations) an instrument to provide a detailed description of Mercury's magnetosphere and of its interaction with the planetary magnetic field and the solar wind, and led by Austrian Space Science (Baumjohan et al. 2020).
- (2) MPPE (Mercury Plasma Particle Experiment) to study low-and high-energetic particles in the magnetosphere with different sensors: an electron and an ion analyzer, high energy particles sensors for ion and electrons, an energetic neutrals analyzer and a mass spectrum analyzer, and led by JAXA, Japan (Saito et al. 2021).
- (3) PWI (Plasma Wave Investigation) to make a detailed analysis of the structure and dynamics of the magnetosphere. It contains several sensors including two sets of electric field sensors and two kinds of magnetic field sensors, and is led by Tohoku University, Japan (Kasaba et al. 2020).
- (4) MSASI (Mercury Sodium Atmospheric Spectral Imager) to measure the abundance, distribution, and dynamics of sodium in Mercury's exosphere, and led by JAXA, Japan (Murakami et al. 2021).
- (5) MDM (Mercury Dust Monitor) to study the distribution of interplanetary dust in the orbit of Mercury, and led by Chiba Institute of Technology, Japan (Kobayashi et al. 2020).

The MTM provides propulsion means for cruise (Fig. 8.6). Apart from dual-mode bipropellant chemical propulsion, it features electric propulsion with four moveable thrusters based on the Kaufman-type electric bombardment ion motor (125 mN thrust). The high power demand by the MTM electric propulsion (up to 11 kW) is satisfied with large solar arrays (area of over 40 square meters in total), using the same high-temperature technology as for the MPO. In addition, the design must still allow significant flexibility for its orientation to the Sun and emergency situations (for example the ability to survive short-term losses of attitude). More details on the thermal design of the MTM can be found in Tuttle and Cavallo (2009). The MTM structure is based on a CFRP (carbon-fiber reinforced plastic) conical primary structure interfacing with the Launch Vehicle Adapter and the MPO (Kempkens et al. 2019).

8.2.3 *Solar Electric Propulsion to Mercury*

BepiColombo was launched on 20 October 2018 from the European spaceport in Kourou, French Guiana and is now en route to Mercury. The design of an interplanetary trajectory towards Mercury—and most other planets—is driven by energy considerations. To avoid the launch and transport of several hundreds of kilograms of propellant, ESA decided to implement the Solar Electric Propulsion System (SEPS, Sutherland et al. 2019), which in combination with carefully designed planetary swing-bys, allows the spacecraft to enter into a Mercury orbit within less than seven

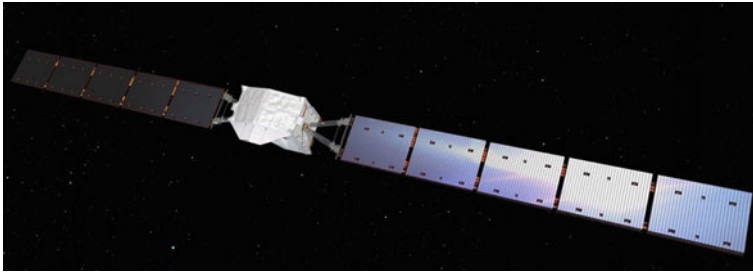


Fig. 8.6 Artist's impression of the Mercury Transfer Module, MTM, with its two solar array wings which provide the power for the SEPS (Solar Electric Propulsion System) during cruise

years. Solar electric propulsion thrust combined with using the gravity of Earth, Venus and Mercury was the final enabling concept to bring the two BepiColombo orbiters, MPO and Mio, to Mercury.

A planetary flyby is a balance of the planet's gravitational potential and the spacecraft's kinetic energy as soon as the spacecraft is in the direct gravitational influence of the planet – its sphere of influence. The radius of the sphere of influence indicates the upper boundary of a swing-by distance; for Venus this is around 600,000 km and for Mercury 110,000 km.

A spacecraft traveling through a planet's sphere of influence is energetically in balance, meaning it does not gain or lose energy. Before the closest approach (CA), the spacecraft gains velocity (kinetic energy) and the gravitational, potential energy decreases. Compensation occurs during the time after the closest approach, and when the spacecraft leaves the sphere of influence, the energy balance towards the planet has not changed. Careful selection of the swing-by trajectory, however, causes the spacecraft to gain or lose energy with respect to the Sun. As a consequence, the planet will lose or gain the same amount of energy, also with respect to the Sun. As the mass of a planet is considerably larger than the mass of a spacecraft, the impact of such an energy transfer on the planet is negligible, but on the spacecraft it is large.

BepiColombo's route to Mercury requires a 7.2-year-long cruise phase, with one Earth flyby, two Venus flybys and six Mercury flybys before arriving at Mercury at the end of 2025. At the time of arrival, the MTM is no longer needed and will be jettisoned. After a weak gravity capture maneuver, both spacecraft will be placed into their dedicated orbits. Their altitude will be adjusted using MPO's thrusters until Mio's desired elliptical polar orbit of $590 \times 11\,640$ km above the planet is reached. Then MPO will separate and descend to its own 480×1500 km orbit using its thrusters. The fine-tuning of the orbits is expected to take three months. The initial mission at Mercury was designed for one Earth year—about four Mercury years—with the possibility of a further one-Earth-year extension. The most limiting constraint to extending the mission still further is the availability of fuel and the possible damage to instruments due to the harsh environment around Mercury.

After launch the spacecraft and all the instruments were successfully tested and commissioned; the spacecraft and instruments are in a good condition. At the time of

writing BepiColombo had already performed two flybys, one at Earth on 10th April 2020 and one at Venus on 15 October 2020. The next flybys are a second Venus flyby in August 2021 and 6 flybys at Mercury distributed over four years (October 1st 2021, June 23rd 2022, June 20th 2023, September 5th 2024, December 2nd 2024 and January 9, 2025).

8.2.4 Operational Constraints During the Cruise Phase

The attitude of BepiColombo during its cruise phase is mainly conditioned by thermal constraints. The radiator panel of MPO has to be directed anti-sunwards and only small deviations are allowed. The constraints on the deviations become harsher (smaller deviation angles) the closer BepiColombo gets to the Sun. Orbit changes are achieved by the planetary flybys, the Solar Electric Propulsion System (SEPS) operations, and the use of the chemical thruster. During the SEPS firing periods no instrument operations are foreseen.

As the Mio spacecraft is protected by a Sun shield during the cruise phase, none of the instruments hosted on Mio are in their final scientific configuration because either the instruments are not deployed yet or their field of view is obstructed by the sunshield. However, the MPO spacecraft does not need Sun shielding.

During nominal cruise operations, one communication pass is scheduled per week, which limits the available data volume for instrument operations. During check-out periods or planetary flybys, the number of antenna passes is increased depending on operational needs. Often, the science campaigns are fostered during these periods, e.g., for planetary swing-by operations. The main influence on the available data volume is, however, the data rate, which is proportional to BepiColombo's distance from the Earth. At a distance closer than 0.54 AU (one AU is the average distance from Earth to the Sun) the maximum data rate is achieved, reducing to a data rate 50 times lower at distances larger than 1.6 AU (Montagnon et al. 2021).

8.3 Cruise Science with BepiColombo During Venus Flybys and Other Opportunities

After BepiColombo's successful launch, both spacecraft (MPO and Mio) are now traveling packed in a composite with the propulsion element (Mercury Composite Spacecraft, also called MCS, see Fig. 8.2). The MCS configuration, due to the Sun shielding, will allow cruise science operations of the Mio instruments only partially and only for those sensors that will be on the low edge of the shield. Onboard the Mio spacecraft, four instruments will be operative: MPPE (partly, not all sensors), MGF, PWI, and MDM. On the MPO spacecraft eight instruments will be able to operate but unfortunately not those pointing in the direction of the planet, since their

view is blocked by the transfer module. The instruments that can operate during the cruise phase are: BERM, ISA, MERTIS, MGNS, MORE, MPO-MAG, PHEBUS, SERENA-MIPA and SERENA-PICAM, SIXS. In addition, there are three “selfie” cameras mounted on the transfer module (Mcam 1,2,3), initially used for monitoring the correct deployment of the solar arrays and the antennas. These can also be used to image a planet during its flyby.

The science investigations during the seven-year-long cruise will be executed by several working groups: Venus Flyby Working Group; Earth Flyby Working Group; Mercury Flyby Working Group; Ground Based Observations Working Group; and the Young Scientist Study Group (Mantagnon et al. 2021), led by representatives from the wider BepiColombo science community.

8.3.1 Science Objectives During Cruise

BepiColombo was launched at the end of a solar cycle, and at Mercury arrival in December 2025, the new solar cycle is expected to be at its peak (Nandy 2021). The cruise phase will thus cover half of a solar cycle, with the chance to study interplanetary processes in this time period.

Coronal mass ejections (CME), solar flares (SF), corotating interaction regions (CIRs), and solar energetic particles (SEPs) change the ions and electrons originating from the Sun and are accelerated (from tens of keV up to 100 s of MeV). They are thus carrying information about the composition of the Sun.

Most measurements of the solar wind and corresponding events originate from observations obtained by Earth-orbiting satellites. Especially in the inner solar system, only a small number of measurements is currently available. Due to the lack of observations, the properties and the evolution of large-scale interplanetary structures needs further study to understand the phenomena themselves, their evolution throughout the heliosphere, and their changes over a solar cycle.

In fact, even limited instrument operations could contribute to a wide range of solar corona and solar wind-related studies. Observations by single spacecraft are not always suited to analyzing the time-varying solar wind and related events as observations provide only single instance observations. Multi-spacecraft observations provide a great opportunity to overcome these shortcomings. Coordinated science observations together with other spacecraft in Earth orbit and in the inner solar system, e.g., Akatsuki (JAXA), Parker Solar Probe (NASA), Solar Orbiter (ESA/NASA) and at a later stage JUICE (ESA) can provide new measurements and/or additional vantage points. The advantage of having several measurement locations in the inner solar system lies in the several geometrically interesting observation opportunities to analyze solar wind as well as transient events. An important geometry constellation is when both spacecraft are aligned on the same magnetic field line. This constellation allows the study of solar wind dynamic behavior. Other geometries are also favorable, for example, to study solar energetic events (CME, flares)

passing through the solar system. As the observational data are of a statistical nature, the data analysis needs to be supported by solar wind models.

Other studies during the cruise phase will encompass the analysis of cometary composition and detection of dust particles of different origins. It is also possible to monitor the local radiation background due to bombardment by energetic particles of galactic cosmic rays as well as the detection and localization of gamma ray burst (GRB). The superior solar conjunctions of BepiColombo can be used for measurements to test general relativity.

As part of the cruise phase, BepiColombo will flyby Earth, Venus twice and Mercury six times. Measurements of the plasma composition, ion flux and density, and magnetic fields in the vicinity of planets will be possible. For more details see Mantagnon et al. (2021).

The Earth flyby took place on 10 April 2020 with the closest approach (CA) at 04.25 UTC (Coordinated Universal Time) at a distance of 12,700 km. When the spacecraft approached the Earth, it passed the bow-shock, a boundary at which the speed of the solar wind abruptly decreases as a result of its approach to the planet's magnetosphere. The CA as well as the exit from the magnetosphere and the ionosphere took place on the night side. All in-situ instruments were operational during the flyby and able to identify the plasma boundary, the ionosphere, and radiation belt crossings. Investigations include the cross-calibration between Mio and MPO instruments as well as with instrumentation from other spacecraft around Earth. The MERTIS spectrometer onboard MPO obtained first spectra from the Moon, which allowed a first in-flight calibration of the instrument.

In addition, two Venus flybys (see below) and six Mercury flybys are needed before the spacecraft can be placed into a polar orbit around Mercury. The six Mercury flybys will take place from late 2021 to early 2025, with various geometries that offer a variety of science investigations to be performed by instruments onboard the MPO and Mio spacecraft. The CA distances of the first four flybys will be on the order of 200 km from the surface, the fifth flyby CA is planned at 40,000 km, and the last flyby at 400 km. The flyby geometries will probe various parts of the exosphere and magnetosphere of Mercury, a critical aspect for improving the understanding of the dynamics of Mercury's environment.

8.3.2 *Science During Venus Flybys*

BepiColombo's first Venus flyby was an opportunity to use instruments for scientific purposes in a planetary environment. This flyby took place on October 15th, 2020, almost two years after launch, at 03:58:31 UTC (CA).

The image below (Fig. 8.7) shows the flyby trajectories of BepiColombo using Venus solar orbital (VSO) coordinates at Venus on October 15, 2020 (Bepi I, red) and August 10, 2021 (Bepi II, blue). Numbers plotted along the orbits give the expected universal time. Green and cyan profiles show average positions of bow-shock and ion composition boundary respectively.

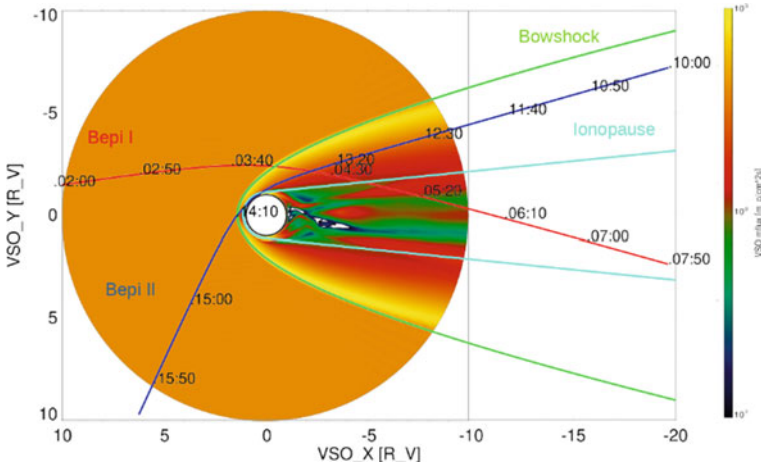


Fig. 8.7 Venus flybys trajectories in XY plan, VSO system (courtesy of Markus Fraenz, MPI, Goettingen, Germany). Units are in Venusian Radii

As Fig. 8.7 shows, BepiColombo approached the planet from the dayside and—given the retrograde rotation of Venus—it had the CA on the evening side, almost at the time of crossing the bow-shock. Minimum distance from the center of the planet was about 16,800 km (that is, an altitude of about 10,750 km above the planet surface). Distance from the Earth was about 1.16 AU and from the Sun 0.71 AU. Hence, the flyby occurred at an altitude of about 2 Venus radii. The spacecraft crossed the bow-shock near the CA time and in the following the ionopause at <10 radii and stayed in the ionotail for several hours. Exit from the bow-shock occurred even later (Fig. 8.8).

The first Venus flyby configuration is optimal for both atmospheric and ionospheric/magnetospheric investigations of the Venus close environment. As far as

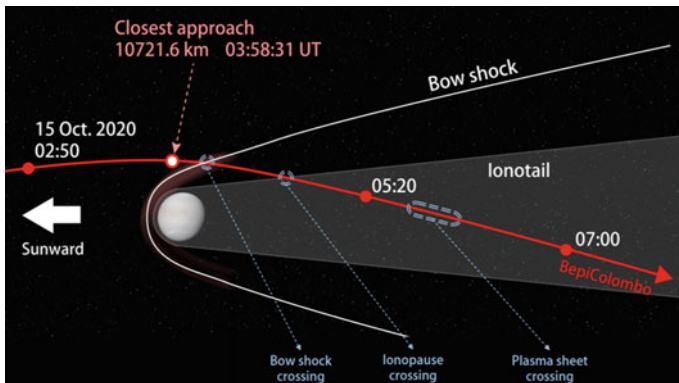


Fig. 8.8 Boundary crossings at Venus by the BepiColombo spacecraft during the first flyby

planetary investigations of Venus are concerned, some instruments were acquiring data only in the hours around the flyby (the magnetometer and other sensors onboard MMO, PHEBUS, SERENA-MIPA and -PICAM, MERTIS, SIXS), while others like MAG, ISA, MGNS, BERM and MORE are operating almost continuously during the cruise. The scientific outputs of these observations are currently being analyzed and first results are already published in relevant journals (e.g. Volwerk et al. 2021).

The second flyby at Venus took place about 10 months after the first on 10 August 2021. The much lower altitude above the surface (552 km, see Fig. 8.7) allowed a different range of investigations. Another feature that mainly differs the second flyby from the first is that Venus was approached from the nightside, even if still approaching the planet from the evening terminator.

Given the BepiColombo instrumentation that can operate during cruise and considering attitude and field of view limitations due to the stacked configuration of BepiColombo during the cruise (Fig. 8.1), the topics of investigation at Venus are the following:

Atmosphere. In the 2020 and 2021 close encounter periods, and in selected periods where BepiColombo is between 0.3 and 0.8 AU from Venus, Venus's atmosphere was observed by MERTIS (the 7–14 μm infrared spectrometer and 7–40 μm radiometer) and PHEBUS (the 4-channel ultra violet (UV) spectrometer). In October 2020 MERTIS observed Venus' dayside near-continuously over the 55-h period leading up to closest approach. MERTIS sensed Venus' middle atmosphere and cloud layers, providing temperature profiles and detecting CO_2 , SO_2 , H_2SO_4 cloud aerosol properties. MERTIS observations will contribute to studies of Venus' radiative balance, atmospheric structure, cloud level chemical processes, and the impact of global-scale atmospheric waves on Venus' weather patterns. At closest approach on 15 October 2020 the PHEBUS spectrometer was used to complete nightside atmosphere stellar occultations in the F(ar)UV channel (145–315 nm), providing density profiles of bulk and trace gas species, and detecting NO emission on Venus' nightside. These data are useful for studies of the atmospheric structure between 80 and 200 km altitude, and the zonal transport processes in the thermosphere (at about 110 km). The MGNS neutron and gamma spectrometers also contributes to the investigation of Venus' atmospheric chemistry and dynamics by detecting elemental composition in terms of C, O, N and H_2O . The BepiColombo-enabled Venus atmosphere investigations were further supported by complementary observations completed by the Japanese Akatsuki spacecraft presently orbiting around Venus, and other Earth-based telescopes, expanding the scope of studying Venus' climate.

Ionosphere—Induced magnetosphere. Venus has no intrinsic dipole magnetic field and thus the solar wind plasma can easily approach the planet and interact directly with its upper atmosphere, leading to the formation of ion plasma boundaries and an induced magnetosphere. The ionopause is created by direct interaction with the EUV flow from the Sun.

During the BepiColombo passage, the investigation of the extended region from the outer unperturbed solar wind, through the bow-shock, the magneto-sheath, ionosphere and magnetotail, was performed by several instruments onboard. First of all,

the two magnetometers (MAG and MGF) who provided magnetic field measurements throughout the whole period. Then, a wide set of ion, electron and plasma sensors at different energy ranges of the particle instruments suites, SERENA, PWI, and MPPE, extended these measurements. Their joint measurements are fundamental to detect the position of the different plasma boundaries and composition of the different layers of the complex ionized environment. Moreover, detection of draped dayside magnetic fields, low-frequency wave activity in the ion composition boundary, flux ropes and tail lobe polarity were all interesting goals of these measurements. In addition, measurements of energetic neutral atoms as derived from solar wind ion sputtering over the exobase, electron shielding effects and detection of atmospheric pick-up ions could help to relate the ionized and neutral components of the Venus environment, hence the induced magnetosphere with the atmosphere themselves. First results of these combined measurements are published by Persson et al. (2022).

8.4 Summary and Conclusions

BepiColombo is set to build on the achievements of previous spacecraft missions (Mariner 10, MESSENGER) to provide the best understanding of the solar system's innermost planet to date. Investigations of these missions raised many questions that scientists did not even consider when designing these missions, which are now left open for BepiColombo to answer. Not only will a new mission provide complementary observations (separated by more than a decade) that will allow any changes to be compared and constrained, but it will also make many new observations.

One obvious difference from previous missions is that BepiColombo comprises two spacecraft in different orbits, affording new science possibilities. In particular, dual observations are key to understanding solar-wind-driven magnetospheric processes, and this will allow unprecedented observations of the planet's magnetic field and the interaction of the solar wind with the planet at two different locations at the same time.

Because MPO's orbit is not highly elliptical, and the instruments will be mainly pointing directly towards Mercury, errors in height measurements of the surface topography will be reduced. This will allow, for example, improvement of gravity and topography models, and also give a higher-resolution coverage of surface features.

In addition, during its seven-year-long cruise, BepiColombo can perform science investigations at Earth, Venus, and Mercury during its flybys, heliophysics studies of the solar wind, and the test of general relativity during solar superior conjunctions. Especially the two flybys at Venus will give scientists some opportunities to study the atmosphere and ionosphere of the other planet between the Earth and the Sun in more detail.

References

- Anderson, D.J., G. Colombo, P.B. Esposito, E.L. Lau, and G.B. Trager. 1987. The mass, gravity field, and ephemeris of Mercury. *Icarus* 71:337–349. [https://doi.org/10.1016/0019-1035\(87\)90033-9](https://doi.org/10.1016/0019-1035(87)90033-9).
- Baumjohann W., A. Matsuoka, Y. Narita, W. Magnes, D. Heyner, K.-H. Glassmeier, R. Nakamura, D. Fischer, F. Plaschke, M. Volwerk, T.L. Zhang, H.-U. Auster, I. Richter, A. Balogh, C. Carr, M. Dougherty, T.S. Horbury, H. Tsunakawa, M. Matsushima, M. Shinohara, H. Shibuya, T. Nakagawa, M. Hoshino, Y. Tanaka, B.J. Anderson, C.T. Russell, U. Motschmann, F. Takahashi, and A. Fujimoto. 2020. The BepiColombo–mio magnetometer en route to mercury. *Space Science Reviews* 216.
- Benkhoff, Johannes, van Casteren, Jan, Hayakawa, Hajime, Fujimoto, Masaki, Laakso, Harri, Novara, Mauro, Ferri, Paolo, Middleton, Helen R. and Ziethe, Ruth. 2010. BepiColombo—Comprehensive exploration of Mercury: Mission overview and science goals.
- Benkhoff, J., G. Murakami, W. Baumjohann, S. Besse, E. Bunce, M. Casale, G. Cremosese, K.-H. Glassmeier, H. Hayakawa, D. Heyner, H. Hiesinger, J. Huovelin, H. Hussmann, V. Iafolla, L. Iess, Y. Kasaba, M. Kobayashi, A. Milillo, I.G. Mitrofanov, E. Montagnon, M. Novara, S. Orsini, E. Quemerais, U. Reininghaus, Y. Saito, F. Santoli, D. Stramaccioni, O. Sutherland, N. Thomas, I. Yoshikawa, J. Zender. 2021. BepiColombo—mission overview and science goals. *Space Science Reviews* 217. doi:10.1007/s11214-021-00861-4
- Blewett, D.T., et al. 2011. Hollows on mercury: Evidence for geologically recent volatile-related activity. *Science* 333:1856–1859. <https://doi.org/10.1126/science.1211681>.
- Blewett, D.T., et al. 2016. Analysis of MESSENGER high-resolution images of Mercury’s hollows and implications for hollow formation. *Journal of Geophysical Research Planets* 121:1798–1813. <https://doi.org/10.1002/2016JE005070>.
- Bunce, Emma J., Adrian Martindale, Simon Lindsay, Karri Muinonen, David A. Rothery, Jim Pearson, Ivor McDonnell, Chris Thomas, et al. 2020. The BepiColombo mercury Imaging X-ray spectrometer: science goals, instrument performance and operations. *Space Science Reviews* 216.
- Chabot, N.L., C.M. Ernst, B.W. Denevi, H. Nair, A.N. Deutsch, D.T. Blewett, S.L. Murchie, A.N. Deutsch, J.W. Head, S.S. Solomon. 2016. Imaging Mercury’s polar deposits during MESSENGER’s low-altitude campaign. *Geophysical Research Letters* 43(18):9461–9468.
- Chabot, N.L., E.E. Shread, and Harmon, J.K., Investigating Mercury’s south polar deposits: Arecibo radar observations and high-resolution determination of illumination conditions. *Journal of Geophysical Research Planets* 123:666–681. <https://doi.org/10.1002/2017JE005500>.
- Cremonese, G., F. Capaccioni, M.T. Capria, A. Doressoundiram, P. Palumbo, M. Vincendon, M. Massironi, S. Debei, et al. 2020. SIMBIO-SYS: cameras and spectrometer for the BepiColombo mission. *Space Science Reviews* 216.
- Ferrero, A., D. Battaglia, T. Malosti, D. Stramaccioni, J. Schilke. 2016. The challenges of the thermal design of BepiColombo mercury planet orbiter. In *ICES-2016–212, 46th International Conference on Environmental Systems*, Vienna.
- Genova, Antonio, Hauke Hussmann, Tim Van Hoolst, Daniel Heyner, Luciano Iess, Francesco Santoli, Nicolas Thomas, Patrick Kolhey, Benoit Langlais, Johannes Z. D. Mieth, Joana S. Oliveira, Alexander Stark, Nicola Tosi, Johannes Wicht, and Johannes Benkhoff, Geodesy. 2021. Geophysics and fundamental physics investigations of the BepiColombo mission. *Space Science Reviews* 216.
- Harmon, J.-K., and M.-A. Slade. 1992. Radar mapping of mercury: Full-disk images and polar anomalies. *Science* 258: 640–643.
- Harmon, J.K., M.A. Slade, R.A. Vélez, A. Crespo, M.J. Dryer, and J.M. Johnson. 1994. Radar mapping of Mercury’s polar anomalies. *Nature* 369: 213–215.
- Heyner, D., et al. 2021. The fluxgate magnetometer of the BepiColombo mercury planetary orbiter. *Space Science Reviews* 217. 10.1007/s11214-021-00822-x.
- Hiesinger, H., J. Helbert, G. Alemano, K.E. Bauch, M. D’Amore, A. Maturilli, A. Morlok, M.P. Reitze, C. Stangarone, A.N. Stojic, I. Varatharajan, I. Weber and the MERTIS Co-I Team. 2020.

- Studying the composition and mineralogy of the Hermean surface with the mercury radiometer and thermal infrared spectrometer (MERTIS) for the BepiColombo mission: an update. *Space Science Reviews* 216.
- https://ui.adsabs.harvard.edu/link_gateway/2021AA...656A..11V/https://doi.org/10.1051/0004-6361/202140910
- Huovelin, J., R. Vainio, E. Kilpua, A. Lehtolainen, S. Korpela, E. Esko, K. Muinonen, E. Bunce, A. Martindale, M. Grande, H. Andersson, S. Nenonen, J. Lehti, W. Schmidt, M. Genzer, T. Vihavainen, J. Saari, J. Peltonen, E. Valtonen, M. Talvioja, P. Portin, S. Narendranath, R. Jarvinen, T. Okada, A. Milillo, M. Laurenza, E. Heino, P. Oleynik. 2020. Solar intensity X-ray and particle spectrometer SIXS: instrument design and first results. *Space Science Reviews* 216.
- Iess, L., et al. 2021. *Space Sci. Rev.*
- Jarvinen, Stavro Lambrov Ivanowski, Ákos Madár, Géza Erdős, Christina Plainaki, Tommaso Alberti, Sae Aizawa, Johannes Benkhoff, Go Murakami, Eric Quemerais, Harald Hiesinger, Igor G. Mitrofanov, Luciano Iess, Francesco Santoli, Stefano Orsini, Herbert Lichtenegger, Gunther Laky, Stas Barabash, Richard Moissl, Juhani Huovelin, Yasumasa Kasaba, Yoshifumi Saito, Masanori Kobayashi, and Wolfgang Baumjohann. 2021. BepiColombo science investigations during cruise and flybys at the Earth, Venus and Mercury. *Space Science Reviews* 216.
- Kasaba, Yasumasa; Hirotsugu Kojima, Michel Moncuquet, Jan-Erik Wahlund, Satoshi Yagitani, Fouad Sahraoui, Pierre Henri, Tomas Karlsson, Yoshiya Kasahara, Atsushi Kumamoto, Keigo Ishisaka, Karine Issautier, Gaetan Wattieaux, Tomohiko Imachi, Shoya Matsuda, Janos Lichtenberger, and Hideyuki Usui. 2020. Plasma Wave Investigation (PWI) aboard BepiColombo Mio on the trip to the first measurement of electric fields, electromagnetic waves, and radio waves around Mercury. *Space Science Reviews* 216.
- Kempkens, K., F. Striedter, H. Gray, S. Clark, K. Chan, and N. Wallace. 2019. BepiColombo—The mercury transfer module. In *36th International Electric Propulsion Conference*, Sept 15–20, University of Vienna, Austria.
- Kobayashi, M., K. Nogami, M. Fujii, H. Ohashi, T. Miyachi, S. Sasaki, S. Hasegawa, H. Yano, H. Shibata, T. Iwai, S. Minami, S. Takechi, E. Grun, R. Srama. 2020. Development of the mercury dust monitor (MDM) on board BepiColombo mission. *Space Science Reviews* 216.
- Mangano, V., Valeria Mangano, Melinda Dósa, Markus Fraenz, Anna Milillo, Joana S. Oliveira, Yeon Joo Lee, Susan McKenna-Lawlor, et al. 2004. An international program for Mercury exploration: synergy of MESSENGER and BepiColombo. *Advance in Space Research* 33:2126–2132.
- McNutt Jr, Ralph L., Sean C. Solomon, Réjean Grard, Mauro Novara, and Toshifumi Mukai. 2004. An international program for Mercury exploration: synergy of MESSENGER and BepiColombo. *Advances in Space Research* 33 (12): 2126–2132 S0273117703004393. [https://doi.org/10.1016/S0273-1177\(03\)00439-3](https://doi.org/10.1016/S0273-1177(03)00439-3).
- McNutt, R. L., S.C. Jr., Solomon, R.E. Gold, J.C. Leary and the MESSENGER Team. 2006. The MESSENGER mission to Mercury: Development history and early mission status. *Advance Space Research* 38:564–571. <https://doi.org/10.1016/j.asr.2005.05.044>.
- Milillo, A., M. Fujimoto, G. Murakami, J. Benkhoff, J. Zender, S. Aizawa, M. Dósa, L. Grifton, et al. 2020. Investigating Mercury’s environment with the two-spacecraft BepiColombo mission. *Space Science Reviews* 216.
- Mitrofanov, I.G., A.S. Kozыrev, D.I. Lisov, M.L. Litvak, A.A. Malakhov, M.I. Mokrousov, J. Benkhoff, A. Owens, R. Schulz, and F. Quarati. 2021. Mercury gamma-ray and neutron spectrometer MGNS for planetary orbiter of BepiColombo mission: updates of design and the first measurements in space. *Space Science Reviews* 216.

- Montagnon, E., F. Budnik, M. Casale, S. de la Fuente, S. Martinez, G. Murakami, M. Ogawa, C. Steiger and M. Yamashita 2021. BepiColombo ground segment and mission operations. *Space Science Reviews* 216.
- Murakami, G., Hajime Hayakawa, Hiroyuki Ogawa, Shoya Matsuda, Taeko Seki, Yasumasa Kasaba, Yoshifumi Saito, Ichiro Yoshikawa, Masanori Kobayashi, Wolfgang Baumjohann, Ayako Matsuoka, Hirotsugu Kojima, Satoshi Yagitani, Michelle Moncuquet, Jan-Erik Wahlund, Dominique Delcourt, Masafumi Hirahara, Stas Barabash, Oleg Korablev, and Masaki Fujimoto. 2020. Mio–First comprehensive exploration of Mercury’s space environment: mission overview. *Space Science Reviews* 216.
- Murakami, G., I. Yoshikawa, S. Kameda, O. Korablev, V. Kottsov, M. Kuwabara, T. Sato, Y. Suzuki, K. Yoshioka, and A. Tavrov. 2021. Mercury sodium atmosphere spectral imager (MSASI) onboard the BepiColombo/Mio spacecraft: overviews, calibrated performances, and observation plans, The Mercury Sodium Atmospheric Spectral Imager for the MMO Spacecraft of BepiColombo. *Space Science Reviews* 216.
- Nandy, D. 2021. Progress in solar cycle predictions: Sunspot cycles 24–25 in perspective. *Solar Physics* 296. <https://doi.org/10.1007/s11207-021-01797-2>.
- Orsini, S., S. Livi, H. Lichtenegger, S. Barabash, A. Milillo, E. De Angelis, M. Phillips, G. Laky, et al. 2020. SERENA: particle instrument suite for Sun-Mercury interaction insights on-board BepiColombo. *Space Science Reviews* 216.
- Persson, M., S. Aizawa, N. André, S. Barabash, Y. Saito, Y. Harada, D. Heyner, S. Orsini, A., Fedorov, C. Mazelle, Y. Futaana, L.Z. Hadid, M. Volwerk, G. Collinson, B. Sanchez-Cano, A. Barthe, E. Penou, S. Yokota, V. Génot, J.A. Sauvaud, D. Delcourt, M. Fraenz, R. Modolo, A. Milillo, H.-U. Auster, I. Richter, J.Z.D. Mieth, P. Louarn, C.J. Owen, T.S. Horbury, K. Asamura, S. Matsuda, H. Nilsson, M. Wieser, T. Alberti, A. Varsani, V. Mangano, A. Mura, H. Lichtenegger, G. Laky, H. Jeszenszky, K. Masunaga, C. Signoles, M. Rojo, and G. Murakami. 2022. BepiColombo mission confirms stagnation region of Venus and reveals its large extent. *Nature Communications* 13(1): 7743. <https://doi.org/10.1038/s41467-022-35061-3>.
- Quemerais, Eric, Jean-Yves Chaufray, Dimitra Koutroumpa, Francois Leblanc, Aurélie Reberac, Benjamin Lustrement, Christophe Montaron, Jean-Francois Mariscal, Nicolas Rouanet, Ichiro Yoshikawa, Go Murakami, Kazuo Yoshioka, Oleg Korablev, Denis Belyaev, Maria G. Pelizzo, Alan J. Corso, and Paola Zuppella, PHEBUS on Bepi-Colombo: Post-launch Update and Instrument Performance. *Space Science Reviews* 216 (2020).
- Rothery, Dave A., Matteo Massironi, Giulia Alemanno, Océane Barraud, Sebastien Besse, Nicolas Bott, Rosario Brunetto, Emma Bunce, et al. 2020. Rationale for BepiColombo studies of Mercury’s surface and composition. *Space Science Reviews* 216.
- Saito, Y., J.A. Sauvaud, M. Hirahara, S. Barabash, D. Delcourt, T. Takashima, K. Asamura and the BepiColombo MMO/MPPE Team. 2021. Scientific objectives and instrumentation of Mercury Plasma Particle Experiment (MPPE) onboard MMO. *Space Science Reviews* 216.
- Santoli, F., E. Fiorenza, C. Lefevre, D.M. Lucchesi, M. Lucente, C. Magnafico, A. Morbidini, R. Peron, V. Iafolla. 2020. ISA, a high sensitivity accelerometer in the interplanetary space Updates after the Near-Earth Commissioning Phase of Italian Spring Accelerometer/ISA Space. *Space Science Reviews* 216.
- Solomon, S.C., and B.J. Anderson 2018. The MESSENGER Mission: Science and implementation overview. In *Mercury-The View after MESSENGER*, eds., Solomon S.C., L.R. Nittler, B.J. Anderson, 1–29. Cambridge, Cambridge University Press.
- Solomon, S.C., R.L. Jr. McNutt, R.E. Gold, D.L. Domingue. 2007. MESSENGER mission overview. *Space Science Review* 131:3–39. <https://doi.org/10.1007/s11214-007-9247-6>.
- Solomon S.C., L.R. Nittler, B.J. Anderson (eds.). 2018. *Mercury-the view after MESSENGER*, 581 p., Cambridge University Press, Cambridge, <https://doi.org/10.1017/9781316650684>.
- Strom, R.G., and A.L. Sprague. 2003. *Exploring mercury: The iron planet*. New York, Springer Verlag Heiderberg, ISBN 1852337311.

- Sutherland O., D. Stramaccioni, J. Benkhoff, N. Wallace, A. Rocchi, R. Jehn. 2019. BepiColombo: ESA's Interplanetary Electric Propulsion Mission to Mercury. In *36th international electric propulsion conference*, Sept 15–20, Austria, University of Vienna.
- Thomas N., H. Hussmann, T. Spohn, L.M. Lara, U. Christensen, M. Affolter, T. Bandy, T. Beck, et al. 2021. The BepiColombo laser altimeter. *Space Science Reviews* 216.
- Tuttle, S., G. Cavallo. 2009. Thermal design of the mercury transfer module. In *39th International conference on environmental Systems*, Savannah, Georgia.

Chapter 9

Analysis of Smart Dust-Based Frozen Orbits Around Mercury



Generoso Aliasi, Lorenzo Niccolai, Alessandro A. Quarta,
and Giovanni Mengali

Abstract According to the classical two-body Keplerian model, the orbital parameters of a spacecraft are constant during a mission. However, real-life spacecraft motion is different from a classical Keplerian model due to the presence of perturbing forces, whose effects are usually undesirable, especially for observation and communication spacecraft that require accurate pointing capabilities. Therefore, active control systems are usually required to maintain the working orbit. However, an alternative strategy consists of suitably selecting the initial orbital elements to generate a “frozen orbit”, which on average maintains some of the design orbital elements. The utilization of spacecraft with large area-to-mass ratio could extend the flexibility on the initial choice of orbital parameters. In this context, a novel option is represented by smart dusts (SDs), which are femto-satellites with large area-to-mass ratio (or millimeter-scale solar sails). In this chapter, a double-averaging technique is used to determine planetocentric frozen orbits maintained by SDs. In particular, a numerical analysis of frozen orbits is discussed, with a special application focused on orbits around Mercury, which are fit for an SD-based scenario due to their closeness to the Sun and the absence of atmospheric drag.

9.1 Introduction

The Keplerian model of gravitation confines the motion of a spacecraft relative to a primary body to a conical trajectory, whose plane is fixed with respect to an inertial reference frame. The actual spacecraft motion is, however, different from a classical Keplerian model, due to the presence of perturbing forces, such as those

G. Aliasi · L. Niccolai · A. A. Quarta · G. Mengali (✉)
Department of Civil and Industrial Engineering–Aerospace Division, University of Pisa, Pisa, Italy
e-mail: g.mengali@ing.unipi.it

L. Niccolai
e-mail: lorenzo.niccolai@ing.unipi.it

A. A. Quarta
e-mail: a.quarta@ing.unipi.it

induced by the inhomogeneity of the gravitational field of the primary body, the atmospheric drag, or the force exerted by other celestial bodies (Capderou 2005). The effects of these perturbing forces are usually undesirable, especially for observation and communication spacecraft that require accurate pointing capabilities. Traditionally, the most commonly used strategy to overcome this issue is the use of sophisticated systems on board dedicated to active orbit control. Possible alternatives are obtained either by suitably selecting the initial orbital elements to generate a “frozen orbit”, which on average maintains some of the initial orbital elements, or by exploiting the perturbations to generate a desirable variation of a specific orbital element. Typical examples of such applications are the classical Molniya orbits or the Sun-synchronous orbits, designed by considering the effects of the primary body oblateness (Chobotov 2002).

The existence of frozen orbits around different bodies in the solar system has been studied by several authors, who have addressed the problem by considering different perturbative effects, apart from the primary’s oblateness. For example, Coffey et al. (1994) studied frozen orbits with a mathematical model including zonal perturbations up to the ninth zonal harmonic using a Lie transformation to average the system Hamiltonian. Within the same gravitational model, Lara et al. (1995) used a numerical continuation method of periodic orbits to find new families of frozen orbits. Park and Junkins (1995) proposed lunar mapping by means of selenocentric frozen orbits, which were obtained by neglecting the moon’s oblateness, but accounting for the Sun’s and Earth’s gravitational perturbations. The effects of the J_2 coefficient combined with the perturbation of a third body on a circular orbit were studied by Scheeres et al. (2001), using both analytical averaging and numerical techniques, for a spacecraft orbiting around the Galilean moon Europa. Later, Paskowitz and Scheeres (2006) included the effect of the third zonal harmonic, and Lara and Russel (2008) used a high-fidelity geopotential both to search for geocentric frozen orbits and to investigate their orbital stability. Abad et al. (2009) and San-Juan et al. (2009) showed that families of frozen orbits exist for a lunar orbiter when the gravitational force of the Earth and the zonal perturbations of the moon up to the seventh zonal harmonic are considered. Using a power series expansion of the Hamiltonian function, Lara et al. (2010) studied frozen orbits around Mercury in the elliptic restricted three-body problem, including both the J_2 and J_3 terms in the gravity field of Mercury. Delsate et al. (2010) developed a simple analytical theory based on an averaging method of the Hamiltonian to find frozen orbits around an oblate primary body, including the effect of a third body on an elliptic orbit around the primary. Scheeres (2012) investigated the existence of frozen orbits around small celestial bodies, as comets and asteroids. More recently, Masoud et al. (2018) studied families of geocentric frozen orbits considering the Earth’s oblateness and the perturbation due to solar radiation pressure but neglecting the problem of orbital stability. In this context, Alessi et al. (2019) proposed a phase-space description of the spacecraft dynamics perturbed by planetary oblateness and solar radiation pressure, while Circi et al. (2019) proposed

to exploit frozen orbits for mapping an irregularly shaped asteroid surface. Finally, Colombo (2020) used models with either single- or double-averaged potential to identify highly elliptical geocentric orbits, which could be exploited as graveyard orbits for satellites after the end of their operational life.

All of the above works involve the effect of natural perturbations on the spacecraft motion. Typically, the results obtained show the existence of a number (or a family) of frozen orbits whose classical elements are constrained within a small set of values. A possible way to extend the admissible solutions is by means of a continuous thrust system, which allows an increased set of achievable values of orbital elements for frozen or Sun-synchronous orbits (Anderson and Macdonald 2013; Macdonald et al. 2010). In this context, Russell (2012) published a comprehensive survey of spacecraft dynamics in a strongly perturbed environment, also accounting for the possibility of providing a propulsive acceleration. The continuous thrust required for maintaining such orbits could be provided by a solar sail, which exploits the solar radiation pressure to generate a propulsive acceleration without consuming any propellant (McInnes 1999). In that case, the solar radiation pressure is no longer a perturbation that has to be counteracted, but is instead used as a source of propulsive acceleration.

A number of papers exist in which a solar sail is used to design a mission on the basis of a continuous natural thrust, including Sun-synchronous polar orbits (Leipold and Wagner 1996), orbits around Sun-Earth Lagrangian points (Farrés and Jorba 2008; Niccolai et al. 2020; Sood and Howell 2019), and non-Keplerian orbits (Gong and Li 2014; Quarta et al. 2019a, 2020). In this context, Gong et al. (2012) proposed the generation of geocentric Sun-synchronous orbits by means of a solar sail, accounting for Earth's oblateness and atmospheric drag, but not for perturbations induced by a third body. Recently, Tresaco et al. (2016, 2018) investigated the possibility of generating frozen orbits for planetary observations by means of a solar sail-based spacecraft. In particular, their approach showed that the application of a Hamiltonian averaging method could be extended to a situation in which the solar radiation pressure is included in the mathematical model. However, their analysis was confined to conventional solar sails and, most importantly, the orbital stability problem has not been addressed (Tresaco et al. 2016), or only checked through specific numerical simulations (Tresaco et al. 2018). Finally, Khattab et al. (2020) extended the analysis of solar sail-based frozen orbits to selenocentric scenarios by means of a double-averaging technique, but they did not investigate the orbital stability.

A novel concept in terms of photonic propulsion is represented by smart dusts (SDs), which are femto-satellites (or millimeter-scale solar sails) with a large area-to-mass ratio and, as such, are extremely sensitive to pressure forces. Actually, a SD may exploit the solar radiation pressure to suitably modify its orbital parameters, similarly to the way a solar sail is able to. In its typical configuration, a SD is equipped with solar cells, a payload (sensor), a processor, and a communication system, all being composed of miniaturized components and integrated on a common substrate

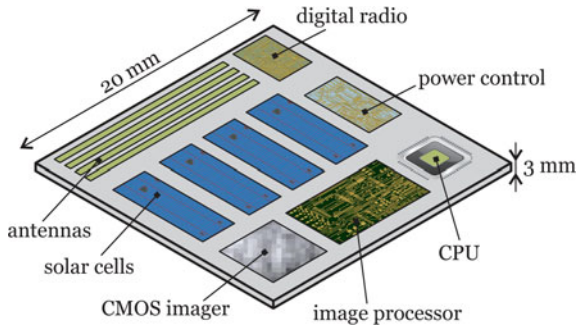


Fig. 9.1 Basic scheme of a SpaceChip, a precursor of Smart Dust concept. Adapted from Barnhart et al. (2007) and Niccolai et al. (2019)

(see Fig. 9.1). The peculiar characteristics of SDs allow innovative mission scenarios to be envisaged (Lücking et al. 2012; Mengali et al. 2018; Quarta et al. 2019b, c). In particular, a distributed-payload mission, in which a constellation of SDs released by a mother spacecraft cover one or more frozen orbits, could be successfully used for planetary observation and mapping, as suggested by Colombo and McInnes (2012).

A SD manufactured with a special design may also be able to passively maintain a Sun-facing attitude (Atchison and Peck 2010), thus providing a propulsive acceleration always aligned along the Sun–spacecraft direction. A comprehensive review of the SD concept and its possible applications is beyond the scopes of this chapter, but the interested reader may refer to the recent review of Niccolai et al. (2019).

The mathematical model discussed in this chapter uses the “doubly-averaging technique” proposed by Delsate et al. (2010) to investigate the possibility of maintaining planetocentric frozen orbits with a SD device, thus extending existing results (Carbone et al. 2020; Khattab et al. 2020; Tresaco et al. 2016, 2018) to the SD case and including the stability analysis of such a frozen orbit. The capability of SDs to self-stabilize a Sun-facing attitude and the possibility of performing a scientific observation with a distributed payload make these satellites perfectly fit for planetocentric observation missions that exploit frozen orbits. It is shown how the assumption that the SD passively maintains a Sun-facing attitude permits the use of the averaging technique developed by Tremaine et al. (2009). Then, a numerical analysis of the frozen orbits is presented with an application to the orbits around Mercury to emphasize the effects of the solar radiation pressure and to verify whether this concept is well suited for a SD-based planetary observation mission. Indeed, a mission around Mercury seems appropriate for an SD-based scenario, because the atmospheric drag (to which SDs are very sensitive) is absent and the short Sun–spacecraft distance makes solar radiation pressure more effective. The stability of the families of frozen orbits obtained is analytically checked. Finally, some numerical simulations of orbits around Mercury are presented, and the limitations of the theoretical results are discussed.

9.2 Mathematical Preliminaries

Consider a SD orbiting around a non-spherical planet with gravitational parameter μ_P , as sketched in Fig. 9.2. The SD motion is perturbed by the gravitational effects of the Sun and by the thrust induced by the solar radiation pressure. Assume that the SD is always in a Sun-facing configuration, that is, the outer normal to the SD surface is parallel to the Sun–SD vector. Note that this configuration can be achieved in a passive way by suitable SD design (Atchison and Peck 2010). The motion of the SD is therefore described by the Hamiltonian function

$$H = \frac{1}{2}v^2 - \phi_P - \phi_S - \phi_T \tag{9.1}$$

where v is the SD inertial velocity with respect to the planet, whereas ϕ_P , ϕ_S , and ϕ_T represent the potentials induced by the planet’s gravity, the Sun’s gravity, and the propulsive acceleration, respectively.

The potential of the planet is usually written using an expansion in terms of associated Legendre functions $P_{lm}(\sin\delta)$ and spherical harmonic coefficients C_{lm} and S_{lm} as (Beutler 2005):

$$\phi_P = \frac{\mu_P}{r} \sum_{l=0}^{\infty} \sum_{m=0}^l \left(\frac{R_P}{r}\right)^l P_{lm}(\sin \delta)[C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda)] \tag{9.2}$$

where R_P is the equatorial radius of the planet, whereas $\{r, \lambda, \delta\}$ are the planetocentric coordinates of the SD.

The other two potentials can be written as (Aliasi et al. 2011; Beutler 2005):

$$\phi_S = \mu_S \left(\frac{1}{\|\mathbf{r} - \mathbf{r}_S\|} - \frac{\mathbf{r} \cdot \mathbf{r}_S}{\|\mathbf{r}_S\|^3} \right) \text{ and } \phi_T = -\beta \frac{\mu_S}{\|\mathbf{r} - \mathbf{r}_S\|} \tag{9.3}$$

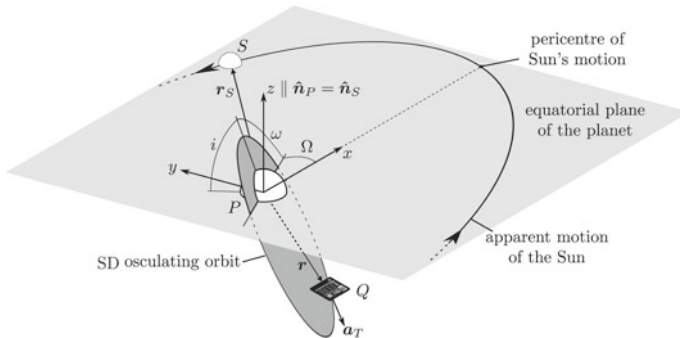


Fig. 9.2 Geometrical model of the system used for numerical integrations

where \mathbf{r} and \mathbf{r}_S are the position vectors of SD and Sun with respect to the planet, μ_S is the Sun's gravitational parameter, and β is the SD lightness number. In particular, the latter is a parameter that characterizes the performance of a classical solar sail (McInnes 1999), given by

$$\beta \triangleq 2\eta P_{\oplus}(A/m)$$

where η is the sail surface reflection coefficient, $P_{\oplus} = 4.56 \times 10^{-6}$ Pa is the solar radiation pressure at 1 au from the Sun, and (A/m) is the spacecraft area-to-mass ratio. The definition of the lightness number can be easily extended to a SD. In this case, it is the ratio of the SD propulsive acceleration to the Sun's local gravitational acceleration and, accordingly, it is a constant parameter as long as no reflectivity control devices are used (Niccolai et al. 2019). Because the SD propulsive acceleration is small when compared to the gravitational acceleration from the Sun, the value of the lightness number is assumed to be (considerably) less than one.

Taking into account the approach described in previous works (Farago and Laskar 2010; Tremaine et al. 2009; Tresaco et al. 2016), the potential functions in Eqs. (9.2) and (9.3) may be conveniently approximated to get a simplified form of the Hamiltonian function. In particular, the planet's gravitational potential ϕ_P is expanded up to the $C_{20} = -J_2$ term to obtain

$$\Phi_P = \frac{\mu_P}{r} - J_2 \frac{\mu_P R_P^2}{2r^5} \left[3(\mathbf{r} \cdot \hat{\mathbf{n}}_P)^2 - r^2 \right] \quad (9.4)$$

which corresponds to the potential of an ellipsoidal planet, where $\hat{\mathbf{n}}_P$ is the unit vector parallel to the spin axis of the planet in the direction of its north pole. A classical expansion in Legendre polynomials of $\|\mathbf{r} - \mathbf{r}_S\|^{-1}$ is used for the Sun and the solar radiation pressure potentials. Because $r \ll r_S$, the expansion can be limited to the second order Legendre polynomial (Beutler 2005). With such an assumption, the sum of the two potentials $\{\Phi_S, \Phi_T\}$ becomes

$$\Phi_S + \Phi_T = (1 - \beta)\mu_S \left\{ \frac{1}{r_S} + \frac{1}{2r_S^5} [3(\mathbf{r} \cdot \mathbf{r}_S)^2 - (r r_S)^2] \right\} - \beta \frac{\mu_S}{r_S^3} (\mathbf{r} \cdot \mathbf{r}_S) \quad (9.5)$$

Resorting to the averaging relations introduced by Kozai (1959) and Boué and Laskar (2006), it is first necessary to average the simplified Hamiltonian function over the apparent motion of the Sun around the planet (with eccentricity e_P and semi-major axis a_P), using the planet's mean anomaly as the independent variable. An average over the SD orbit is then made, assuming constant eccentricity and semi-major axis along a single orbit. Note that the averaging technique requires that the non-Keplerian forces acting on the SD must be regarded as perturbations, that is, several orders of magnitude less than the (Keplerian) gravitational force. This assumption will later be seen to have important consequences and will be better explained when the obtained orbits are analyzed using numerical simulation techniques.

By virtue of the orbit averaging, the semi-major axis is a constant of motion. It is therefore possible to obtain an averaged non-dimensional Hamiltonian H_{av} by dividing the averaged Hamiltonian by a factor $\mu_P J_2 R_P^2 / a^3$, viz.

$$H_{av} = -\frac{a^2}{2J_2 R_P^2} + \frac{1 - 3(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_P)^2}{4(1 - e^2)^{3/2}} + \frac{3}{8}\gamma(1 - \beta)\left[5(\mathbf{e} \cdot \hat{\mathbf{n}}_S)^2 + (1 - e^2)(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_S)^2 - 2e^2 + 3 - \frac{8a_P^2(1 - e_P^2)^{3/2}}{3a^2}\right] \tag{9.6}$$

where \mathbf{e} is the eccentricity vector, $\hat{\mathbf{n}}$ is the unit vector normal to the osculating orbit, and $\hat{\mathbf{n}}_S$ is the unit vector normal to the plane of the apparent motion of the Sun. In Eq. (9.6), the coefficient γ is defined as (Delsate et al. 2010)

$$\gamma = \frac{\mu_S a^5}{\mu_P a_P^3 (1 - e_P^2)^{3/2} J_2 R_P^2} \tag{9.7}$$

Because the semi-major axis is constant, and under the assumption that $\hat{\mathbf{n}}_S \equiv \hat{\mathbf{n}}_P$ (i.e., neglecting the orbital inclination of the planet), a set of dimensionless canonical Delaunay variables $\{\omega, \Omega, G, K\}$ is now introduced.

In particular, the variables ω and Ω are the argument of the pericenter and the longitude of the ascending node of the SD osculating orbit, respectively, while $G = \sqrt{1 - e^2}$ and $K = G \cos i$ are the dimensionless conjugate momenta, where i is the inclination of the orbit with respect to $\hat{\mathbf{n}}_P$. Accordingly, the Hamiltonian function in Eq. (9.6) becomes

$$\begin{aligned} \bar{H} = & \frac{1}{4G^3} - \frac{3K^2}{4G^5} + \frac{3}{8}\gamma(1 - \beta)\left[5(1 - G^2)\left(1 - \frac{K^2}{G^2}\right)\sin^2\omega - K^2 + 2G^2\right] \\ & + \frac{3}{8}\gamma(1 - \beta)\left[1 - \frac{8a_P^2(1 - e_P^2)^{3/2}}{3a^2}\right] - \frac{a^2}{2J_2 R_P^2} \end{aligned} \tag{9.8}$$

which is consistent with the results of Delsate et al. (2010) [see Eq. (9.13) of the referenced manuscript] when the lightness number is zero (i.e., without propulsive acceleration), being different only by some constraints. The latter are uninfluential because the equations of motion, obtained from the Hamiltonian in Eq. (9.8), are invariant under the addition of a constant term to the function. That means that when the lightness number is zero, all of the results by Delsate et al. (2010) are recovered. It is explicitly stated that in the following $0 \leq K \leq G \leq 1$ and all of the (uninfluential) constants will be set to zero.

9.3 Frozen Orbit Conditions

In general, due to perturbations, the altitude of a spacecraft relative to its given subsatellite point on the planet varies with time, from one orbit to the succeeding one. When the orbit is chosen such that its altitude (apart from short-period oscillations) only depends on the latitude of the subsatellite point, without any variation with time, that orbit is said to be *frozen* (Capderou 2005). Such an orbit has constant a , e , i and ω on average or, equivalently, constant a , ω , G and K .

Since the equations of variation of the SD orbital parameters are derived from the Hamiltonian in Eq. (9.8), terms a and K are constants of motion because the Hamiltonian function is independent of the mean anomaly and the longitude of ascending node. The remaining parameters ω and G are constants of motion if their combination is an equilibrium point of the Hamiltonian function. Therefore, possible frozen orbits are found as the equilibrium points of the Hamiltonian function in Eq. (9.8). Taking the first derivatives of the averaged Hamiltonian with respect to ω and G , the frozen orbits are found as the solutions of the equations

$$\frac{\partial \bar{H}}{\partial \omega} = \frac{5}{4} \gamma (1 - G^2) \left(1 - \frac{K^2}{G^2} \right) (1 - \beta) \sin \omega \cos \omega = 0 \quad (9.9)$$

$$\frac{\partial \bar{H}}{\partial G} = \frac{3}{4G^4} \left(\frac{5K^2}{G^2} - 1 \right) + \frac{3}{4} \gamma (1 - \beta) \left[5 \sin^2 \omega \left(\frac{K^2}{G^3} - G \right) + 2G \right] = 0 \quad (9.10)$$

Note that the conditions expressed by Eqs. (9.9) and (9.10) correspond to imposing that the time derivatives of ω and G are equal to zero.

The orbital stability depends on the sign definiteness of the Hessian matrix associated to the Hamiltonian function, as stated by the Lagrange-Dirichlet criterion (Bloch 2005). In particular, the frozen orbit stability is related to the following second derivatives

$$\frac{\partial^2 \bar{H}}{\partial \omega^2} = \frac{15}{4} \gamma (1 - G^2) (1 - \beta) \left(1 - \frac{K^2}{G^2} \right) (\cos^2 \omega - \sin^2 \omega) \quad (9.11)$$

$$\frac{\partial^2 \bar{H}}{\partial G^2} = \frac{3}{4G^5} \left(2 - \frac{15K^2}{G^2} \right) + \frac{3}{4} \gamma (1 - \beta) \left[2 - 5 \sin^2 \omega \left(1 + \frac{3K^2}{G^4} \right) \right] \quad (9.12)$$

$$\frac{\partial^2 \bar{H}}{\partial \omega \partial G} = \frac{15}{2} \gamma (1 - \beta) \sin \omega \cos \omega \left(\frac{K^2}{G^3} - G \right) \quad (9.13)$$

Equation (9.9) is the same as that found by Delsate et al. (2010), and its solutions are $G = 1$ (or $e = 0$), $\sin \omega \cos \omega = 0$ (or $\omega = \{0, \pm\pi, \pi\}$) and $K = G$ (or $i = \{0, \pi\}$). In the following, the solutions of the remaining Eq. (9.10) will be studied, along with its stability properties, for the cases $G = 1$ and $\sin \omega \cos \omega = 0$. The results of the analysis will be converted in terms of classical orbital parameters considering Mercury as the primary body and using the physical data of Table 9.1.

Table 9.1 Physical data of Mercury

Parameter	Value	Units	Ref
$\mu_S/\mu_{\text{Mercury}}$	6.0236×10^6		Luzum et al. (2011)
R_{Mercury}	2439.7	km	Archinal et al. (2011)
$J_{2\text{Mercury}}$	6×10^{-5}		Anderson et al. (1987)
e_{Mercury}	0.20563593		JPL
a_{Mercury}	0.38709927	au	JPL

9.3.1 Case $e = 0$: Circular Orbits

When $G = 1$, or $e = 0$, the argument of pericenter ω is not defined, and so Eq. (9.10) cannot be used. In this case, by means of the transformation

$$x = \sqrt{2(1 - G)} \cos \omega, X = \sqrt{2(1 - G)} \sin \omega \tag{9.14}$$

it is possible to eliminate such an indetermination, so that the Hamiltonian function in Eq. (9.8) becomes

$$\begin{aligned} \bar{H}_c = & \frac{\left(1 - \frac{x^2+X^2}{2}\right)^2 - 3K^2}{4\left(1 - \frac{x^2+X^2}{2}\right)^5} + \frac{3}{8}\gamma(1 - \beta) \left\{ \frac{5X^2}{2} \left(2 - \frac{x^2 + X^2}{2}\right) \times \right. \\ & \left. \times \left[1 - \frac{K^2}{\left(1 - \frac{x^2+X^2}{2}\right)^2} \right] - K^2 + 2 \left(1 - \frac{x^2 + X^2}{2}\right)^2 \right\} \end{aligned} \tag{9.15}$$

The Hamiltonian in Eq. (9.15) has an equilibrium point for $x = 0$ and $X = 0$ (that is, for $G = 1$ or $e = 0$), independent of the value of K . Hence, circular orbits are always frozen for all inclinations. However, by means of the Lagrange-Dirichlet criterion, it can be shown that they are stable only when

$$\cos^2 i < \frac{1 - 2\gamma(1 - \beta)}{5} \text{ or } \cos^2 i > \frac{1 + 3\gamma(1 - \beta)}{5[1 + \gamma(1 - \beta)]} \tag{9.16}$$

In this context, Fig. 9.3 shows the stability regions in terms of orbital parameters obtained from Eq. (9.16) for different values of the lightness number in a Mercury-based mission scenario. The gray regions in Fig. 9.3 represent the pairs (altitude–inclination) for which the orbit is stable when the lightness number is zero. Note that the presence of the solar radiation pressure induces a stretching of the stability regions, thus giving the possibility of obtaining stable frozen orbits around Mercury at higher altitudes for a given orbital inclination.

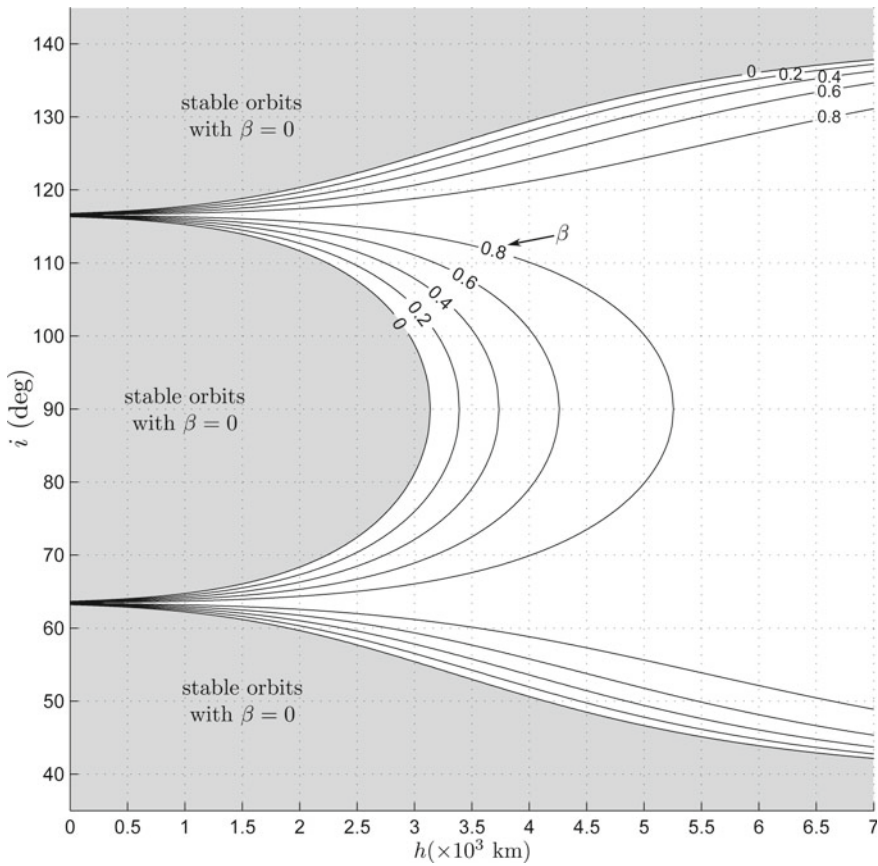


Fig. 9.3 Stable circular frozen orbits for Mercury as a function of the inclination i and the altitude h

9.3.2 Case $\omega = 0$ or $\omega = \pi$

Keeping in mind that $K = G \cos i$, the condition to satisfy Eq. (9.10) with $\omega = 0$ or $\omega = \pi$ is

$$\cos^2 i = \frac{1 - 2\gamma(1 - \beta)G^5}{5} \tag{9.17}$$

Note that, being $G > 0$ and $\beta < 1$, the condition given by Eq. (9.17) can be conservatively approximated by $\cos^2 i < 1/5$. Consequently, when $\omega = 0$ or $\omega = \pi$, frozen orbits exist only when $63.435^\circ < i < 116.565^\circ$.

Substituting Eq. (9.17) with the condition $\sin \omega = 0$ into Eqs. (9.11), (9.12) and (9.13), the following stability condition is recovered:

$$G > \sqrt[5]{\frac{1}{7\gamma(1-\beta)}} \tag{9.18}$$

Indeed, the derivative in Eq. (9.13) is trivially zero and the derivative in Eq. (9.11) is always positive, then for orbit stability the derivative in Eq. (9.12) is required to be positive too. From Eqs. (9.17) and (9.18), a further limitation of the orbital inclination is obtained, whereby it is found that the frozen orbits are stable when $\cos^2 i < 1/7$, corresponding to $67.792^\circ < i < 112.208^\circ$.

Figure 9.4 gives the possible frozen orbits for Mercury with $\omega = 0$ or $\omega = \pi$ when the eccentricity is 0.1 or 0.3. Each curve corresponds to a fixed value of the sail lightness number and gives all the frozen orbits obtainable with that fixed value of the lightness number when ω is either 0 or π . Note that unstable orbits are also represented according to Eq. (9.18). The SD propulsive acceleration permits the natural frozen orbits ($\beta = 0$) to be displaced, thus giving more freedom in choosing the altitude of pericenter when the remaining orbital parameters are fixed.

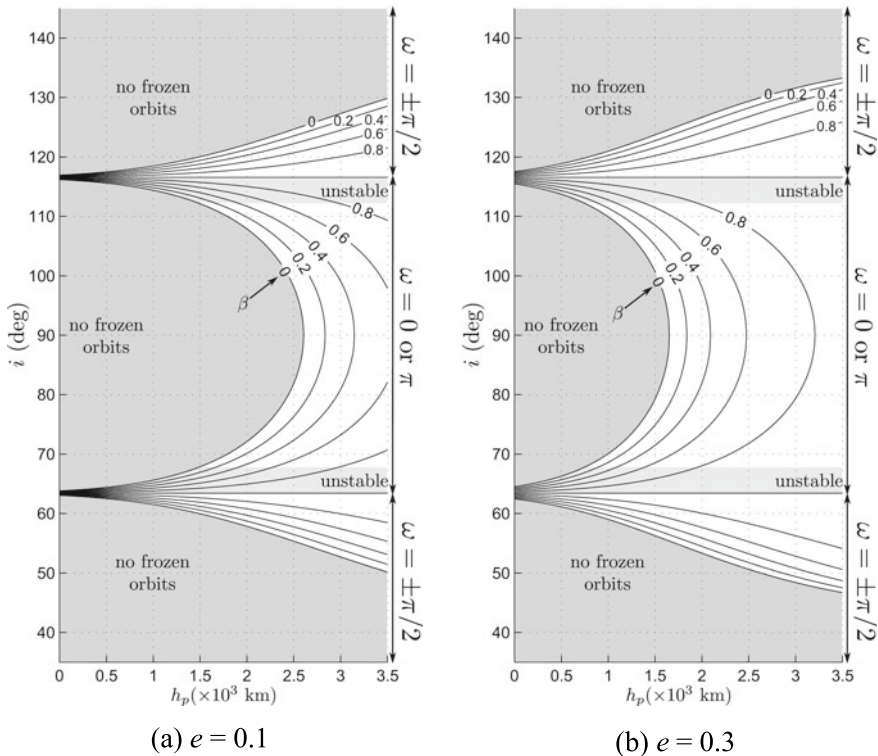


Fig. 9.4 Elliptical frozen orbits for Mercury as a function of the inclination i and the pericenter altitude h_p with eccentricity 0.1 and 0.3

9.3.3 Case $\omega = \pm\pi/2$

In this case, Eq. (9.10) gives

$$\cos^2 i = \frac{1 + 3\gamma(1 - \beta)G^5}{5[1 + \gamma(1 - \beta)G^3]} \quad (9.19)$$

It can be shown that, for realistic values of the lightness number (less than 1), the quantity $\cos^2 i$ is an increasing function of γ as long as $G > \sqrt{1/3}$ or, equivalently, if $e < \sqrt{2/3} \approx 0.8165$.

Assuming the eccentricity to be less than 0.8165, the limitation $1/5 < \cos^2 i < (3G^2/5)$ is found. Therefore, in this case, frozen orbits exist only when $i < 63.435^\circ$ or $i > 116.565^\circ$. As far as stability is concerned, the derivative in Eq. (9.13) is again trivially zero and the derivative in Eq. (9.11) is always negative for $\cos^2 i < (3G^2/5) < 3/5$. Therefore, the derivative in Eq. (9.12) is required to be negative.

It can be shown that this is verified if

$$2 + 21\gamma(1 - \beta)G^5 - 15\gamma(1 - \beta)G^3 \cos^2 i > 0 \quad (9.20)$$

which is always met when $\cos^2 i < (3G^2/5)$. If the eccentricity is greater than 0.8165, a frozen orbit exists only if $\cos^2 i < 1/5$, or equivalently, when $63.435^\circ < i < 116.565^\circ$ and the stability condition given by Eq. (9.20) has to be checked on a case-by-case basis. In this context, Fig. 9.4 shows the frozen orbits when $\omega = \pm\pi/2$ and the eccentricity is 0.1 or 0.3 as a function of the lightness number. Again, the natural frozen orbits are displaced towards higher pericenter altitude for a fixed value of the inclination.

9.4 Frozen Orbit Period

The Hamiltonian in Eq. (9.8), considered as a function of the two variables ω and G , can be approximated in the neighborhood of an equilibrium point using a Taylor expansion up to the second order, viz.

$$\bar{H} = \bar{H}_{\text{eq}} + \frac{1}{2} \left. \frac{\partial^2 \bar{H}}{\partial \omega^2} \right|_{\text{eq}} (\omega - \omega_{\text{eq}})^2 + \frac{1}{2} \left. \frac{\partial^2 \bar{H}}{\partial G^2} \right|_{\text{eq}} (G - G_{\text{eq}})^2 \quad (9.21)$$

where the subscript “eq” denotes a value calculated at the equilibrium point, that is, along a frozen orbit. Note that in Eq. (9.21) the first derivatives with respect to G and ω and the mixed second derivatives are omitted. In fact, they are zero for the problem at hand since Eqs. (9.9), (9.10), (9.11), (9.12) and (9.13) hold.

When a stable frozen orbit is considered, Eq. (9.21) is the Hamiltonian of a harmonic oscillator with a (dimensionless) period τ given by

$$\tau = \frac{2\pi}{\sqrt{\left. \frac{\partial^2 \bar{H}}{\partial \omega^2} \right|_{\text{eq}} \left. \frac{\partial^2 \bar{H}}{\partial G^2} \right|_{\text{eq}}}} \tag{9.22}$$

A similar expression involving the second derivatives with respect to x and X is obtained when the Hamiltonian of Eq. (9.15) is expanded in the circular case. The corresponding dimensional period can be written as

$$T = \frac{\tau a^{7/2}}{J_2 R_p^2 \sqrt{\mu_P}} \tag{9.23}$$

recalling that the averaged Hamiltonian is made non-dimensional using $\mu_P J_2 R_p^2 / a^3$.

Note that Eq. (9.23) represents the period of oscillation of a point around an equilibrium of the Hamiltonian in the $(\omega - G)$ plane or, equivalently, the period of oscillation of the eccentricity vector of a frozen orbit.

9.5 Numerical Validation

The analytical results obtained by means of the averaging technique described in the previous sections are based on the assumption that non-Keplerian forces acting on the SD can be considered as perturbations with respect to the gravitational force of the planet. In other words, the instantaneous effect of non-Keplerian forces can be neglected, and the mean motion of the spacecraft is well approximated by considering the average effect over an orbit and over the apparent motion of the Sun.

Numerical simulations of the behavior of the frozen orbits are used to investigate when the analytical model is adequate. In the following, the non-averaged equations of motion of the SD are integrated with the aim of understanding the limitations of the obtained results.

In this context, numerical simulations are performed with reference to the model of the system illustrated in Fig. 9.2. This simplified model takes into account the accelerations due to the gravitational force of the planet (including the J_2 term) and the Sun (whose apparent motion is assumed to be elliptical and to lie in the equatorial plane of the planet) and the acceleration \mathbf{a}_T provided by the solar radiation pressure acting on the SD surface. Mercury is used as the reference planet in the simulations, so that a non-negligible effect of the planet eccentricity is also included; see Table 9.1. Note that the simplified model neglects the shadowing effect of the planet, the effect of the obliquity angle of the planetary orbit plane with respect to the equatorial plane, and the effect of all the terms of Eq. (9.2) apart $\{C_{00}, C_{20}\}$.

The Gauss form of the equations of motion in terms of modified equinoctial elements (Hintz 2008; Walker et al. 1985; Walker 1986) has been integrated, using a variable-order Adams–Bashforth–Moulton PECE solver with absolute and relative tolerances of 10^{-13} and 10^{-9} , respectively. The initial state of the spacecraft,

Table 9.2 Initial orbital parameters for numerical simulations

a_0 (km)	e_0	i_0°	ω_0°	β	T (years)
6439.86	0.5	90	0	0	29.58
6465.93	0.5	90	0	0.02	30.00
6479.21	0.5	90	0	0.03	30.21

given in terms of modified equinoctial elements of the osculating orbit, is obtained starting from the Keplerian orbital elements for a SD on a frozen orbit. Four of those orbital elements (a_0 , e_0 , i_0 , ω_0) are calculated using the analytical model discussed previously. The true anomaly ν_0 and the ascending node Ω_0 can be freely chosen, as well as the initial mean anomaly of the Sun M_{S0} . Indeed, they are not defined by the analytical model as they disappear during the averaging process. When not differently stated, they are all set equal to zero, without loss of generality. Note that the frame with respect to which the orbits are described is the planet equatorial frame, with the z -axis along the spin axis of the planet, the x -axis in the equatorial plane and pointing toward the pericenter of the apparent motion of the Sun, and the y -axis forming a right-handed frame; see Fig. 9.2.

For comparison purposes, the simulations are run for frozen orbits with given values of $\{e_0, i_0, \omega_0\}$ and different values of the SD lightness number β , to which correspond different values of the semi-major axes a_0 . The data for the simulated frozen orbits are reported in Table 9.2 together with the period of oscillation of the eccentricity vector from Eq. (9.23). Note that the first simulation neglects the effects of the solar radiation pressure, whereas in the other two cases, a set of values of the lightness number have been considered (Mengali et al., 2018), which are compatible with a SD having a low-medium performance level.

The simulation results are plotted in Figs. 9.5, 9.6 and 9.7. Other simulations performed with different set of initial parameters have provided similar results and are not reported for the sake of conciseness.

9.6 Discussion of the Results

For each of the simulations of Figs. 9.5, 9.6 and 9.7, it is possible to recognize the superimposition of three oscillations at different frequencies. The highest frequency of oscillation is on the order of the inverse of the orbital period and corresponds to the variation of orbital elements due to the variation of the perturbing force experienced by the spacecraft during one orbit. The medium frequency of oscillation, whose period is on the order of the revolution period of Mercury, is an effect of the motion of the orbit together with Mercury around the Sun. The lowest oscillation frequency corresponds to the oscillation of a harmonic oscillator around its equilibrium point; see Eq. (9.21).

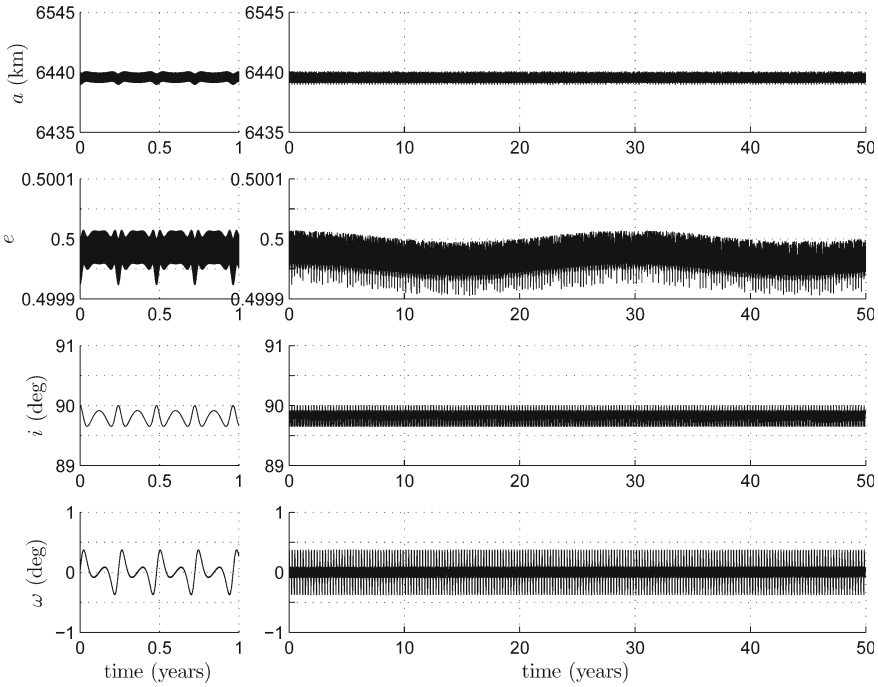


Fig. 9.5 Simulation for a frozen orbit around Mercury with $\beta = 0$

Figure 9.5 shows the simulation of a natural frozen orbit ($\beta = 0$). It presents small oscillations of the orbital parameters, with a low frequency component having a period of about 30 years, in agreement with Eq. (9.23). When the lightness number is increased to $\beta = 0.02$ while maintaining the same value of eccentricity and inclination, the semi-major axis of the new frozen orbit tends to increase. Figure 9.6 shows the simulation for that new case. Even though the average value of the orbital elements is constant, the amplitude of the oscillations is now large and eventually results in the spacecraft crashing on Mercury’s surface. Furthermore, the period of oscillation is about 50 years with an error of 20 years with respect to the value of Table 9.2. Therefore, Eq. (9.23) is inadequate for this case, implying that the theoretical model based on the averaging technique is not able to correctly describe the average behavior of the orbit.

For a greater lightness number ($\beta = 0.03$), the argument of pericenter and the eccentricity are continuously increased (see Fig. 9.7). Therefore, the actual orbit is not frozen, and the analytical model fails. In general, a similar trend characterizes all of the orbits with fixed $\{e_0, i_0, \omega_0\}$ as long as β is increased. Typically, the analytical model fails to predict the right behavior of an orbit when values of β of order 0.01–0.02 are considered, thus suggesting that frozen orbits cannot be found using the proposed method for values of the lightness number beyond those limits.

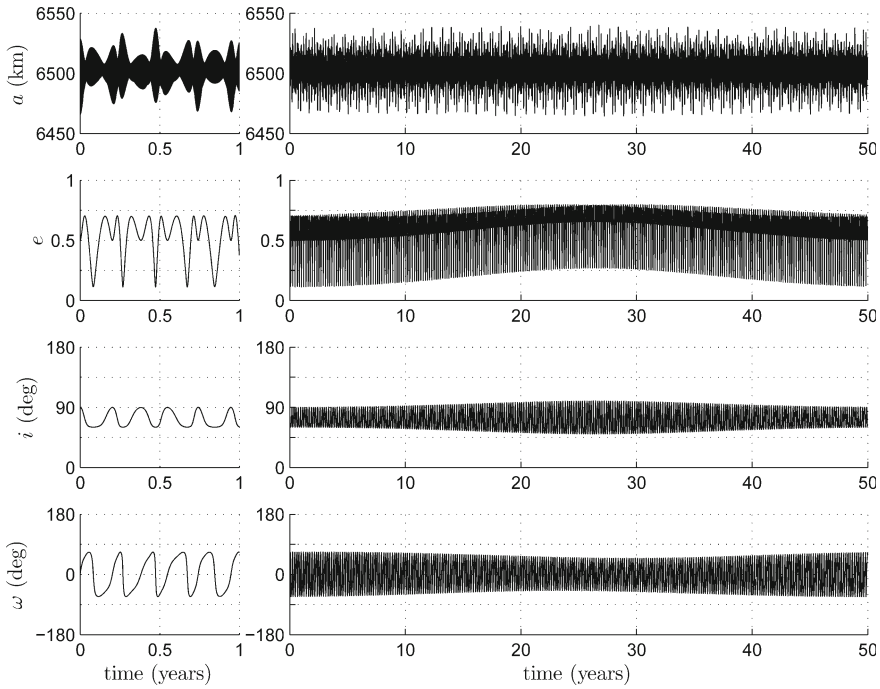


Fig. 9.6 Simulation for a frozen orbit around Mercury with $\beta = 0.02$

The main reason for this discrepancy between the analytical model and the numerical simulations seems to be that the mean behavior predicted by the averaging analysis cannot be considered as a good approximation of the real behavior when the SD lightness number becomes greater than a value beyond which the effect of the solar radiation pressure cannot be considered as a simple perturbation effect. In this regard, Tresaco et al. (2018) show that such a value should be about $\beta \simeq 0.01 - 0.02$ at least one order of magnitude greater than $\beta \simeq 0.001$, corresponding to the lightness number of the IKAROS mission. Furthermore, the analytical model does not take into account the effects of the initial position of the planet with respect to the Sun, which is lost due to the orbit averaging. The simulations show that significant effects can be associated with the initial condition in some cases. Figure 9.8 shows the evolution of an orbit when two different initial positions of Mercury with respect to the Sun are set. With Mercury starting at the perihelion (Fig. 9.8a), the orbit is not frozen. By setting the initial position of Mercury at the aphelion, where the initial propulsive acceleration is minimized, the orbit turns out to be frozen (Fig. 9.8b). Other notable effects involve the different amplitudes of oscillation and the different starting slopes in the time evolution of the orbital parameters, which result in a shift of the mean values of the parameters with respect to the desired ones.

To sum up, for values of the lightness number on the order of 0.01–0.02, corresponding to SDs of low-medium performance (or to the current state of the art for

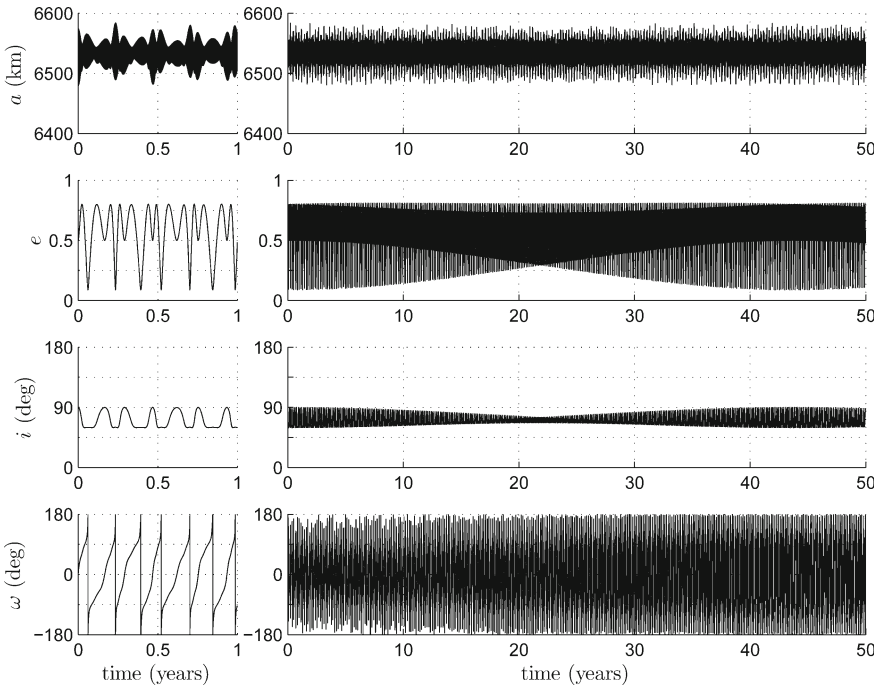


Fig. 9.7 Simulation for a frozen orbit around Mercury with $\beta = 0.03$

a solar sail), the numerical results show some discrepancies with respect to those predicted by the analytical method (as expected from previous considerations), and the orbits obtained are not exactly frozen. Such differences increase with β and could still be compatible with mission requirements. Instead, when $\beta > 0.02$, corresponding to a high-performance SD (or a medium- or far-term solar sail), the analytical model fails and the families of frozen orbits generated do not have a practical application.

Ultimately, feasible long-term frozen orbits can only be obtained using the analytical method discussed for small lightness numbers (on the order of $\beta \leq 0.001$), implying that the possible choices in terms of orbital parameters can be only slightly extended when the solar radiation pressure is used as a propulsive acceleration source. Therefore, the use of a spacecraft equipped with a conventional solar sail is discouraged, since the advantages obtained thanks to the photonic propulsion do not justify the increased complexity in the spacecraft design. On the other hand, such values of lightness number are comparable with those of femto-satellites that have already been manufactured or even launched (Niccolai et al. 2019). A constellation of such femto-satellites used as low-performance SDs and placed on planetocentric frozen orbits could combine the advantages of a distributed-payload mission (due to the large number of SDs that could be carried by a mother spacecraft) with a slight increase in orbital parameter combinations for frozen orbits (due to the fact that SDs could exploit their high area-to-mass ratio to generate otherwise unfeasible frozen

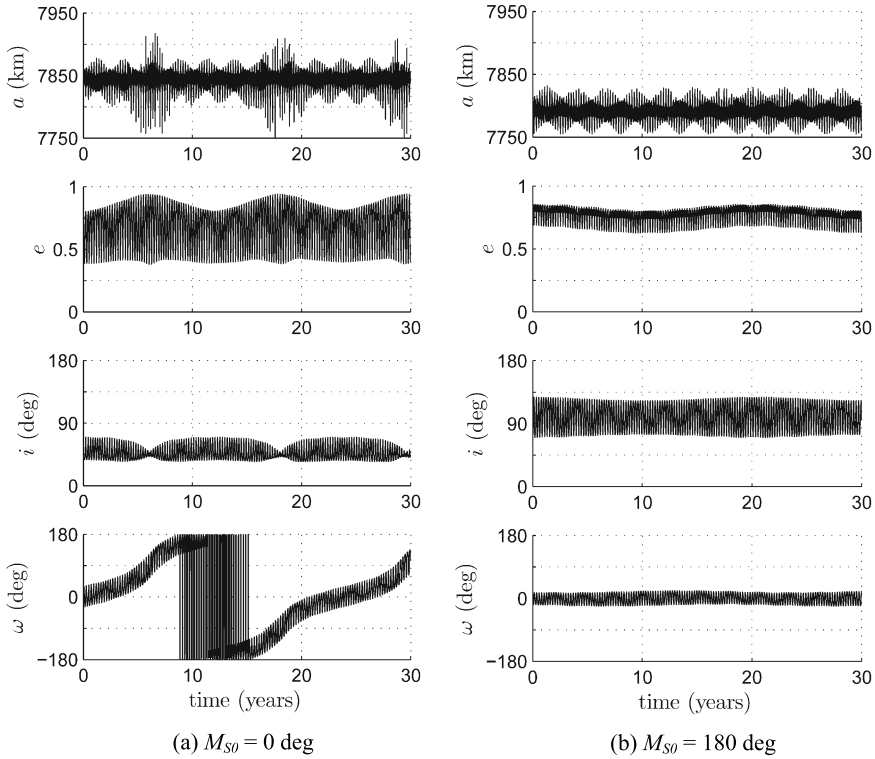


Fig. 9.8 Effect of the initial conditions on the orbital parameters for an orbit defined by $a_0 = 7811.99$ km, $e_0 = 0.8$, $i_0 = 70^\circ$, $\omega_0 = 0^\circ$

orbits). Moreover, the requirement of low lightness number could be a technological advantage, since it allows the area-to-mass ratio to be reduced, so that the payload mass can be increased.

9.7 Conclusions

An analytical model based on Hamiltonian averaging technique has been used to derive the conditions for tracking planetocentric frozen orbits by means of a Sun-facing smart dust. The analysis takes into account the planet’s oblateness, the perturbation due to Sun’s gravity, and the solar radiation pressure. The stability of frozen orbits is investigated, in order to understand whether they are suitable for mission scenarios involving planetary observations.

To validate the theoretical results, high-precision numerical simulations of frozen orbits have been performed for different values of the smart dust lightness number.

Simulations have revealed some differences between the numerical results and the results expected on the basis of a theoretical model.

Generally, in contrast with the analytical results, no frozen orbit has been found when the lightness number becomes larger than a value on the order of 0.02. Moreover, for increasing values of the lightness number, both short- and long-period oscillations of the orbital parameters show growing amplitudes, resulting in a possible crash of the smart dust on the planet surface.

Although the requirement of a small lightness number could be a strong limitation for a classical solar sail, a constellation of smart dusts could represent a promising option for mission scenarios consisting in the scientific observation of Mercury by means of a distributed payload. Indeed, the required performance level is small and a smart dust is capable of passively maintaining a Sun-pointing attitude, making the design complexity reasonable. On the other hand, the slight increase in terms of orbital parameter flexibility could be useful for a planetary observation mission with a constellation of smart dusts. Therefore, the analytical tool discussed in this work could be useful for a preliminary choice of the initial orbital parameters to obtain frozen orbits for a low- or medium-performance smart dust.

Acknowledgements This work is supported by the University of Pisa, Progetti di Ricerca di Ateneo (Grant no. PRA_2018_44).

List of Acronyms

SD Smart Dust

References

- Abad, A., A. Elife, and E. Tresaco. 2009. Analytical model to find frozen orbits for a lunar orbiter. *Journal of Guidance, Control and Dynamics* 32(3):888–898. <https://doi.org/10.2514/1.38350>.
- Alessi, E.M., C. Colombo, and A. Rossi. 2019. Phase space description of the dynamics due to the coupled effect of the planetary oblateness and the solar radiation pressure perturbations. *Celestial Mechanics and Dynamical Astronomy* 131(9). <https://doi.org/10.1007/s10569-019-9919-z>.
- Aliasi, G., G. Mengali, and A.A. Quarta. 2011. Artificial equilibrium points for a generalized sail in the circular restricted three-body problem. *Celestial Mechanics and Dynamical Astronomy* 110(4):343–368. <https://doi.org/10.1007/s10569-011-9366-y>.
- Anderson, P., and M. Macdonald. 2013. Extension of highly elliptical earth orbits using continuous low-thrust propulsion. *Journal of Guidance, Control, and Dynamics* 36(1):282–292. <https://doi.org/10.2514/1.55304>.
- Anderson, J.D., et al. 1987. The mass, gravity field, and ephemeris of mercury. *Icarus* 71(3):337–349. [https://doi.org/10.1016/0019-1035\(87\)90033-9](https://doi.org/10.1016/0019-1035(87)90033-9).
- Archinal, B.A., et al. 2011. Report of the IAU working group on cartographic coordinates and rotational elements: 2009. *Celestial Mechanics and Dynamical Astronomy* 109(2):101–135. <https://doi.org/10.1007/s10569-010-9320-4>.

- Atchison, J.A., and M.A. Peck. 2010. A passive, Sun-pointing, millimeter-scale solar sail. *Acta Astronautica* 67(1–2):108–121. <https://doi.org/10.1016/j.actaastro.2009.12.008>.
- Barnhart, D.J., T. Vladimirova, and M.N. Sweeting. 2007. System-on-a-chip design of self-powered wireless sensor nodes for hostile environments. In *IEEE aerospace conference*. Big Sky, MT, USA. <https://doi.org/10.1109/AERO.2007.352640>.
- Beutler, G. 2005. Methods of celestial mechanics. In: *Astronomy and Astrophysics Library*, vol. 1 ISBN: 3–540–40749–9. Springer-Verlag. Chap. 3, p. 54.
- Bloch, A. M. (2005). Nonholonomic mechanics and control. In: *Interdisciplinary Applied Mathematics*, Chap. 3, p. 152. Springer-Verlag, ISBN: 0–387–95535–6.
- Boué, G., J. Laskar. 2006. Precession of a planet with a satellite. *Icarus* 185(2):312–330. <https://doi.org/10.1016/j.icarus.2006.07.019>.
- Capderou, M. 2005. *Satellites orbit and missions*. Springer-Verlag, ISBN: 2–287–21317–1.
- Carbone, A., et al. 2020. Observing Mercury by a quasi-propellantless mission”. In: *Celestial Mechanics and Dynamical Astronomy* 132.1. DOI: <https://doi.org/10.1007/s10569-020-9950-0>.
- Chobotov, V.A. 2002. *Orbital mechanics*. Ed. by AIAA. ISBN: 978–1563475375. <https://doi.org/10.2514/4.862250>.
- Circi, C., et al. 2019. Global mapping of asteroids by frozen orbits: The case of 216 Kleopatra. *Acta Astronautica* 161: 101–107. <https://doi.org/10.1016/j.actaastro.2019.05.026>.
- Coffey, S.L., A. Deprit, and E. Deprit. 1994. Frozen orbits for satellites close to an earth-like planet. In *Celestial Mechanics and Dynamical Astronomy*, 59.1. <https://doi.org/10.1007/BF00691970>.
- Colombo, C., and C.R. McInnes. 2012. Orbit design for future SpaceChip swarm missions in a planetary atmosphere. *Acta Astronautica* 75: 25–41. <https://doi.org/10.1016/j.actaastro.2012.01.004>.
- Colombo, C. 2020. Long-term evolution of highly-elliptical orbits: Luni-solar perturbation effects for stability and re-entry. *Frontiers in Astronomy and Space Sciences* 6. <https://doi.org/10.3389/fspas.2019.00034>.
- Delsate, N., et al. 2010. Frozen orbits at high eccentricity and inclination: Application to mercury orbiter. *Celestial Mechanics and Dynamical Astronomy* 108(3):275–300. <https://doi.org/10.1007/s10569-010-9306-2>.
- Farago, F., and J. Laskar. 2010. High-inclination orbits in the secular quadrupolar three-body problem. *Monthly Notices of the Royal Astronomical Society* 401(2):1189–1198. <https://doi.org/10.1111/j.1365-2966.2009.15711.x>.
- Farrés, A., A. Jorba. 2008. A dynamical system approach for the station keeping of a solar sail. *Journal of the Astronautical Sciences* 56(2). <https://doi.org/10.1007/BF03256549>.
- Gong, S., J. Li. 2014. Solar sail heliocentric elliptic displaced orbits. *Journal of Guidance, Control, and Dynamics* 37(6). <https://doi.org/10.2514/1.G000660>.
- Gong, S., et al. 2012. A new solar sail orbit. *Science China Technological Sciences* 55(3). <https://doi.org/10.1007/s11431-011-4691-7>.
- Hintz, G.R. 2008. Survey of orbit element sets. *Journal of Guidance Control and Dynamics* 31(3):785–790. <https://doi.org/10.2514/1.32237>.
- Khattab, E.H., M. Radwan, and W.A. Rahoma. 2020. Frozen orbits construction for a lunar solar sail. *Journal of Astronomy and Sciences* 37(1):1–9. <https://doi.org/10.5140/JASS.2020.37.1.1>.
- Kozai, Y. 1959. The motion of a close earth satellite. *Astronomical Journal* 64(1274):367–377. <https://doi.org/10.1086/107957>.
- Lara, M., and R.P. Russel. 2008. Fast design of repeat ground track orbits in high-fidelity geopotentials. *Journal of the Astronautical Sciences* 56(3). <https://doi.org/10.1007/BF03256555>.
- Lara, M., A. Deprit, and A. Elipe. 1995. Numerical Continuation of families of frozen orbits in the zonal problem of artificial satellite theory. *Celestial Mechanics and Dynamical Astronomy* 62(2):167–181. <https://doi.org/10.1007/BF00692085>.
- Lara, M., et al. 2010. Analytical theory for spacecraft motion about Mercury. *Acta Astronautica* 66(7–8):1022–1038. <https://doi.org/10.1016/j.actaastro.2009.10.011>.

- Leipold, M.E., and O. Wagner. 1996. Mercury sun-synchronous polar orbits using solar sail propulsion. *Journal of Guidance Control and Dynamics* 19(6):1337–1341. <https://doi.org/10.2514/3.21791>.
- Lücking, C., C. Colombo, and C.R. McInnes. 2012. Electrochromic orbit control for smart-dust devices. *Journal of Guidance, Control, and Dynamics* 35(5):1548–1558. <https://doi.org/10.2514/1.55488>.
- Luzum, B., et al. 2011. The IAU 2009 system of astronomical constants: The report of the IAU working group on numerical standards for fundamental astronomy. *Celestial Mechanics and Dynamical Astronomy* 110(4):293–304. <https://doi.org/10.1007/s10569-011-9352-4>.
- Macdonald, M., et al. 2010. Extension of the sun-synchronous orbit. *Journal of Guidance Control and Dynamics* 33(6):1935–1939. <https://doi.org/10.2514/1.49011>.
- Masoud, A., et al. 2018. Construction of frozen orbits using continuous thrust control theories considering earth oblateness and solar radiation pressure perturbations. *Journal of the Astronautical Sciences* 65(4). <https://doi.org/10.1007/s40295-018-0135-y>.
- McInnes, C.R. 1999. *Solar sailing: technology, dynamics and mission applications*. Springer-Verlag. ISBN: 978–3540210627. <https://doi.org/10.1007/978-1-4471-3992-8>.
- Mengali, G., A.A. Quarta, and E. Denti. 2018. Relative motion of Sun-pointing smart dust in circular heliocentric orbits. *Journal of Guidance, Control, and Dynamics* 41(4). <https://doi.org/10.2514/1.G003088>.
- Niccolai, L., et al. 2019. A review of smart dust architecture, dynamics, and mission applications. *Progress in Aerospace Sciences* 106. <https://doi.org/10.1016/j.paerosci.2019.01.003>.
- Niccolai, L., et al. 2020. Feedback control law of solar sail with variable surface reflectivity at Sun-Earth collinear equilibrium points. *Aerospace Science and Technology* 106. <https://doi.org/10.1016/j.ast.2020.106144>.
- Park, S.-Y., and J.L. Junkins. 1995. Orbital mission analysis for a lunar mapping satellite. *Journal of the Astronautical Sciences* 43(2):207–217. <https://doi.org/10.2514/6.1994-3717>.
- Paskowitz, M.E., D.J. Scheeres. 2006. Design of science orbits about planetary satellites: application to Europa. *Journal of Guidance Control and Dynamics* 29(5):1147–1158. <https://doi.org/10.2514/1.36220>.
- Quarta, A.A., G. Mengali, and M. Bassetto. 2019a. Optimal solar sail transfers to circular Earth-synchronous displaced orbits. *Astrodynamics* 4(3). <https://doi.org/10.1007/s42064-019-0057-x>.
- Quarta, A.A., G. Mengali, and E. Denti. 2019b. Optimal in-orbit repositioning of Sun-pointing smart dust. *Acta Astronautica* 154. <https://doi.org/10.1016/j.actaastro.2018.03.036>.
- Quarta, A.A., G. Mengali, and L. Niccolai. 2019c. Smart dust option for geomagnetic tail exploration. *Astrodynamics* 3(3). <https://doi.org/10.1007/s42064-019-0048-3>.
- Quarta, Alessandro A., Giovanni Mengali, and Lorenzo Niccolai. 2020. Solar sail optimal transfer between heliostationary points. *Journal of Guidance, Control, and Dynamics* 43: 1935–1942. <https://doi.org/10.2514/1.G005193>
- Russell, R.P. 2012. Survey of spacecraft trajectory design in strongly perturbed environments. *Journal of Guidance, Control, and Dynamics* 35(3):705–720. <https://doi.org/10.2514/1.56813>.
- San-Juan, J., et al. 2009. Analytical model for lunar orbiter revisited. In *Workshop on Computer Algebra and Differential Equations*, vol. 130(pt 2):1281–1299.
- Scheeres, D.J., M.D. Guman, and B.F. Villac. 2001. Stability analysis of planetary satellite orbiters: application to the Europa orbiter. *Journal of Guidance Control and Dynamics* 24(4):778–787. <https://doi.org/10.2514/2.4778>.
- Scheeres, D.J. 2012. Orbit mechanics about asteroids and comets. *Journal of Guidance, Control, and Dynamics* 35(3):987–997. <https://doi.org/10.2514/1.57247>.
- Sood, R., K. Howell. 2019. Solar sail transfers and trajectory design to sun-earth L_4 , L_5 : Solar observations and potential earth trojan exploration. *Journal of the Astronautical Sciences* 66(3). <https://doi.org/10.1007/s40295-018-00141-4>.
- Tremaine, S., J. Touma, and F. Namouni. 2009. Satellite dynamics on the laplace surface. *The Astronomical Journal* 137(3):3706–3717. <https://doi.org/10.1088/0004-6256/137/3/3706>.

- Tresaco, E., A. Elipe, and J.P.S. Carvalho. 2016. Frozen orbits for a solar sail around Mercury. *Journal of Guidance, Control, and Dynamics* 39(7). <https://doi.org/10.2514/1.G001510>.
- Tresaco, E., et al. 2018. Averaged model to study long-term dynamics of a probe about Mercury. *Celestial Mechanics and Dynamical Astronomy* 130(2). <https://doi.org/10.1007/s10569-017-9801-9>.
- Walker, M.J.H., J. Owens, and B. Ireland. 1985. A set of modified equinoctial orbit elements. *Celestial Mechanics* 36(4):409–419. <https://doi.org/10.1007/BF01227493>.
- Walker, M.J.H. 1986. Errata: A set of modified equinoctial orbit elements. *Celestial Mechanics* 38(4):391–392. <https://doi.org/10.1007/BF01238929>.

Part III
The Moon, a Steppingstone to Planetary
ISRU

Chapter 10

Simulants in In-Situ Resource Utilization Technology Development



Hunter Williams

Abstract Regolith, the surface material from terrestrial planetary bodies such as the Moon, Mars, and asteroids, is extremely expensive and rarely available for scientists and engineers to use when developing new technology. Different technologies work better in different materials, so since failure is not an option in space missions it is necessary to understand how a technology will behave in the materials it may encounter in space. To address this, planetary surface simulants have been developed to offer a low-cost alternative to planetary materials. They are made to reflect either physical, mineralogical, or chemical properties of lunar, Martian, and asteroidal rocks and regolith. A wide variety of regolith and volatile simulants have been used since the Apollo program that have differed based on the planetary body of interest, available material, and physical attribute tested. In-situ resource utilization (ISRU) technology development requires simulant with geotechnical attributes unlike those required before. Interaction between regolith simulant and the test environment has also proven more important for ISRU technology development than for purely mechanical technology for exploration. This chapter examines the background of regolith simulant development for a variety of planetary bodies and the efficacy of the ways they have been used.

Regolith, the surface material from terrestrial planetary bodies such as the Moon, Mars, and asteroids, is extremely expensive and rarely available for scientists and engineers to use when developing new technology. Different technologies work better in different materials, so when failure is not an option in space missions it is necessary to understand how a technology will behave in the materials it may encounter in space. To address this, planetary surface simulants have been developed to offer a low-cost, readily available alternative to planetary materials; they are made to reflect either physical, mineralogical, or chemical properties of lunar, Martian, and asteroidal rocks and regolith (Taylor 2016). A wide variety of regolith and volatile simulants have been used since the Apollo program, depending on the planetary body

H. Williams (✉)
Honeybee Robotics, 2408 Lincoln Ave, Altadena, CA 91001, USA
e-mail: hjwilliams@honeybeerobotics.com

of interest, available material, and physical attribute tested. In-situ resource utilization (ISRU) technology development requires simulant with geotechnical attributes unlike those required before. Interaction between regolith simulant and the test environment has also proven more important for ISRU technology development than for purely mechanical technology for exploration. This chapter examines the background of regolith simulant development for a variety of planetary bodies and the efficacy of the ways they have been used.

10.1 Introduction

Regolith is an unconsolidated, granular, non-organic material that covers terrestrial planetary bodies and is made of material that has been weathered away from the underlying bedrock by meteorite impacts, winds, or water (McKay et al. 1991). Regolith simulant is a material crafted to simulate important properties of planetary regolith for testing technologies for applications on that planetary body. Different simulants replicate different properties, including chemical, geological, geotechnical, optical, volatile content, behavior under low gravity, and many others. Some simulants grossly approximate several properties by starting from a similar geology, while other simulants approximate single properties more closely or at a lower cost. To describe in detail these properties for even one simulant would be beyond the scope of this chapter—doctoral dissertations have been written on geophysical property measurement—but this chapter will cover *how* to approach the simulant selection process based on those important properties.

There are several additional key factors in testing technology made for use in space, including the chemical content and pressure of the gaseous environment, the real or simulated gravity, the electromagnetic and energetic environment, and the simulated regolith's interaction with any of these. Depending on the type of technology, its sensitivity to any one factor and to the level of interaction between the factors will determine the fidelity required for testing. Unless a test is conducted on returned native regolith samples in a deep vacuum chamber with a simulated space electromagnetic environment on a parabolic flight, some aspect of the test will be a compromise.

It is therefore important to note two points:

1. All testing requires compromise in one factor or another due to cost and availability.
2. Paying attention to how a simulant and test environment combination reproduces what the technology will encounter in space is important for accurate results.

A key concept for making new technology for unknown circumstances and avoiding the extreme cost of testing at the highest fidelity while still testing in relevant conditions is called “bracketing. Bracketing is testing under a variety of potential conditions including different simulants and environments to determine how a technology

works under edge cases when it is not possible to test under the exact conditions faced in space. Failure to bracket by testing with only a single simulant has contributed to several mission failures. The most common mistake made when using regolith simulant is overestimating the use cases of a specific simulant. NASA's Insight Mars Lander burrowing heat probe could have been redesigned to avoid getting stuck had it been tested in several regolith simulants to bracket its potential use cases. Example simulants could include one that is easy to get stuck in, one that is easy to penetrate like a loose granular sand, one that is the most geophysically accurate, a high cohesion simulant, a high water-content simulant, and a worst-case hard compacted volcanic basalt. Even if getting stuck was unavoidable, it would have been possible to measure the geophysical properties of what it encountered and determine the closest analog from its test matrix of simulants. This could have shortened the time it took to develop an alternative operational concept to keep it burrowing without getting stuck.

Engineers and scientists often bracket the extremes of possible conditions that a new technology will face (Atkinson and Zancy 2018). Examples of this are testing under both atmospheric and vacuum conditions, testing with low- and high-cohesion simulant, and comparing tests performed on the ground or in a reduced gravity environment. Figure 10.1 shows bracketing tests in action: the technology is tested for utility in extremely dense, rocky simulant as well as low-density simulant to judge performance with various geotechnical properties of the regolith and to estimate performance in the lower gravity of the Moon.

There are many available regolith simulants and each of them simulates the planetary body's regolith in a different way for different reasons. From large rocky planets to small asteroids, every planetary body has variation in composition and geotechnical properties in different areas, whether between different geological regions or between the surface and subsurface. A low-cost simulant that reproduces the cohesion and bearing strength of an area of a planetary body's surface will not necessarily approximate its geochemistry. An example of this is using sand as a simulant for



Fig. 10.1 Bracketing use cases for the PlanetVac system. Left: tests in dense, rocky simulant. Right: tests in low-density aircrete. *Source* Courtesy of Honeybee Robotics

Martian regolith when testing rover wheels; the chemistry and weathering mechanisms that formed Martian regolith and Earth's sands are different, but testing at Honeybee Robotics has shown the geotechnical properties of some Martian regolith can be bracketed by testing with a variety of sands.

Geotechnical simulants can be appropriate for use as crude geochemical simulants (since a material on Earth geologically similar to the material in space will sometimes be similar in both aspects), but retrospective studies show simulants with incorrect properties have been used in inappropriate ways (Sibille et al. 2006). Similarly, tests have been conducted in air due to an assumption that there is little significance in the difference in geotechnical properties between regolith in air and under vacuum, but studies show even regolith simulant beds with and without trapped gas under vacuum have different penetration curves under atmospheric and vacuum conditions at similar levels of compaction (Kleinhenz 2014).

This chapter provides ISRU practitioners with information on the historical and modern steps for simulant development, information on the simulants used for different planetary bodies with a focus on Earth's moon, Mars, asteroids, and other celestial bodies, and the interaction of simulants and their testing environment for technology development.

10.2 Simulant Development

Regolith simulant is a constantly evolving field. Dozens of simulants over the decades have come into use and gone out of favor based on material availability and evolving understanding of the materials they simulate. Examples of these can be found in the following sections on lunar and Martian simulants. Regardless of target body, the process for developing a new simulant often follows these steps (Battler et al. 2006):

1. Study a region of a planetary body and determine its geological properties.
2. Select analog materials to replicate as closely as possible its mineralogy/chemistry *or* geotechnical properties based on the simulant use case. If the use case is unconcerned with composition, cheaper stand-in materials that match key geotechnical properties are used. Other important factors in analog material selection are cost, availability, and production speed.
3. Crush feedstock of the analog materials to approximate geotechnical properties. Perform any other processing techniques (drying, sieving, partial melting, addition of small percentage particles, etc.).
4. Test simulant for utility using the target technology for which the simulant was developed.

Because some popular simulants do not have data available for the original utility tests performed, simulant users must examine the literature on technology development studies that used the simulant. In the aerospace industry, technology developers have run into problems when they have not taken into account previous studies on a simulant's efficacy in simulating certain key compositional and geotechnical

properties (Taylor et al. 2016). Simulant users should therefore be careful to note if a study *found* a simulant effective for certain uses or if the simulant was *assumed* to be effective for the studied use case.

The following sections on lunar, Martian, asteroidal, and other simulants describe how ISRU specifically has driven simulant improvement and specialization. ISRU requires simulants with a wider variety of simulated geotechnical properties and at times closer adherence to regolith composition. ISRU is much wider in scope than the mostly mechanical interactions with dry regolith of previous exploration activities, such as loose material sampling, the development of lander feet and rover wheels, or early dust mitigation strategies.

10.3 Lunar Simulant

Lunar simulant was likely the first simulant ever made. Simulants have been made for a variety of use cases, and the quality of simulants has increased dramatically based on sample availability. In the early days of lunar exploration, scientists only had remote data and lunar meteorite composition to base their simulants on. In the late 1960s, during preparation for the Apollo missions, NASA scientists developed an inaccurate simulant because they assumed meteorite impacts would give the surface of the Moon an almost purely chondritic composition and cause the surface dust to have smoothed edges (Salisbury and Glaser 1964). In the first sample collection missions, scientists found that the surface dust is highly angular and is only composed of a few percent of meteoritic material (Fries et al. 2010). Lunar simulants were the first for which NASA designed a “figure of merit” for judging simulants. This initial document focused on geological accuracy, cost, producibility, and time frame of production (Rickman et al. 2010). It did not cover other aspects which have since become dominant, such as use-case fidelity and site specificity, but it served as a good starting point for considering future simulant development. Because the Moon has served as humanity’s cradle to the stars, it has also served as the inspiration for its test beds for consideration of how best to make the simulants that will help humanity get there.

Two long-time popular lunar simulants were JSC-1 and MLS-1. They have not been used regularly in more than a decade, but it is worth discussing their previous use as lunar simulants and the resulting pitfalls. JSC-1 was mined from a basaltic ash flow near Flagstaff, Arizona (McKay et al. 1994). It was developed as a low-titanium complement to MLS-1, claimed by the NASA workgroup at Johnson Space Center who developed it to be suitable for geotechnical experiments such as excavation and material handling and for practical work such as dust control and spacesuit durability. Even the team at JSC refrained from saying it was a suitable chemical simulant, saying it was currently being used in studies of oxygen production and sintering (McKay et al. 1994).

MLS-1 has a somewhat similar chemical composition to JSC-1 and lunar mare, but its chemistry is incorrect due to a 12% higher percentage of ilmenite (used in

several proposed ISRU processes (Sanders 2018)) and its mineralogy is inaccurate because its grain size is significantly skewed to larger sizes (Sibille et al. 2006). Crystal-setting velocity and temperature-dependent melt viscosity differ significantly between JSC-1 and similar chemistry reference materials such as BCR-1, meaning there is a potentially key difference in relevance between the simulants' melting behavior and what will be encountered on the Moon, even though the simulants are geochemically similar. Because representative lunar material has not been melted in bulk to observe its molten thermal properties, this is a use case where bracketing by using several simulants is key.

JSC-1 eventually became harder to acquire, giving rise to the need for a replacement. Its mare simulant successor JSC-1A (Gustafson 2009) is even now one of the most often used lunar simulants though it is no longer available for purchase. NU-LHT, which became popular as a highlands simulant, may have been the first simulant to use the “root and branch” model from its inception: the simulant uses a mix of several readily available materials rather than a premixed single-source feed-stock, and allows for site-specific and use-specific differentiation in future “branch” versions of the simulant (Stoeser et al. 2011). Though neither JSC-1A nor NU-LHT had published geotechnical properties when they were first formulated and introduced, they did have compositional information and particle size distribution, and many researchers have characterized their geotechnical properties since then (He 2010). JSC-1A was used for a number of ISRU-related tests, including several demonstrations of carbothermal reduction (Muscatello and Gustafson 2010) to show reduction of silicates and ilmenite. While JSC-1A may be useful for reduction of silicates, it contains no ilmenite. This is an example of ISRU technology development using a potentially inadequate simulant to demonstrate the full effectiveness of the technology (Taylor et al. 2010). A simulant with a realistic quantity of ilmenite may have proven that carbothermic reduction produces a greater amount of oxygen than previous tests showed. Another issue present in much of the testing done with JSC-1A and NU-LHT is the lack of nanophase iron. This material may be of particular importance for testing with the microwave sintering technology used for ISRU additive construction.

A number of modern simulants have been sold in recent years. Most of them have either been extremely inexpensive, such as GRC or BP-1 (Suescun-Florez et al. 2015), or of higher mineralogical and geotechnical fidelity, following the root-and-branch approach to specialization (Blewett et al. 2020). The low-cost modern simulants have been used for full-scale rover testing in large facilities such as the Swamp-Works lab at Kennedy Space Center. The higher-fidelity simulants have been used in testing of ISRU-related technology requiring stricter adherence to geochemical and geotechnical properties, such as concentrated solar melting (Williams 2021).

Because many technology developers have recognized the importance of testing under a wide variety of geotechnical properties, they have begun mixing their own simulants to bracket the properties they may encounter in space (Fig. 10.2). Custom root-and-branch simulants can be made either of low-cost materials for large-scale tests, or of compositionally and geotechnically accurate expensive material for testing



Fig. 10.2 Left to right: simulant materials in buckets before combination, mixing simulant base materials for desired geotechnical properties, testing in a thermal vacuum environment using custom simulant, bore hole after test. *Source* Honeybee Robotics

simulated geology of specific sites not represented in other simulants. These custom simulants are also useful for testing specific scenarios such as rocks buried deep in the surface or sudden changes in geology.

10.4 Martian Simulant

Mars has been a target since the early days of the space race between the United States and the Soviet Union (Perminov 1999). Though spectra of regolith have been taken and studied through telescopes, orbital spacecraft, from meteorites, and at the Martian surface, no samples have been returned to Earth. Mars also has a comparatively varied geology (Tanaka et al. 2014) mostly due to the wind and water erosion which the Moon does not have. This has left scientists in a predicament: what part of Mars (if any) should they try to model with simulant, and what will it be useful for?

The previously most popular Martian simulant was a spectrally matched simulant called JSC-Mars-1. This simulant was developed based on visible/near-IR reflectance spectroscopy, X-ray fluorescence, and loss-on-ignition analysis (Allen et al. 1997). A comparison of its geochemistry with a Martian surface sample collected by the Viking Lander 1 was made (though JSC-Mars-1 differed significantly from the sample). Its particle size distribution was noted in the original publication detailing its announcement but was not compared to particle size distribution data from any Martian sample. The authors also did not quote what this simulant was meant to be used for besides support of scientific research and engineering.

Clearly, this is a simulant that simply *appears* Martian and shares vaguely similar geochemical properties to a landing site. If the geotechnical properties of a site on the extraterrestrial surface are not studied, it is impossible to say whether the simulant matches those properties. JSC-Mars-1 has a significant percentage of organic and

volatile content, unlike what is likely to be encountered on Mars (Allen et al. 1998). Potentially the only purpose for which JSC-Mars-1 is more accurate than common hardware-store sand is optical technology development (such as cameras or spectrometers). This leads to a conundrum: how can engineers safely develop technology for space when they need data from surface missions to make accurate simulants? The answer to this lies in bracketing. Testing in a variety of sample materials lowers the need for any one simulant to be exactly like what will be encountered. Honeybee Robotics used this approach in developing the Mars 2007 Phoenix Lander Icy Soil Acquisition Device (ISAD) (Bonitz et al. 2007).

ISAD is a scoop with two blades and a rasp used for acquiring icy and dry regolith samples. Because ISAD was made for use with a variety of materials of different bearing strength, hardness, and chemical makeup, it had to be tested under a variety of circumstances. Rather than find exact matches for the overburden and expected icy regolith below, Honeybee Robotics tested ISAD in pure ice, fine basaltic sand, the hard soil in Death Valley, and JSC-Mars-1 mixed with a variety of weight percentages of water ice (ranging from 0 to 30% and with icy chunks on top). Because ISAD was tested under and optimized for such varied conditions, when problems arose on the Martian surface due to more rapid sublimation of excavated samples than expected, the ISAD was able to alter its excavation technique and acquire samples to complete its mission (Fig. 10.3).

Several other Martian simulants have been developed since JSC-Mars-1 was released. Mojave Mars Simulant (MMS) was developed specifically because JSC-Mars-1 was unrealistically hygroscopic for water sublimation permafrost loss tests (Peters et al. 2008). In one of the introductory papers for MMS, the development team described its optical, geotechnical, and compositional properties in relation to those of several sites on Mars. MMS was the best Martian simulant at the time based on the number of data points available, but as new data from Mars sampling missions

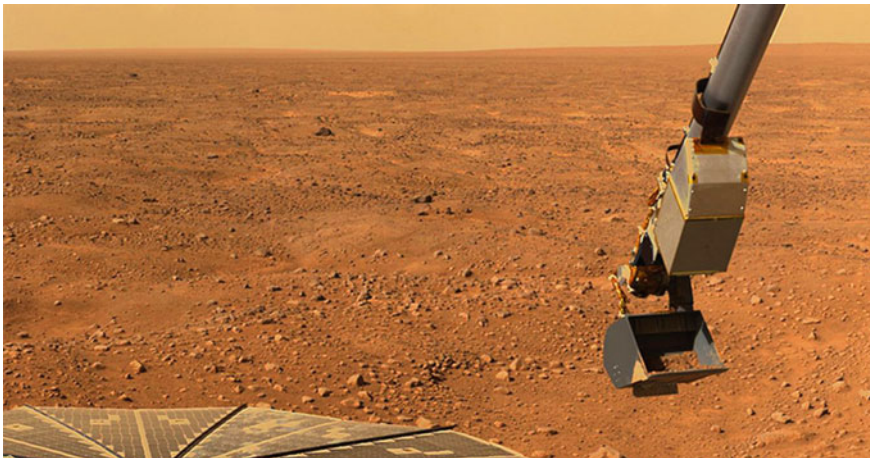


Fig. 10.3 ISAD on the Martian surface. *Source* NASA

has emerged over the course of a decade, new simulants have been made available that more closely match bulk Martian regolith properties and characterize the simulant for specific uses. Mars Global Simulant (MGS-1) is a modern simulant that took a similar development approach to MMS: it improved upon previous simulants by integrating new data from Mars that had come in over the years, described its feed-stock clearly, and gave a full explanation of its differences in properties and use cases when compared with other simulants and sites on Mars (Cannon et al. 2019). MGS-1 addressed the need for future production by using the root-and-branch model.

These are key factors in ISRU technology development on Mars when investigating use cases such as 3D printing or subsurface digging for water ice. The first natively constructed habitats on Mars will likely be built by melting Martian regolith or mixing it with polymer binders. 3D printing in powder is a difficult enough task on Earth, with powder development companies keeping strict control over the material composition and particle size used in printers. No truly material-agnostic printers exist, so the utility of polymer binder printing on Mars will depend on tailoring the technology for use with a specific regolith material from a specific site at a specific grain size. This concept of using the minimum standards of analogous Earth technology is important for safety and realistic ISRU technology utility evaluation (Williams and Butler-Jones 2019). The closer simulant manufacturers can get to simulating such a specific Martian regolith, the lower the risk will be for future missions to the Martian surface.

Martian simulants will continue to improve as data emerges from current and future science missions. Because of the complex and highly non-homogeneous geology of Mars as well as the varied use cases for Martian ISRU technology, a variety of Martian simulants will be necessary in the future. Through regular simulant quality improvements and bracketing by choosing simulants with different properties, technology developers can avoid mission failures on the red planet.

10.5 Asteroid Simulant

Asteroid simulant development has been approached in a fundamentally different way from other simulants because of three factors: (i) many asteroid samples exist on Earth in the form of meteorites (though meteorites do not match particular asteroids); (ii) asteroids have much more varied geological properties and formation mechanisms than most lunar and planetary surfaces; and (iii) until recently there have not been any missions to land on an asteroid. The wide variation in asteroid geology means the technology for harvesting their materials relies on a variety of physical properties. Scientists have proposed pneumatic, mechanical, and concentrated solar mining systems for resource extraction technologies (Zacny et al. 2013).

Asteroids are roughly categorized into chondritic clay/silicate or C-types, stony mixes of silicates and nickel/iron called S-types, and metallic asteroids called M-types. Different asteroid types have different formation characteristics and as a result vary highly in density and structure. Asteroids can be hard and dense, fluffy and dusty,

have surface structures dominated by electrostatic forces or unusual impact craters, and exist in microgravity, so a purely geotechnical simulant is not effective for all applications (Housen and Holsapple 2003). Because the low albedo of asteroids makes it difficult to study them, many of their characteristics were unknown until recent decades. Therefore until recently, simulants to match the properties needed for geochemical or microgravity behavior tests, or technology development were not made (Zacny et al. 2020).

Early work in asteroid simulant in the 1960s in support of lunar surface formation mechanism modeling used a variety of materials such as sand shot at high velocity at a simulated lunar surface (Gault et al. 1963). The field saw little growth until the 1980s when an asteroid collision study was performed, comparing impact clouds from different materials with dry ice (Schultz and Gault 1986). In subsequent years it was found that asteroids were not just singular hard bodies with dusty surfaces, but bodies with cracks and fissures, and potentially more dust than previously thought (Housen 1992). Before the Hayabusa mission, a few simulants were tested for use with mechanical impactors including limestone as a simulant for S-type and tuff (dry volcanic ash) as a simulant for C-type asteroids (Yano et al. 2000). Scientists expected less sample to be collected from C-type, but the 25,143 Itokawa asteroid, an S-type, required two attempts to collect enough sample. Part of this difficulty was due to system-level issues, but another part was simply that not enough representative simulants were tested to bracket the potential use cases.

The modern need for asteroid simulant became clear when companies such as Deep Space Industries and Planetary Resources released public statements on their intentions to mine asteroids for their materials. Subject matter experts created a figures-of-merit grading system for asteroid simulants like the lunar simulant figures of merit previously released by NASA (Metzger et al. 2019). The simulant made by Deep Space Industries to enable asteroid mining technology testing followed the root-and-branch method previously described by creating a recipe of various base minerals to grossly match the chemistry of a representative asteroid of a specific type (Britt et al. 2017). Dr Daniel Britt, recipe designer for the Deep Space Industries simulant, acknowledged the impossibility of matching meteorites to asteroids on a one-to-one basis, but used optical methods and matched estimated asteroid composition to what is known in both meteorites and accessible simulant materials (Britt et al. 2019). These asteroid simulants were used in the construction and testing of a concentrated solar asteroid mining project at the Colorado School of Mines (Fig. 10.4). Cobbled chunks of asteroid simulant of varying densities and varying levels of cohesion were tested. By using the same feedstock simulant and varying the construction methods, it was possible to test the utility of concentrated solar mining with a variety of asteroid formations. Because asteroids hold enough water and metals to allow humanity to one day spread throughout the solar system, they will remain a target for ISRU technology developers. To deal with the difficulty of finding, landing on, and extracting resources from asteroids, new and varied simulants will also remain important until astronauts are regularly testing new technology on asteroid material directly.



Fig. 10.4 Left: asteroid simulant spalling under a concentrated sunlight simulator. Right: Various densities, water concentrations, and structures of asteroid simulant. *Source* Colorado School of Mines

10.6 Other Simulants

For over a decade NASA has had ISRU as a goal with a specific focus of harnessing the resources of the Moon, and one of the most important resources is water ice (Sanders and Larson 2012). Because the exact nature and percentage of ice in the lunar poles is still unknown, a variety of weight percentages and forms of ice have been investigated. The differences in physical properties between these varieties are even more extreme than the differences in base regolith simulants, ranging in bearing strengths from that of weak coal to limestone (Gertsch et al. 2006). Ice particles in the cold shadows of the Moon likely arrive there through migration across the lunar surface; a simplification of the process is that vapor particles get to the surface, move randomly across it during the day, freeze during the night, eventually arriving at a shadow which the sunlight does not reach and remaining there (Taylor et al. 2007). This ice may form at extremely cold temperatures, leading to amorphous ice or higher phases of ice with odd crystalline and electrical properties (Fukazawa et al. 2006; Debenedetti 2003). Ice is also present on the icy ocean moons Europa and Enceladus. For current exploration and ISRU activities farther into the future requiring penetration of the surface ice, icy and underwater simulants have also been developed (Fig. 10.5). Drills that can handle a variety of icy, slushy conditions and can operate with semi-collapsing boreholes are being developed for these future ISRU and exploration activities.

Because it is not known if ice will be fluffy and mixed into the regolith without cementing, re-melted between regolith grains to form a contiguous structure, or something of unknown qualities in between, any technology for ice excavation or even locomotion through the ice fields must be developed through bracketing these potential cases. Availability is also a concern; while it is possible to premix water and simulant and put it in a freezer, grind ice using a commercial machine, or use liquid nitrogen to create frozen droplets and get down to colder temperatures, those specific methods are some of the only ones easily available, and each of them has specific temperatures and specific qualities of icy simulant mixtures. Icy regolith made by adding water to regolith then freezing it becomes extremely hard at weight

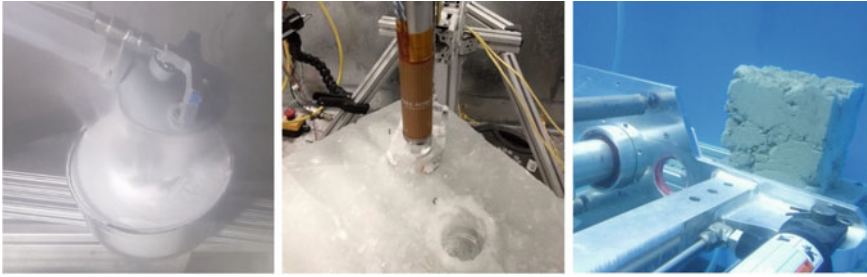


Fig. 10.5 Left: extremely low-density material for microgravity simulation. Middle: ice for Enceladus drill tests. Right: concrete blocks underwater for Europa drill tests. *Source* Honeybee Robotics

percentages from 2% upwards. Sublimation speed depends on temperature and pressure, so being limited to either freezer or liquid nitrogen pressures means that some temperature and pressure combinations are not easily tested. These problems can be addressed by ISRU technology developers, but it is important to take into account how much is unknown and how much is being bracketed in the test preparation.

Microgravity simulants are also useful for space technology development. These materials are generally extremely low density and are useful for demonstrating technology to be used on bodies with low gravity such as small asteroids. For ISRU purposes, at times it is necessary to use chemically accurate simulants to determine the utility of technology for harvesting materials, but at times it is more important to determine how the dusty surfaces will behave when developing anchoring and capturing technology. Another area of interest is the behavior of dust in microgravity: when gravity is significantly lower than on Earth, even at lunar strengths, electrostatic effects can take precedence and dust can behave like gel, sticking to and creeping up surfaces (Sercel et al. 2018). Dust mitigation techniques are important for ISRU applications: dust is one of the biggest threats to both human and mechanical systems in ISRU applications (Gaier and Creel 2005). Many fundamental issues with ISRU technology development boil down to energy issues and thermal issues. If dust layers reduce the combined effectiveness of multi-layer insulation and solar panels, the battery systems required to keep ISRU technology alive during cold nights make the systems economically unfeasible. To examine these combined microgravity dust effects, scientists have used materials such as 3 M glass bubbles with densities less than 0.2 g/cc. For lower-gravity scenarios such as the Moon's gravity, which is 1/6th that of Earth, materials such as aircrete have also been used. Low-density simulants such as aircrete and glass bubbles are also used for bracketing purposes with other materials such as sand when developing technology for entirely unknown low-gravity bodies such as the moons of Mars (Zacny et al. 2020).

10.7 Environmental Requirements

The gaseous environment under which testing occurs has an impact on a variety of tests. For materials scientists, testing and reporting material behavioral differences under vacuum and ambient conditions is always necessary if the material will be used in a vacuum. However, while the materials scientists will often report the changes in behavior between ambient and vacuum, spacecraft technologists are often simply concerned with whether it works under both. Thermal vacuum testing is part of any NASA technology readiness-level advancement plan. NASA's vacuum test facilities often go to the molecular flow regime of 10^{-6} Torr and higher (JSC 2011), so clearly NASA takes vacuum testing seriously. However, space technologists generally do not report differences in behavior between ambient and vacuum conditions because a technology that works in ambient but not in vacuum must be redesigned and retested. Thermal vacuum testing is treated as a necessary step for verifying a technology's readiness for use in space, but not for examining physical behavior changes with the technology or the simulant. This is an unfortunate characteristic of the industry because of the wealth of knowledge not published and examined as a result. However, some scientists have looked at material behavior changes under vacuum and ambient conditions for both mechanical and melting properties of regolith simulant.

Cone penetrometers are common, inexpensive tools for geotechnical property investigation. Though cone penetrometer tests do not show an effective difference between ambient and vacuum conditions during penetration, relaxation curves can be used to determine a difference (Atkinson et al. 2019). Certain physical properties such as bearing strength may not be affected by the pressure of the testing environment, but other properties such as shear strength measurements may be, because the cone penetrometer's relaxation curves can determine differences in shear strength, temperature, water content, and other soil characteristics. This is as important to exploration technology developers as it is to ISRU practitioners because of the correlation between material and physical properties. At first, it will not be possible to put samples from every location of interest through chemical evaluation at the lunar surface, but resource extraction companies will need detailed maps of ground truth data to back up remote sensing data from satellites. Simple tools like penetrometers can serve as intermediate providers of rough ground truth data on geotechnical properties when chemical analysis is unavailable.

For melting technology tests, the presence or absence of air causes a significant difference in strength, volume, test repeatability, and coloration. At Colorado School of Mines (Williams 2021), tests to determine the effects of changes in melting behavior of regolith simulant in ambient and vacuum environments were performed (Fig. 10.6). JSC-1A was melted in air and at a low vacuum (around 0.01 Torr). LMS-1, a modern, more mineralogically accurate simulant from Exolith Lab, was also melted in vacuum. All samples were melted using a xenon arc lamp and focusing lens. Each sample was then pressed to failure and the failure force required was recorded. The JSC-1A melted in air had a bimodal distribution of failure force loosely correlated

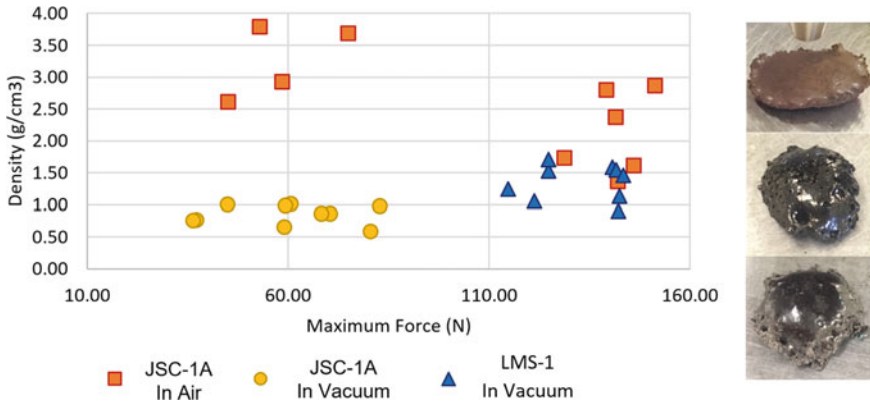


Fig. 10.6 Left, table showing maximum force before failure vs density of samples melted in air and in vacuum. Top right: JSC-1A sample melted in air. Middle: JSC-1A melted in vacuum. Bottom: LMS-1 melted in vacuum. *Source* Colorado School of Mines

to density. The JSC-1A melted in vacuum had more regular densities but high variability in maximum force. The LMS-1 had the highest repeatability for maximum force and density. Among the important discoveries of this investigation was that the most commonly used material for regolith melting tests (JSC-1A), in the most common test environment (atmospheric conditions), produced the least technically valuable results. For ISRU technology developers this has the problem of making potentially good technology seem weaker than it is. Because many 3D printing and oxygen extraction technologies rely on full or partial regolith melting, this is a potentially serious issue. It is not always economically or materially feasible to run tests under vacuum, but it is important for technology developers not to rule out promising technology that has not been tested in a more realistic (and potentially easier) vacuum environment.

Sample preparation technique is also important. If a simulant has been used for many years it may have lost its initial grain jaggedness and its particle size distribution may be skewed to lower-sized particles as large particles are skimmed off the top from settled containers. Compaction levels and density can be standardized before simulant is put into a vacuum chamber by slowly introducing simulant into a container, using a surcharge to compact it a desired amount, then adding another layer and repeating the process until a desired height and general compaction is reached. However, once the vacuum pump is turned on and air begins rushing out of the chamber, air also rushes out from between the grains of the simulant, often causing the simulant to lose its compaction. Sometimes pockets of air can be trapped in a closely packed section until the pressure differential between the chamber and inside the pocket is great enough for it to burst. This causes several effects, from total loss of compaction of the top layer while the bottom layer stays compacted, to thinner vents of gas forming “volcanos” in the sample container, to miniature dust explosions in the chamber. Similarly, the speed at which water is introduced to a sample will change the

clumpiness and localized cohesion and weight percentage of water of a sample. For drilling operations, a well distributed wet sample will behave significantly differently from a mostly dry sample with very wet clumps mixed in. This is not an exhaustive list of sample preparation concerns, but it does show that simply getting the simulant and the gaseous environment right is only part of the problem for a repeatable and useful test setup.

10.8 Conclusion

The purpose of this chapter is to give a general overview of the way simulants are developed and characterized, to describe how simulants were made in the past and how they differ from the simulants of today, to explore the differences between simulants for the Moon, Mars, asteroids, and other bodies, and to examine how the ambient pressure environment can affect tests on regolith simulant in non-intuitive ways. Key concepts are: bracketing the use cases of a technology by testing with different simulants that have a variety of geotechnical and compositional properties; evaluating the effectiveness of a simulant for an activity by investigating how well the properties that are important to the test align with those of the simulated body; and using root-and-branch simulant development methods to create simulants for specific use cases that can be evaluated against similar simulants for other use cases. As more data comes in from direct regolith sampling experiments on target celestial bodies, simulants will likely become more accurate and site specific. As ISRU grows as a field, simulant users will need this accuracy and specificity to reduce risk and increase the breadth of technologies available to those using space resources.

References

- Allen, Carlton C., et al. 1997. JSC Mars-1-Martian regolith simulant. *Lunar and Planetary Science Conference*, vol 28.
- Allen, Carlton C., et al. 1998. Martian soil simulant available for scientific, educational study. *Eos, Transactions American Geophysical Union* 79(34):405–409.
- Atkinson, J., and K. Zacny. 2018. Mechanical properties of icy lunar regolith: application to ISRU on the Moon and Mars. In *Earth and Space 2018: Engineering for Extreme Environments*. Reston, VA.
- Atkinson, Jared, et al. 2019. Penetration and relaxation behavior of dry lunar regolith simulants. *Icarus* 328:82–92.
- Battler, M.M., et al. 2006. Developing an anorthositic lunar regolith simulant. In *57th International Astronautical Congress*.
- Blewett, David T., et al. Lunar simulant assessment. Johns Hopkins University Applied Physics Laboratory.
- Bonitz, Robert G., et al. NASA Mars 2007 Phoenix lander robotic arm and icy soil acquisition device. *Journal of Geophysical Research: Planets* 113(E3).

- Britt, Daniel, Steven D. Covey, and Cody Schultz. 2017. University of central Florida/deep space industries asteroid regolith simulants. *DPS*.
- Britt, Daniel T., et al. 2019. Simulated asteroid materials based on carbonaceous chondrite mineralogies. *Meteoritics & Planetary Science* 54(9):2067–2082.
- Cannon, Kevin M., et al. 2019. Mars global simulant MGS-1: A Rocknest-based open standard for basaltic Martian regolith simulants. *Icarus* 317:470–478.
- Debenedetti, Pablo G. 2003. Supercooled and glassy water. *Journal of Physics: Condensed Matter* 15(45):R1669.
- Fries, Marc, et al. 2010. Extralunar materials in lunar regolith. *A White Paper Submitted for the NRC Decadal Survey*.
- Fukazawa, Hiroshi, et al. 2006. Existence of ferroelectric ice in the universe. *The Astrophysical Journal Letters* 652(1):L57.
- Gaier, James R., and Ronald A. Creel. 2005. The effects of lunar dust on advanced EVA systems: Lessons from apollo. Presentation Jan.
- Gault, Donald E., Eugene M. Shoemaker, and Henry J. Moore. 1963. *Spray ejected from the lunar surface by meteoroid impact*. National Aeronautics and Space Administration.
- Gertsch, Leslie, Robert Gustafson, and Richard Gertsch. 2006. Effect of water ice content on excavatability of lunar regolith. In *AIP conference proceedings*, vol. 813(1), American Institute of Physics.
- Gustafson, G. 2009. JSC-1A lunar regolith simulant: availability and characterization. In *2009 Lunar Regolith Simulant Workshop*.
- He, Chunmei. 2010. *Geotechnical characterization of lunar regolith simulants*. Diss: Case Western Reserve University.
- Housen, Kevin R., and Keith A. Holsapple. 2003. Impact cratering on porous asteroids. *Icarus* 163 (1): 102–119.
- Housen, Kevin R. 1992. Crater ejecta velocities for impacts on rocky bodies. In *Lunar and Planetary Science Conference*, vol 23.
- Kleinhenz, Julie E., and R. Allen Wilkinson. 2014. Development and testing of an ISRU soil mechanics vacuum test facility.
- McKay, David S., et al. 1991. The lunar regolith. In *Lunar sourcebook*, vol 7. New York: Cambridge Univ. Press, 285–356.
- McKay, David S., et al. 1994. JSC-1: A new lunar soil simulant. In *Engineering, construction, and operations in space IV 2*, 857–866.
- Metzger, Philip T., et al. 2019. Measuring the fidelity of asteroid regolith and cobble simulants. *Icarus* 321:632–646.
- Muscattello, Anthony, and Robert Bob Gustafson. 2010. The 2010 field demonstration of the solar carbothermal reduction of Regolith to produce oxygen. In *11th Space Resources Roundtable: Planetary and Terrestrial Mining Sciences Symposium*; 6–8 June 2010, Golden, CO.
- Perminov, V. G. 1999. *The difficult road to Mars: a brief history of Mars exploration in the Soviet Union*. National Aeronautics and Space Administration Headquarters.
- Peters, Gregory H., et al. 2008. Mojave Mars simulant—Characterization of a new geologic Mars analog. *Icarus* 197(2):470–479.
- Rickman, D.L., C.M. Schrader, and J.E. Edmunson. 2010. Generation of requirements for simulant measurements. In NASA Technical Memorandum 2010–216445, NASA Marshall Space Flight Center, Ala.
- Salisbury, John W., and Peter E. Glaser. 1964. *Studies of the characteristics of probable lunar surface materials*. Little (Arthur D) Inc., Cambridge, MA.
- Sanders, Gerald B., and William E. Larson. 2012. Progress made in lunar in situ resource utilization under NASA's exploration technology and development program. In *Earth and space 2012: Engineering, science, construction, and operations in challenging environments*.
- Sanders, Gerald B. Overview of past lunar in situ resource utilization (ISRU) Development by NASA.

- Schultz, Peter H., and Donald E. Gault. 1986. Momentum transfer from oblique impacts. In *Lunar and planetary science conference*, vol 17.
- Sercel, J., et al. Microgravity granular material research facility for ISS. In *Earth and Space 2018: Engineering for Extreme Environments*, Reston, VA.
- Sibille, Laurent, et al. 2006. Lunar regolith simulant materials: recommendations for standardization, production, and usage.
- Stoeser, D., D. Rickman, and S. Wilson. 2011. Design and specifications for the highland regolith prototype simulants NU-LHT-1M and-2M.
- Suescun-Florez, Eduardo, et al. 2015. Geotechnical properties of BP-1 lunar regolith simulant. *Journal of Aerospace Engineering* 28(5):04014124
- Tanaka, Kenneth L., et al. 2014. Geologic map of Mars.
- Taylor, Lawrence A., Carle M. Pieters, and Daniel Britt. 2016. Evaluations of lunar regolith simulants. *Planetary and Space Science* 126: 1–7.
- Taylor, G.Jeffrey, P.G. Lucey, and N. Schörghofer. 2007. The Science of the Lunar Polar Volatile Deposits. *AAS* 210:71–103.
- Taylor, Lawrence A., and Yang Liu. 2010. Important considerations for lunar soil simulants. In *Earth and space 2010: Engineering, science, construction, and operations in challenging environments*, 106–118.
- Thermal-Vacuum Testing. 2011. *Facility Testing Information*. Johnson Space Center, National Aeronautics and Space Administration. https://www.nasa.gov/centers/johnson/pdf/639724main_Thermal_Vacuum_Testing_FTI.pdf. Accessed 12 Oct 2020.
- Williams, Hunter, and Evan Butler-Jones. 2019. Additive manufacturing standards for space resource utilization. *Additive Manufacturing* 28: 676–681.
- Williams, Hunter. 2021. Simulant and environment requirements for space resources technology development. In *Earth and Space 2021: Engineering for Extreme Environments*.
- Yano, H., et al. 2000. Trajectory planning and sampling technology of asteroid family multiple fly-bys and sample return mission.
- Zacny, Kris, et al. 2013. Asteroid mining. In *AIAA Space 2013 Conference and Exposition*.
- Zacny, Kris, et al. 2020. Pneumatic Sampler (P-Sampler) for the Martian Moons eXploration (MMX) Mission. In *2020 IEEE Aerospace Conference*. IEEE.

Chapter 11

Regolith Processing



Kevin M. Cannon and Robert P. Mueller

Abstract Regolith processing includes a growing set of methods and technologies that will be used to extract resources found within regolith, or to transform the regolith itself into a useful product. In this chapter we describe the basic properties of granular regolith on the surfaces of the Moon, Mars, and asteroids, and the processing steps necessary to upgrade its value as a potential space resource. The mineralogy, chemistry and physical properties of regolith differ by solar system body and according to local geology, and these are necessary to consider in designing a regolith processing architecture. Particularly important is the distinction between mineralogy and bulk chemistry, which often causes confusion because of certain conventions used by geologists to report compositions. Regolith processing techniques are divided into excavation and transport, separation/beneficiation, particle bonding, and extracting resources from regolith. Regolith excavators will be highly automated, and are divided into discrete and continuous methods with trade-offs between the two. Some of the most efficient excavator designs have emerged from trends in successive robotic mining competitions, with bucket ladders, front-end loaders and bucket belts being the most popular choices. After excavation, regolith will likely need to be beneficiated or pre-processed to select specific size fractions or compositions. We discuss methods and challenges for each of these steps, emphasizing the difficulty of beneficiating dry, fine-grained lunar regolith which is a fiendishly complex material. Particle bonding techniques impart strength back to unconsolidated regolith and include sintering, geopolymers, regolith concrete and 3D printing, with 3D printing showing the most promise for large structures constructed in challenging planetary environments. For extracting space resources found within regolith, the challenge is to choose the right type of energy (concentrated solar, microwave, etc.) and system configuration (closed vessel, covered dome, etc.) to

K. M. Cannon (✉)

Department of Geology and Geological Engineering and Space Resources Program, Colorado School of Mines, Golden, USA
e-mail: cannon@mines.edu

R. P. Mueller

Exploration Systems and Development Office, National Aeronautics and Space Administration (NASA) Kennedy Space Center “Swamp Works”, Florida, USA
e-mail: rob.mueller@nasa.gov

deliver that energy to the regolith to extract the resource in question. Finally, techniques with future potential include using biologic processes as a form of “biomining” to extract specific elements from regolith.

11.1 Introduction

Almost all space resources are located within or beneath regolith, and regolith itself can be considered a type of resource that can be put to use in a diverse range of applications. Learning how to excavate, transport, separate, bind together, and chemically process this material at scale is essential for building a robust space economy founded on the use of local materials, because in many cases regolith is the only accessible material present on a planetary body. Regolith composition and behavior differ between bodies (asteroids, Moon, Mars), but there are commonalities in the types of engineering solutions needed to handle and process unconsolidated granular materials in order to extract resources from them, or to use regolith directly by transforming its properties.

A tremendous amount has been learned about the regolith on different worlds from remote sensing observations, in-situ exploration by robotic spacecraft and humans, and by studying returned samples. In many cases the properties of regolith are similar to granular materials commonly processed on Earth, but the unique combination of low gravity, vacuum or near-vacuum conditions, and exotic geologic processes leads to important differences that will factor heavily into the technologies developed to process regolith. To some degree, these features can be captured by terrestrial regolith simulants, vacuum chambers, and low-gravity experiments, but this type of testing is never a perfect substitute for the real thing.

This chapter is focused on regolith processing, ranging from digging it out of the ground to high-energy chemical transformations. We start by discussing the composition and physical properties of regolith on different planetary bodies. Excavation and conveyance are covered in detail, with lessons learned from decades of research at NASA Swamp Works, and from experience with robotic mining competitions. Then we look at techniques to separate regolith into different particle size fractions and different mineral or chemical constituents. Particle bonding techniques are described and compared in the context of building structures from regolith. Finally, we end by considering high-energy techniques for resource extraction, as well as innovative solutions that employ biology to process regolith. Throughout, we draw on real-world results from spacecraft and humans interacting with regolith, from experimental results in laboratory studies on Earth, and from computer modeling simulations that extend those results to other settings.

11.2 The Nature of Regolith

Regolith is defined as the fragmented layer of unconsolidated material that blankets the surfaces of most rocky planetary bodies (Fig. 11.1). This definition is quite broad, and more specific vertical zones and size fractions of the regolith have been identified for practical purposes. On the Moon and Mars, the *megaregolith* is the zone of heavily fractured material that extends kilometers deep, recognizing that the regolith at the surface—explored by humans, landers and rovers—barely scratches the surface of the part of the crust affected by intense impact bombardment early in the history of the solar system (Hartmann 1973). Richardson and Abramov (2020) divide the lunar regolith into three vertical zones: a *surficial regolith* 5–20 m thick followed by 1–3 km of *upper megaregolith* that is characterized by material transported and deposited in layers. This upper megaregolith layer is followed by 20–25 km of fractured-in-place bedrock called the *lower megaregolith*. These divisions can likely be extended to Mars as well, with some adjustments to the specific depths. For asteroids, it is more challenging to define such regions, particularly for rubble pile asteroids where the entire body can be thought of as comprising a single horizon of regolith (Fujiwara et al. 2006; Walsh 2018).

The surficial regolith is most relevant to space resources and regolith processing in particular, so we will restrict ourselves to that zone in this chapter. Surficial regolith is made up of grains that range from sub-microscopic all the way up to house-sized boulders. For convenience, *soil* in a planetary context is used colloquially to refer to the fraction of regolith with particle sizes <1 mm in diameter. There is some controversy surrounding this term, because in a terrestrial context the word *soil* is usually reserved for an organic-rich material including its biota; however, *soil* is in common use for planetary bodies and will be used here as well. *Dust* refers to the very finest particle size fraction of the regolith: there is no hard cutoff, but lunar dust is usually taken to mean particles with <20 or <10 μm diameters, and dust on Mars is the fraction that is fine enough to be lofted high into the atmosphere (microns in size). Standard particle size definitions from terrestrial sedimentology (Table 11.1) are also used for regolith, although this is more common for Mars which has a rich



Fig. 11.1 Regolith on the Moon (left), Mars (center), and the asteroid Bennu (right). *Image credits* NASA/JPL/Malin

Table 11.1 Particle size divisions based on the Wentworth scale

Size fraction	Diameter	Phi scale
Boulder	>256 mm	<-8
Cobble	64–256 mm	-6 to -8
Gravel	2–64 mm	-1 to -6
Sand	62.5 μm to 2 mm	4 to -1
Silt	3.9–62.5 μm	8 to 4
Clay	0.98–3.9 μm	10 to 8
Colloid	0.95–977 nm	20 to 10

record of sedimentary erosion and deposition (e.g., Malin and Edgett 2000). Note that clay in geology can refer to a particle size fraction of any composition, or to clay minerals which are a type of silicate.

11.2.1 *Regolith Petrology and Mineralogy*

Regolith particles can be made up of crushed rocks and breccia fragments, individual mineral or glass fragments, or more complex derivative products formed by surface and space weathering processes (Fig. 11.2). For the Moon, regolith petrology is usually divided into: (1) lithic fragments, (2) breccia fragments, (3) mineral fragments, (4) glass fragments, and (5) agglutinates, which are complex, fractal-shaped particles made of small grains welded together by interstitial glass (Fig. 11.3). Most of these particle types are expected to exist on other rocky planetary bodies, except for agglutinates which are not present on bodies with atmospheres because the micrometeoroids that form agglutinates will not reach the surface on those bodies. Figure 11.4 shows a close-up image of Martian regolith, where lithic fragments and glass beads are prominent, and almost all grains are coated in an orange-tinted dust.

To a large extent, the modal mineralogy of regolith will reflect the local or regional mineralogy of the underlying crust from which the regolith is derived. Exceptions arise from material transported long distances by energetic impacts, and from meteoric material that falls from space and makes up a small fraction of the regolith. Table 11.2 gives the modal mineralogy for different representative terrains on the Moon and Mars, and for several meteorites that are linked to S (stony) and C (carbonaceous) type asteroids.

11.2.2 *Regolith Chemistry*

Like mineralogy, the major element chemistry of the regolith reflects the bulk chemistry of a planetary body's crust, usually with local variations imposed on a relatively

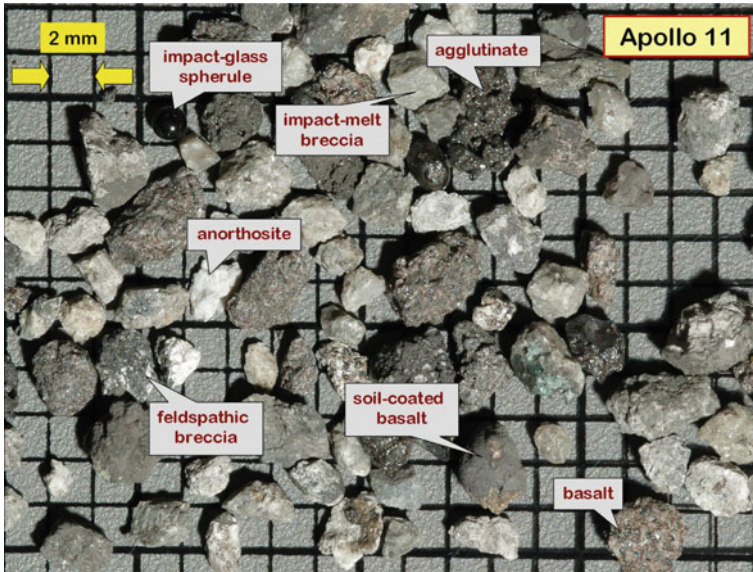


Fig. 11.2 Close-up view of lunar regolith grains that have been size sorted and cleaned. *Image credit NASA*

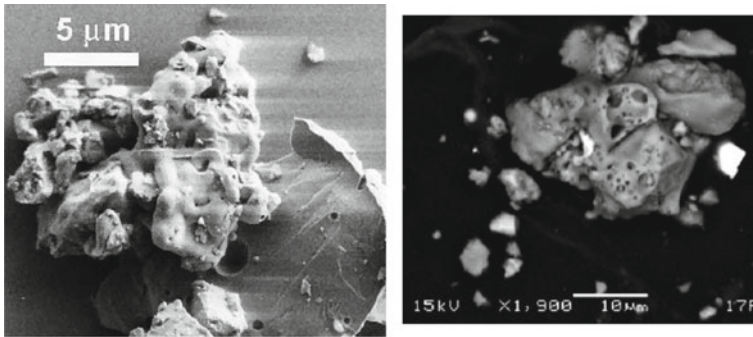


Fig. 11.3 Scanning electron micrographs of lunar agglutinate particles. *Image credit NASA*

homogeneous regional composition. For example, the mare terrains on the Moon vary locally to regionally in the abundance of titanium, but overall have a fairly narrow range in most other elements. Table 11.3 gives representative analyses from returned samples and in-situ measurements.

It is important to note that writing bulk major element chemistry as oxides (Table 11.3) is a standard convention in geology. It is a coincidence that some of these oxides are the chemical formulas for common minerals (e.g., quartz, SiO_2). The values in Table 11.3 are chemistry data and do not imply anything about the presence or absence of oxide minerals like quartz in the regolith.

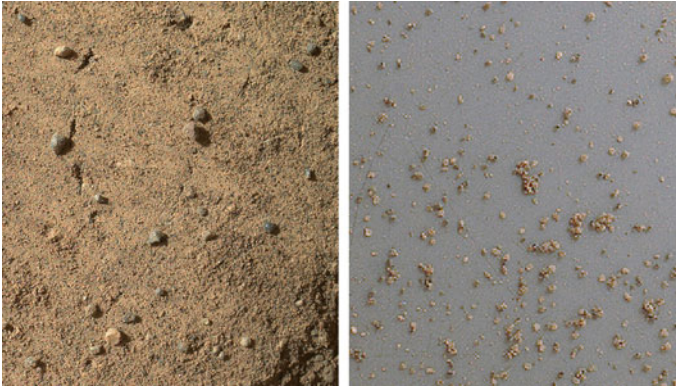


Fig. 11.4 Close-up view of the Martian regolith at the Rocknest site, acquired by the MAHLI imaging camera on MSL Curiosity. *Image credit* NASA/JPL/Malin

Table 11.2 Modal mineralogy of representative regolith or meteorite samples (weight percent in parentheses)

Terrain or material	Modal mineralogy
Lunar highlands ^a	Plagioclase (79.5), Pyroxene (8.6), Olivine (4.0), Ilmenite (0.4), Glass (7.5)
Lunar mare ^a	Plagioclase (50.0), Pyroxene (21.7), Olivine (13.0), Ilmenite (1.4), Glass (13.7)
Mars ^b	Plagioclase (40.8), Pyroxene (28.4), Olivine (22.4), Magnetite (2.1), Anhydrite (1.5), Quartz (1.4), Sanidine (1.3), Hematite (1.1), Ilmenite (0.9)
Carbonaceous chondrite (C-type asteroid) ^c	Phyllosilicates (77.4), Olivine (11.5), Pyroxene (4.9), Sulfides (3.0), Magnetite (2.0), Calcite (1.0)
Ordinary chondrite (S-type asteroid) ^d	Olivine (34.7), Pyroxene (32.8), Plagioclase (9.5), Troilite (6.1), Metal (15.3), Other (1.6)

^a Papike et al. (1982)

^b Bish et al. (2013)

^c Howard et al. (2011)

^d Dunn et al. (2010)

11.2.3 Regolith Physical Properties

The rheological behavior of a granular material like regolith is a highly complex subject that is still not totally understood, although advanced computer models are making good progress. The physical properties of regolith are a function of both particle properties (size, shape, density) and the disordered packing of these particles into a 3D volume. Electrostatic effects and van der Waals forces also play a role in the behavior of static and flowing regolith.

Table 11.3 Bulk elemental chemistry (in wt.%) for representative samples of regolith or meteorites

Oxide	Lunar Highlands ^a	Lunar High-Ti Mare ^a	Mars ^b	Carbonaceous Chondrite (C-type asteroid) ^c	Ordinary Chondrite (S-type asteroid) ^c
SiO ₂	45.5	40.6	43.0	29.1	39.5
TiO ₂	0.32	10.8	1.19	0.13	0.1
Al ₂ O ₃	28.6	9.67	9.43	2.15	2.5
Cr ₂ O ₃	0.10	0.27	0.49	0.48	0.54
FeO	4.25	18.0	19.2	28.3	19.1
MnO	0.06	0.29	0.41	0.20	0.3
MgO	4.38	7.05	8.69	19.9	24.9
CaO	16.4	12.4	7.28	1.89	1.9
Na ₂ O	0.41	0.43	2.72	0.24	1.0
K ₂ O	0.06	0.08	0.49	0.04	0.1
P ₂ O ₅	0.01	–	0.94	0.23	0.27
SO ₃	–	–	5.45	6.59	5.79
Fe (m)	–	–	–	0.1	6.9
Ni	–	–	0.05	–	1.41

^a Taylor and McLennan (2009)^b Blake et al. (2013)^c Jarosewich (1990)

The structure and properties of lunar regolith are better known than regolith on Mars or asteroids because of the in-situ coring, trenching and drilling carried out by the Apollo missions, and extensive study of the subsequently returned samples (Table 11.4). Typically, regolith on the Moon is loose at the surface and then becomes compressively more compacted at depth as shown in Fig. 11.5 (Heiken et al. 1991). The regolith on the Moon has been mixed by billions of years of high-energy impacts from asteroids and comets so that the grain size distribution is fairly uniform across the Moon's surface. On Mars, the regolith is more diverse due to the different geological history of the planet, where there is an atmosphere and evidence indicates that there were large oceans in its early history (Carr and Head 2003). The Martian surface regolith ranges from sand dunes to consolidated sedimentary and other rock surfaces which makes excavation more challenging. Asteroids have a large variety of surfaces ranging from fine-grained regolith to loosely bound cobble-like materials with 1/1000 Earth's gravity.

Table 11.4 Summary of bulk regolith properties taken as representative of typical lunar characteristics based on prior landed missions and sample properties (Roberts (2015), NASA SLS-SPEC-159, Cross-Program Design Specifications for Natural Environments (DSNE))

Property	Value	Units	Notes	DSNE selection	Sources
Bulk density (ρ)	1.58 \pm 0.05: 0–30 cm	g cm ⁻³	Inter crater areas	3.4.2.3.1	Carrier et al. (1991)
	1.74 \pm 0.05: 30–60 cm				
Relative density (D_R)	74 \pm 3: 0–30 cm	%	Inter crater areas	3.4.2.3.2	Carrier et al. (1991)
	92 \pm 3: 30–60 cm				
Specific gravity (G)	3.1	–	Based on limited number of bulk samples	3.4.2.3.3	Carrier et al. (1991)
Porosity (n)	49 \pm 2: 0–30 cm	%	Calculated	3.4.2.3.4	Carrier et al. (1991)
	44 \pm 2: 30–60 cm				
Void ratio (e)	0.96 \pm 0.07: 0–30 cm	–	–	3.4.2.3.4	Carrier et al. (1991)
	0.78 \pm 0.07: 30–60 cm				
Permeability (Q)	1–7 \times 10 ¹²	M2	Firing of Surveyor vernier engines on surface	3.4.2.3.5.1	Choate et al. (1968)
Diffusivity	7.7 He, 2.3 Ar, 1.8 Kr	Cm ² s ⁻¹	Measured on simulant function of gas species	3.4.2.3.5.2	Martin et al. (1973)
Friction angle (φ)	30–50	°	–	3.4.2.4.6	Carrier et al. (1991)
Cohesion (c)	0.1–1	kPa	–	3.4.2.4.7	Carrier et al. (1991)
Compression index (C_c)	0.3: loose	–	Lab measurement on 1.2 to 200 g samples	3.4.2.4.2	Langseth et al. (1973)
	0.05: dense				
	0.01–0.11 range				

(continued)

Table 11.4 (continued)

Property	Value	Units	Notes	DSNE selection	Sources
Recompression index (C_r)	0.003: avg	–	Lab measurement on 1.2 to 200 g samples	3.4.2.4.2.1	Carrier et al. (1991)
	0–0.013: range				
Coefficient of lateral stress (K_θ)	0.45: normally consolidated	–	Lab measurement on 1.2 to 200 g samples	3.4.2.4.3	Carrier et al. (1991)
	3–5: below a few meters				
	0.7: recompacted				
Modulus of subgrade reactions (k)	8: avg 1–100: range	kPa cm ⁻¹	Based on in-situ observations of boot prints	3.4.2.4.5	Carrier et al. (1991)

11.2.4 Other Regolith Properties

There are many other properties of regolith that can be measured either in situ or from returned samples. Table 11.5 (adapted from Metzger et al. 2015) provides a detailed list of these properties in addition to those discussed above. High-quality measurements do not currently exist for all properties for each type of planetary body, but some of these may be important to constrain through further exploration and laboratory or modeling studies in order to test certain regolith processing technologies.

11.3 Excavation and Transport

In order to use regolith, it must first be acquired by an excavation (or digging) device and then the regolith payload must be transported to an end user or destination for stockpiling or further processing.

Each regolith environment must be individually considered before a custom excavation solution can be developed (Fig. 11.6). Excavation in space will be a robotic endeavor, where machines will be teleoperated or will feature semi-autonomous or fully autonomous control systems. It is not likely the excavators will be driven by humans as they are in most terrestrial mines today, since the environment is too extreme with high radiation, large thermal swings, vacuum, and micrometeorite impact dangers. The astronauts will be busy with other, higher-priority tasks. The excavators are also much smaller than those used on Earth due to mission mass constraints and mission requirements for in-situ resource utilization (ISRU) propellant production. Excavation for surface construction purposes will also be necessary but is not addressed here, and excavation for icy volatiles (e.g., water) is also not addressed since their properties and environments are not known yet.

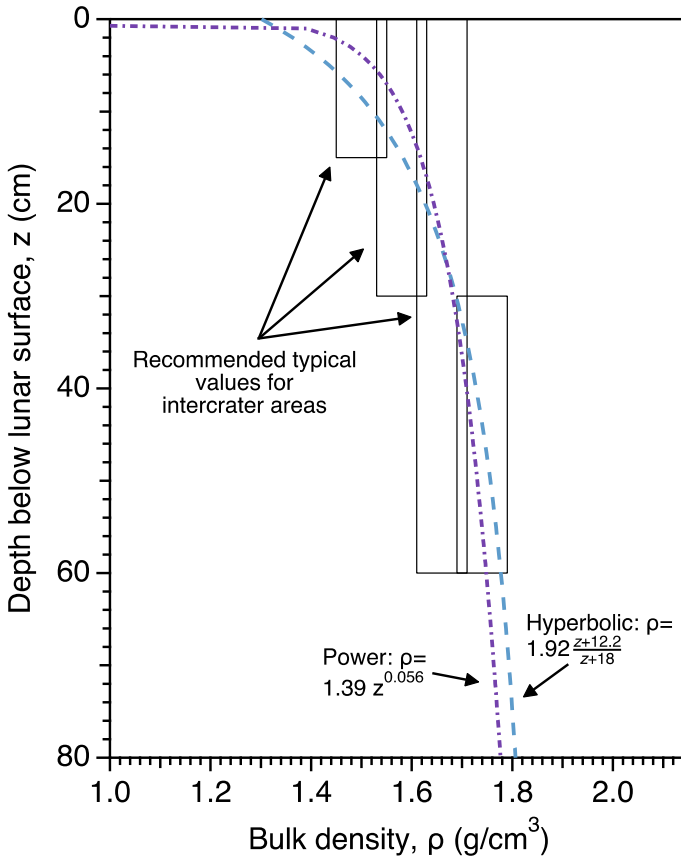


Fig. 11.5 Plots of bulk density as a function of depth in the lunar soil layer based on data from core tube samples and detailed studies of soil samples. The soil, although less dense near the surface (<10 cm deep), quickly becomes dense to very dense with depth (>20 cm). Modified from Heiken et al. (1991), *Lunar Source Book* p. 499

The excavators may be stationary using drag line systems or large boom systems, or they may be mobile with appropriate digging implements attached (Fig. 11.7).

Mobility is more versatile so it is likely the excavators will be mobile, but there are cases where steep slopes, soft terrain and other hazards may require drag lines or other excavation systems such as pneumatic excavation via tubes or hoses. In contrast with the sophisticated science rovers that roam Mars today (e.g., Opportunity, Curiosity and Perseverance) the future robotic regolith excavators will be simple, robust and tough. Sensor systems should be minimized to avoid potential failure modes and mechanisms should be dust tolerant, with relatively high design margins for contingencies and redundancy. Examples of mobile excavator types with a notional mobility platform are shown in Fig. 11.8 (Mueller et al. 2008).

Table 11.5 Comprehensive list of regolith properties. Modified from Metzger et al. (2015)

<p>1. Geomechanical & Grain Properties</p> <ul style="list-style-type: none"> a. Particle size distribution <ul style="list-style-type: none"> i. Mean particle size ii. Broadness of size distribution iii. Minimum and maximum particle size b. Particle shapes c. Specific surface area d. Strength (tensile/shear) e. Angle of repose/internal friction f. Compressibility/compactibility g. Abrasivity 	<p>2. Compositional Properties</p> <ul style="list-style-type: none"> a. Mineralogy <ul style="list-style-type: none"> i. Crystalline components ii. Amorphous components b. Bulk chemistry <ul style="list-style-type: none"> i. Mg/Fe ratio of silicate minerals ii. Redox state of iron iii. Superoxide species c. Organic content <ul style="list-style-type: none"> i. C-to-H ratio (aliphatic/aromatic) ii. Toxicity d. Isotopic ratios
<p>3. Physical Properties</p> <ul style="list-style-type: none"> a. Bulk and grain density b. Porosity and permeability c. Thermophysical properties <ul style="list-style-type: none"> i. Heat capacity/conductivity ii. Thermal cracking behavior 	<p>4. Volatile Properties</p> <ul style="list-style-type: none"> a. Total volatile content <ul style="list-style-type: none"> i. Water and hydroxyl ii. Organic molecules iii. Sulfur compounds b. Volatile release pattern (ambient and vacuum)
<p>5. Optical Properties</p> <ul style="list-style-type: none"> a. Reflectance spectrum b. Thermal emission spectrum c. Radar reflectivity 	<p>6. Chemical Reactivity</p> <ul style="list-style-type: none"> a. Absorptive capacity for volatiles <ul style="list-style-type: none"> i. Deliquescence b. Susceptibility to oxidation c. Reaction to surface damage
<p>7. Texture</p> <ul style="list-style-type: none"> a. Homogeneity and isotopy of texture b. Monomineralic vs. polymineralic grains c. Presence of agglutinates, chondrules, etc 	<p>8. Magnetic Properties</p> <ul style="list-style-type: none"> a. Magnetic susceptibility
<p>9. Electrostatic Properties</p> <ul style="list-style-type: none"> a. Tribocharging b. Bulk resistivity c. Conductance 	<p>10. Aerodynamic Properties</p> <ul style="list-style-type: none"> a. Erodibility by gas (rocket exhaust) b. Particle coefficient of drag

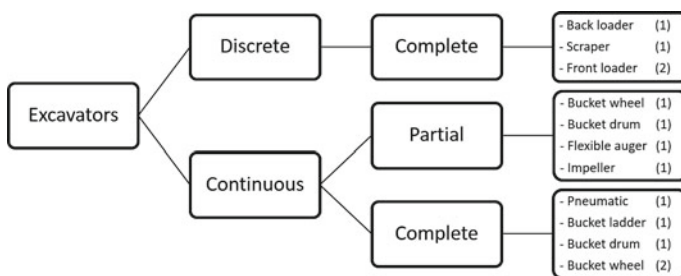


Fig. 11.6 Classification method for 13 investigated regolith excavator systems. The number in brackets indicates the number of systems reviewed per excavation method. *Credit* Just et al. (2020)

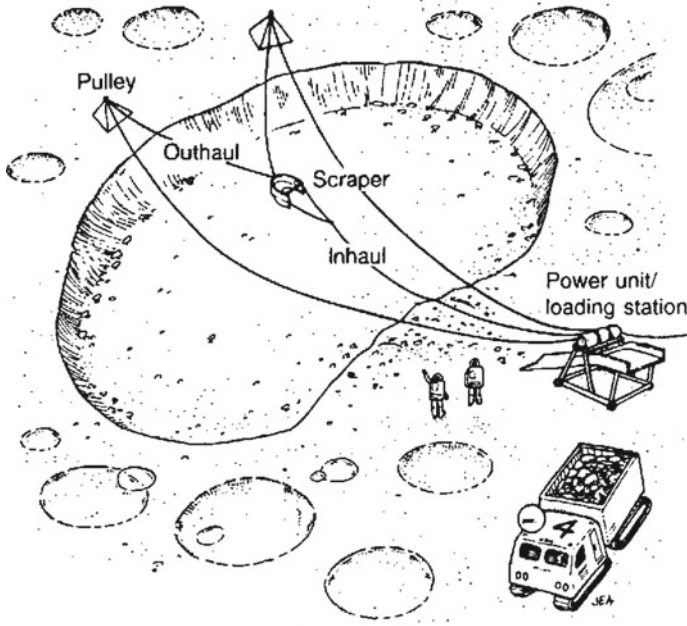


Fig. 11.7 Three-drum slusher system concept using draglines. Credit Gertsch (1983)

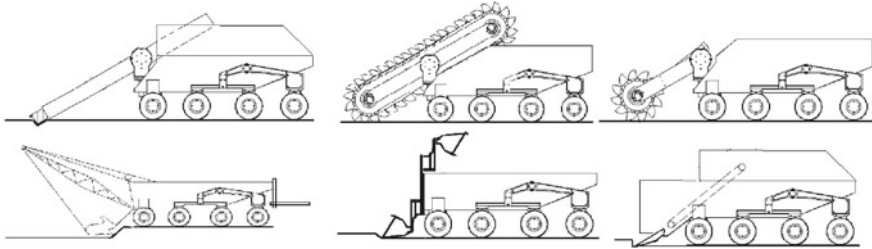


Fig. 11.8 Regolith excavation system concepts: From left to bottom right: auger, bucket ladder, bucket wheel or bucket drum, dragline, overshot loader and scraper. Credit Mueller et al. (2008)

Since 2010, NASA has hosted the “Lunabotics” Robotic Mining Competition (RMC) at Kennedy Space Center for university student teams. It involves a head-to-head excavation competition where teleoperated or autonomous robots traverse an obstacle course of rocks and craters and then regolith simulant is excavated and hauled back to the starting point, where it is deposited in a hopper for mass measurements. The team that is able to deposit the most regolith simulant in 15 min and meet a set of other requirements related to size, dust tolerance, safety, systems engineering, innovation and design is then declared to be the winner. Over the last ten years data has been collected for the various solutions chosen by the teams and these are listed

in Tables 11.6, 11.7 and 11.8 (Mueller et al. 2021). Since the competition rules bias the designs towards rapid excavation in just 15 min and long-term lifetime is not considered, the most popular designs may not be appropriate for an actual space mission, but the data do give an indication of the diversity of designs, mechanisms and mobility that are possible. Transport of regolith can involve small distances such as movement of the acquired regolith at the end effector to the storage mechanism onboard the excavator, or it can refer to hauling the excavated regolith to the destination with a dumping operation included. Examples of various mechanisms used in the RMC are also shown in the tables.

While actual excavation in space has been limited to small scoops such as those used in the NASA lunar Surveyor mission (Agui et al. 2013), the Soviet Luna 16, 20 and 24, NASA Mars Viking (Younse et al. 2009) and NASA Mars Phoenix (Goldstein and Shotwell 2008) missions, future needs will require larger excavation capabilities. Some typical key performance parameters (KPP), state of the art (SOA) and figures of merit (FOM) for excavation to support ISRU propellant initial production are shown in Table 11.9.

As shown in Fig. 11.9, low-gravity excavation can be categorized as discrete or continuous excavation. Testing and analysis have shown that discrete excavation

Table 11.6 Most popular RMC excavation and regolith transportation mechanisms

# sys	Excavation mechanism	# sys	Transportation mechanism
101	Bucket ladder	103	Bucket ladder
37	Front-end loader	40	In scoop
29	Bucket belt	22	Conveyor belt
27	Bucket wheel	21	Bucket belt
17	Bucket drum	15	Auger
15	Snow blower (auger or brush)	11	Over shoulder dump into hopper
12	Auger	8	Chute for guiding regolith
8	Backhoe	7	Bucket drum
8	Bulldozer	7	Drum
8	Scraper	6	Bucket wheel
7	Large single scoop	6	Impeller
4	Dual auger	4	Bucket rim
4	Dual bucket wheel	4	Bucket wheel discharge through bottom
4	Rotating brush	4	In bucket
3	Excavating wheels	4	Rotate scoop to slide simulant in hopper
2	Claw/gripper scoop	3	Throw from impeller
2	Dual bucket ladder	2	Bucket wheel with side discharge
2	Dual counter-rotating bucket drums	2	Paddle conveyor
2	Large bulldozer scoop	2	Raising scraper with chute
2	Paddle conveyor	2	Thrown from brush up ramp

Table 11.7 Most popular RMC regolith storage and regolith dumping mechanisms

# sys	Storage mechanism	# sys	Dumping mechanism
213	Hopper	111	Rotating tilting hopper
41	In scoop	36	Conveyor belt as bottom and inclined side
10	Drum	30	Scoop tilting
6	Bucket drum	7	Auger
6	On conveyor belt	6	Counter-rotate bucket drum
4	Auger	6	Raising/tilting hopper/scissor lift
3	Scraper	5	Conveyor belt
2	In bucket	5	Fixed rotating hopper
1	Bucket drums	4	Raising hopper with back chute
1	Bucket ladder	4	Rotate and lift scoop to slide off back
1	Bulldozer	4	Scissor lift and tilting hopper
1	Drums	4	Tilted raised drum
1	In auger pipe	3	Bucket ladder
1	In clamshell	3	Raising counter-rotating drum
1	Inside tube body	3	Raising hopper with bottom conveyor belt
1	Large conveyor belt with crazy carpet	3	Tilting raising scoop
1	Saddle hopper (two sides)	3	Tilting scoop
1	Scraper scoop	2	Angled vibrating hopper
1	Side hopper	2	Chute
1	Slide	2	Horizontal conveyor belt

(e.g., front-end loader bucket) will stall as the load of regolith increases. Excavation resistance increases with increasing payload for discrete excavation, but is bounded for continuous excavation (e.g., bucket drum excavator). Continuous excavation maintains mobility and high productivity in offloaded gravity. When small scoops of regolith are continuously taken and stored as payload, then the summation of regolith in these scoops exceeds that of discrete systems at a constant excavation force, as shown in Fig. 11.9 (Diaz et al. 2012; Skonieczny et al. 2016).

An example of a continuous lunar regolith excavator being developed by NASA Swamp Works is the Regolith Advanced Surface Systems Operations Robot (RASSOR) which will meet the mission needs of ISRU for NASA's future lunar missions and the KPPs shown above. The design consists of a counter-rotating dual-bucket drum excavator (Fig. 11.10). The continuous excavation from the bucket drum has a low reaction force vector, but even this low force can become excessive as the lunar regolith bulk density increases when it is more compacted at depth with low gravity. The counter-rotating bucket drum design addresses this issue, where the horizontal component reaction forces are horizontally opposing so that they cancel

Table 11.8 Most popular RMC robot mobility mechanisms

# sys	Movement mechanism
173	4 fixed wheels
73	Tracks
21	6 fixed wheels
10	4 steerable wheels with custom profile
6	2 auger drums to propel
5	Stationary with swivel
4	4 fixed track wheels
3	4 digging wheels
2	3 wheels (2 driven, 1 steering)
2	4 six-legged wheels
2	4 wheels with suspension
2	Each of two robots has 4 fixed wheels with grousers
2	4 individual steerable tracks
2	Three robots working together, two transport, one excavator, each with 4 fixed wheels
1	3 fixed wheels (front wheel swivels freely)
1	3 large wheels (2 with grousers, third with scoops)
1	4 medium and 2 large front wheels
1	4 wheels (2 steerable coupled) with grousers
1	4 wheels with grousers, 2 of which have buckets to fill with regolith to increase counterweight
1	4 wheels, of which 2 steerable rear wheels

each other. The vertical reaction force is also reduced or eliminated through the self-anchoring action of each curved scoop as it penetrates the regolith surface. These combined effects allow the RASSOR excavator to be very light (<100 kg) while still being able to dig deep for trenching and cutting below grade on the Moon or Mars.

An analysis using lunar excavation criteria previously established (Skonieczny et al. 2014) indicates that the NASA RASSOR is the current state-of-the-art lunar excavator design (Smith et al. 2019) that is under development and it is a candidate for a NASA lunar technology demonstration mission. The RASSOR 2.0 excavator prototype can travel at a maximum velocity of 44 cm/s and has a payload ratio of 1.34, which are the lunar excavation criteria that affect production the most. Its sensitivity to unknown regolith properties (especially at the lunar poles) is very low, making it a robust and efficient choice for lunar excavation activities.

As excavation forces go up, then excavator mass and traction must also rise to provide the necessary vertical and horizontal reaction forces. Since the excavation force is proportional to the depth of cut (Zacny et al. 2010), then a method of reducing excavation forces is to take many smaller-depth cuts, which can be achieved with a

Table 11.9 Excavation for Lunar ISRU

KPP	SOA	Threshold	Goal	FOM
Amount of regolith moved (surface regolith only)	Surveyor Scoop, Phoenix Scoop <10 kg	5 mT	10 mT	Tonnes (mT) of regolith
Total distance traveled	46 km (Mars Opportunity Rover)	75 km	150 km	Kilometers traveled on lunar surface
Repeated Trafficking	Apollo Sorties deepened loose regolith <5 X	350 X	700 X	Count of repeated trips (X)
Recharge cycles	None 0 X	20 X To Be Reviewed (TBR)	40 X (TBR)	Number of recharges (X)
Lunar days of operation (28 Earth days = 1 lunar day)	0 (Excavator) Yutu (Chinese lunar rover)	>1	Support ISRU Pilot Plant Needs	Lunar Day (28 Earth Days)
Amount of regolith moved (down to 1 m below grade)	Lunar core drills	10 mT	100 mT (NASA Artemis Reusable Landers)	Tonnes (mT) of regolith
Total distance traveled	46 km (Mars Opportunity Rover)	750 km (TBR depends on ISRU architecture)	1500 km (TBR depends on ISRU architecture)	Kilometers traveled on lunar surface
Recharge cycles	None 0 X	>20 X (TBR)	>40 X (TBR)	Number of recharges (X)
Autonomy range	Supervised autonomy 46 km in 10 years (Mars Opportunity Rover)	500 m (TBR)	1 km (TBR)	Kilometers away from lander
Regolith size sorting & beneficiation	Mars—sieves Moon—none	Remove rocks >5 mm diameter No Beneficiation	Customized size distribution profile <1 mm and customized minerology	Regolith particle size and minerology (mm)

bucket wheel or bucket drum system. Other mechanisms such as a bucket ladder can also achieve this outcome, but the reliability of all alternatives must be considered and traded off against each other to create the best low-gravity excavation solution. Other low reaction force excavation methods that have been considered are pneumatic excavators and percussive blade excavators; however, percussive blade systems are discrete with inherent limitations as the payload ratio rises, and pneumatic excavators require a consumable gas. Larger excavators would provide higher efficiency excavation due to scaling effects (Zacny et al. 2010), but transporting these large

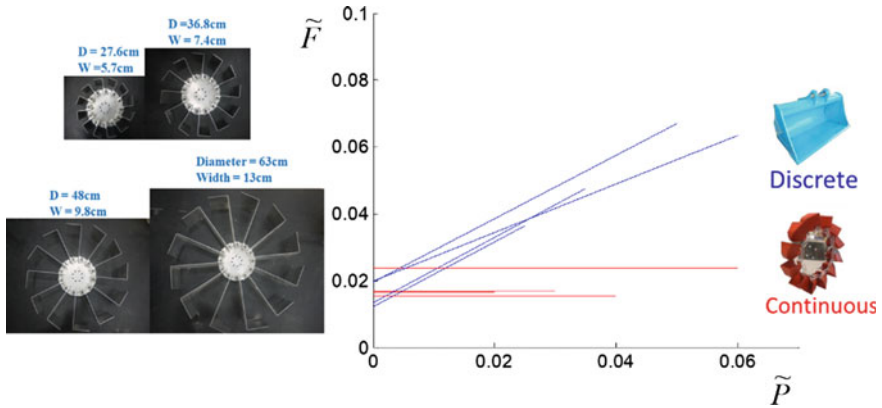


Fig. 11.9 Effect of bucket wheel scale on excavation forces and soil motion (Diaz et al. 2012) showing how force (F) varies with regolith payload ratio (P). Credit Skonieczny et al. (2016)



Fig. 11.10 NASA Regolith Advanced Surface Systems Operations Robot (RASSOR) Excavator 2.0 Prototype during testing. Credit NASA Kennedy Space Center, Swamp Works

and heavy robots to the Moon is currently prohibitively expensive and logistically difficult. The future capability of the SpaceX Starship to deliver 100–200 tons to the lunar surface may alleviate these concerns.

11.4 Regolith Separation

For many applications it may be beneficial or even necessary to separate out different parts of the regolith for use in further processing steps. Various techniques exist or can be adapted from terrestrial practices in order to target specific particle size fractions, particle compositions, or particles that differ in terms of the other properties listed in Table 11.5. An overview of separation techniques is provided here, but readers are also directed to Rasera et al. (2020) who provide a detailed review of separation techniques and experimental results.

11.4.1 Separating Particle Sizes

Regolith tends to be very poorly sorted (or well graded), such that a random sample of regolith will contain grains that span several orders of magnitude in diameter. Some processing techniques may work best with a specific size fraction of grains, in which case that fraction must be efficiently separated out from the rest. For example, gravel and cobbles could be removed from finer particles and spread over the ground to create a prepared surface that mitigates against dust. The fines could then be further separated to isolate the clay and silt fractions, best suited for 3D printing techniques. Finally, the leftover sand fraction could be melted down and used for oxygen extraction, a process that is not highly sensitive to particle size.

In terrestrial mineral processing, sieving methods are the most basic way to sort a mixture of grains into specific size ranges. However, gravity-driven sieving is likely to be slow or ineffectual in low-gravity environments, particularly the Moon and asteroids. Additionally, the higher cohesion caused by fractal-shaped agglutinates in lunar regolith may lead to significant challenges for passing grains through sieves, particularly at smaller particle size fractions when using dry sieving.

Grizzly screens or grizzly bars (Fig. 11.11) are likely to be effective for removing larger cobbles embedded in excavated regolith, even in reduced-gravity environments.

For smaller sizes, Dreyer et al. (2012) invented a centrifugal sieve (Fig. 11.12) that gets around the issue of low gravity; this type of device has since been patented (Walton et al. 2015), and prototypes have been shown to work effectively down to 0.1 mm using JSC-1A lunar simulant.

For sizes much finer than this, it may be necessary to move away from sieving-based techniques and use other methods to separate particles on the order of tens of microns. For example, Adachi et al. (2017) describe a system that uses an electrostatic traveling wave to separate out particles less than 20 microns in diameter. Methods that rely on attractive forces (magnetic, electrostatic, etc.) can be tuned to move only the finest particles, although this may have the consequence of also separating by composition, which could be beneficial or not.



Fig. 11.11 Example of a grizzly screen used to filter out larger rock sizes. Photo by Peter Craven, Creative Commons CC-BY-2.0

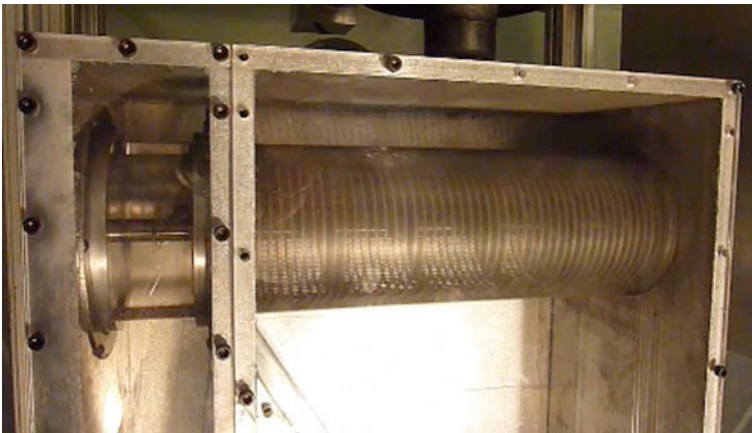


Fig. 11.12 Centrifugal sieve developed by Dreyer et al. (2012). *Image credit* C. Dreyer

11.4.2 Separating Mineral and Chemical Components

Some regolith processing techniques are only effective for certain types or chemistries of particles, such that beneficiation is required to enrich those grains from the rest of the regolith. Examples include: (1) hydrogen reduction of ilmenite for lunar applications requires ilmenite-bearing grains to be highly enriched in order to achieve high

yields of oxygen; (2) metal production from regolith will lead to mixed or mongrel alloys, unless the cation content is tuned by separating out different minerals, for example iron-rich silicates and oxides to form iron-rich alloys; and (3) water extraction from hydrated minerals (clays, sulfates) on Mars will be more effective if those minerals can be separated from the anhydrous basaltic silicate minerals they are mixed with.

Separation techniques make use of the fact that grains differing in mineralogy and/or chemistry will also tend to differ in their derived properties (Table 11.5). Density separation can separate high-density grains like metals or oxides from lighter silicates, although as with sieving this will be less effective on low-gravity bodies. Terrestrial density separation commonly employs fluid mediums, and this would have to be done inside of sealed vessels on planets that lack a significant atmosphere. Adding a centrifugal component could compensate for the low gravity on the Moon and at asteroids.

Magnetic separation can be used to sort grains based on their magnetic susceptibility, which tends to be much higher for metals and metal oxides than for silicate minerals. With high enough power, electromagnets could also separate different silicates, which may only have small differences in susceptibility. Figure 11.13 shows an example of using a terrestrial magnetic separator on a Martian regolith simulant, with some degree of composition-based sorting achieved with a single pass through the system.

In a similar vein, electrostatic separation works by applying an electrostatic charge to grains, which are affected differently based on their electrical conductivity. Triggwell et al. (2013) report on electrostatic beneficiation of lunar simulants and Apollo soils, where mineral enrichments >100% were achieved.

A word of caution is warranted regarding composition-based separation. The lunar regolith grains in Fig. 11.2 have been carefully picked out and cleaned, but regolith in its native form is a phenomenally messy and complex natural material. Almost all grains are coated with strongly clinging fines, different phases are welded together in the same grains, and impurities are the norm. One should not assume that mineral phases, particularly metals or ices, are present as large clean grains that can be plucked out of the regolith like marbles. Beneficiation is likely to be challenging, and this difficulty may not be captured by experiments using terrestrial simulants.

11.5 Binding Regolith Particles

Impact cratering and thermal cycling have done a tremendous amount of geological work to pulverize planetary crusts into fine powders. However, some applications—namely construction—seek to build consolidated structures, in which case the granular regolith must be imbued with strength again. There are many options that have been explored for doing so, each of which has different strengths and weaknesses as described below.

Fig. 11.13 Example of a magnetic separation device for granular materials. *Image credit* K. Cannon



11.5.1 Sintering

Sintering is a process of forming coherent materials by fusing granular material together at the grain boundary contacts. The most common and widely explored technique is liquid-phase sintering, where a granular material is heated in a furnace to some fraction of its melting temperature (commonly ~80%), and incipient melting at grain asperities and along grain boundaries fuses the particles together. Liquid-phase sintering of regolith has been explored in the context of landing pads (Hintze 2010), dust mitigation (Wilson and Wilson 2005), construction (Meurisse et al. 2018; Romo et al. 2018) and regolith shielding (Meurisse et al. 2020). Advantages of sintering include potentially high strength of the final products, low processing times, and a lack of consumables brought from Earth. Potential disadvantages include high energy requirements, and challenges in scaling up from small tiles or test coupons that have been traditionally used in lab experiments. Sintering may also not reach maximum density, and partially sintered materials tend to be crumbly and weak. Most testing on regolith simulants has been done in air in an enclosed furnace; future testing should focus on vacuum sintering for the Moon (e.g., Song et al. 2019), and on sintering mechanisms that do not require heavy enclosed furnaces.

Other types of sintering may show promise for regolith applications: these include cold sintering that makes use of a fluid, and microwave sintering that may be much more effective for lunar regolith (Taylor and Meek 2005), although this technique has not matured much since the early 2000s. In addition, there is active research on using solar concentrated light (Imhof et al. 2017), and lasers (Mueller et al. 2014) as the source of energy for the heating required to sinter regolith particles. The high energy usage concerns still exist with microwave and laser systems, but solar concentrated light could provide a low-energy alternative if the heat transfer can be achieved in a vacuum environment with the highly thermally insulating regolith. However, reflectivity at certain wavelengths is a concern, particularly when the upper surface begins melting. All of these methods require development of a robotic positioning control system which adds further complexity.

11.5.2 Geopolymers

Geopolymers are a type of consolidated material formed by the reaction of highly alkaline fluids with natural minerals and rocks (Davidovits 2005). The action of the alkaline cations in the fluid (usually some combination of sodium or potassium hydroxide, and sodium or potassium silicate) promotes reactions with the minerals that are present to form an interconnected network of chemical bonds which hold the grains together. Geopolymers are similar to and are often conflated with the types of reactions that take place in concretes, but Davidovits (2013) clarifies the specific differences between the two. Geopolymers have been explored for regolith applications, for example by Montes et al. (2015) and Alexiadis et al. (2017). Because the elements in the regolith grains take part in the geopolymer reaction, fewer consumables need to be transported from Earth compared to traditional concrete processes. However, geopolymers use a fluid and a long curing process, which are significant challenges for airless or low-pressure environments like the Moon and Mars. Additionally, the reactions are quite sensitive to the composition of the grains themselves, such that fine tuning would be needed based on the local geology.

11.5.3 Regolith Concrete

Concrete involves using an aggregate and an interstitial cement that binds the aggregate grains together in order to create a strong, solid material. For off-world construction, regolith constitutes a convenient and abundant aggregate that would only require minimal processing to adjust the particle size distribution. The issues with regolith concrete have always come down to the cement and curing. If traditional water-activated cements are brought from Earth, this represents a large amount of additional binder mass that comes with extremely high costs. Various efforts have been made to come up with ways to source cements locally, for example using molten sulfur (Omar

and Issa 1994; Khoshnevis et al. 2016). However, there are significant performance concerns with sulfur concrete in the lunar environment (e.g., Grugel and Toutanji 2008), and contrary to what others have suggested, there are no obvious sources of extractable elemental sulfur on the Moon or elsewhere. Others have proposed using biological materials including urea (Pilehvar et al. 2020) and bovine serum albumin (Roedel et al. 2014) as possible cements for regolith concrete. The ostensible advantages of regolith concrete are the high strength achieved in resulting materials, the low energies involved, and the familiarity from thousands of years of terrestrial construction experience. However, concrete activated with water must cure over a period of days to weeks in a damp environment: this is not practical in the vacuum of space, except in an enclosed environment. Pre-fabricated concrete sections could be produced and cured in a pressurized facility as a potential workaround.

11.5.4 3D Printing with Regolith

Automated additive construction, and 3D printing specifically, have been proposed as a way to process regolith into large, complex structures on the Moon and Mars. The basic process uses a print head to deposit layers of regolith or a regolith mixture directly on the planet's surface, with a way of translating the print head in the XY plane. Multiple types of 3D printing with regolith have been explored. A simple solution is to take traditional plastic filaments such as poly-lactic acid (PLA) that are used in terrestrial 3D printing and use regolith as a filler to reduce the amount of filament that needs to be brought from Earth. The typical ratio of thermoplastic polymer binder to regolith is 1:0.25–0.3 and this has been shown to create good structural materials with tensile strengths up to 96.5 kPa (14,000 psi), (Mueller et al. 2016, 2018) when the regolith is used in the form of basalt glass fibers. Other potential solutions involve extruding a damp regolith slurry that cures after deposition, and methods to deposit powdered regolith layers then go back over those layers with a technique such as microwave or solar sintering to fuse the regolith in place. Potential advantages of 3D printing with regolith include relatively high speed, low energy, and scaling: of all the binding techniques described here, 3D printing has the clearest path to creating large and intricately shaped buildings with a straightforward concept of operations (Fig. 11.14). However, 3D printing large structures on Earth is still in its infancy, and significant development and testing will be needed to refine the best ways of depositing and fusing regolith. A size separation step will likely be needed to deliver only the finest fraction of the regolith to the print head; for example, in selective laser sintering a narrow size range of $50 \pm 10 \mu\text{m}$ is required. Agglutinate-like materials can be formed in a solar printing process and up to 50% “virgin” powder must be added to reduce their effect. When such a large fraction of lunar material is agglutinate, a pre-processing step besides size sorting may be necessary (Williams and Butler Jones 2019). Conveying the regolith and binder materials to the print head on a robotic positioning device still remains a challenge to be solved in a lunar or Martian environment. On Earth, feedstock conveying has been achieved with



Fig. 11.14 A PLA thermoplastic polymer/regolith material being 3D printed at the NASA Centennial Challenge competition for “3D Printing a + Habitat” by the winners: A.I. Space Factory. *Credit NASA*

blended thermoplastic/regolith aggregate pellets that are pneumatically conveyed to the print head, or concrete slurry is pumped through a hose to the print head.

11.6 Extracting Resources from Regolith

Regolith commonly hosts constituents that can be used as space resources, but the raw regolith has to undergo processing to extract those materials. This processing can range in complexity and energy from simple heating to liberate water and other volatiles, to complete melting and chemical reduction in order to extract metals and gases at the atomic level from inside mineral crystal structures.

11.6.1 *Extracting Water*

Water is present within and buried under regolith in the form of ice, as well as hydrated minerals like clays and sulfates on Mars where water molecules are part of the mineral structures themselves. Regolith will have to be processed to extract this water, where it would then be passed on to an electrolysis unit (to create cryogenic propellants), or be used directly in life support systems. Regolith processing techniques for extracting

water can be divided broadly into two categories: vessel-based heating, and direct or in-situ mining. In vessel-based heating, the regolith and water-bearing phase(s) are excavated together and hauled to a central processing plant, where they are heated inside a sealed vessel to drive off the water as vapor or liquid. For ice in Martian or lunar regolith, energy requirements are quite modest. To drive water off sulfates or clay minerals on Mars, higher temperatures of hundreds of degrees or more will be required (Abbud-Madrid et al. 2016). In direct or in-situ extraction (Sowers and Dreyer 2019), energy is applied directly to the regolith to extract water vapor, which is then collected in some type of cold trap. Methods for applying energy include microwave, visible (solar), and thermal infrared radiation, as well as conducting rods inserted directly into the regolith.

11.6.2 Extracting Oxygen and Metals

As described above, the major mineral constituents of regolith on rocky planetary bodies are silicate minerals, including plagioclase, pyroxene and olivine. These minerals are made up of different cations (Si, Mg, Fe, Ca) strongly bonded with oxygen in a crystal structure. While oxygen may make up ~40–50% of the regolith by mass, it is not easy to extract from the solid minerals. Various high-temperature processes have been proposed and tested to extract oxygen directly from regolith, and many of these same techniques can also produce metal as a byproduct or the primary target product. Schlüter and Cowley (2020) and Schwandt et al. (2012) provide excellent reviews of the specific processes for O₂ extraction from regolith, so these are only mentioned briefly here. They include hydrogen reduction of ilmenite (Fig. 11.15), carbothermal reduction, molten regolith electrolysis, and the FCC-Cambridge or FCC-Metalysis process which uses a molten salt electrolyte. In all these techniques, high temperatures up to and exceeding 2000 °C are combined with chemical reduction power to liberate oxygen from its host minerals. The reducing power can come either from chemical phases (hydrogen, methane, solid carbon, etc.) or from electrolysis.

11.7 Biological Processing of Regolith

A novel and relatively unexplored approach for processing regolith is to harness the power of biological systems (i.e., biomining), some of which are already being employed in the mineral processing industry. This may be especially useful for extracting chemical elements found in trace amounts in regolith as a way to enrich and benefit the resource. Cockell et al. (2020) demonstrated biomining techniques in microgravity on the International Space Station, and other laboratory demonstrations using simulated regolith have been carried out by Volger et al. (2020) and Castelein et al. (2020).

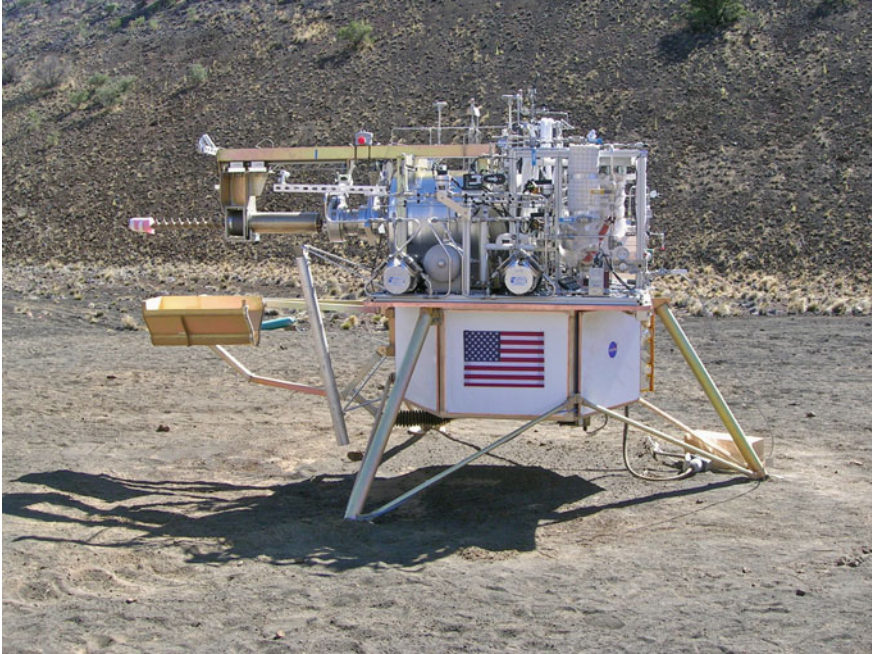


Fig. 11.15 Example of a lunar regolith hydrogen reduction plant prototype for oxygen extraction being tested by NASA in an analog test site in Hawaii. *Credit* NASA ISRU project

11.8 Summary

This chapter described many of the major processes that have been proposed to transform regolith from a simple pile of crushed rock fragments into useful materials, structures, and highly refined end products, all of which constitute space resources. It is useful to understand and appreciate the nature of regolith itself, including its chemistry, mineralogy and physical properties. While there are many commonalities between regolith found on the Moon, Mars and asteroids, there are also important differences. Regolith processing methods may have to be highly tailored to the unique combination of regolith and environment (gravity, atmospheres, temperatures) on these worlds, and testing with appropriately high-fidelity regolith simulants is a must. We described how regolith can first be removed from the ground and transported, which are key initial steps in learning how to work with this material. Various separation processes can then be used to sort the regolith into its constituent grain sizes and compositions, depending on the downstream processing needs. Finally, both simple techniques and advanced or high-temperature processing can be used to transform the beneficiated feedstock into a good material that can be put to use in the space environment.

References

- Abbud-Madrid, Angel, et al. 2016. Mars water in-situ resource utilization (ISRU) planning (M-WIP) study. Report of the Mars Water In-Situ Resource Utilization (ISRU) Planning (M-WIP) Study 90.
- Adachi, Masato, et al. 2017. Particle-size sorting system of lunar regolith using electrostatic traveling wave. *Journal of Electrostatics* 89(2017):69–76.
- Agui, J.H., M. Bucek, A. DeGennaro, R.A. Wilkinson, and X. Zeng. 2013. Lunar excavation experiments in simulant soil test beds: Revisiting the surveyor geotechnical data. *Journal of Aerospace Engineering* 26 (1): 117–133.
- Alexiadis, Alessio, Federico Alberini, and Marit E. Meyer. 2017. Geopolymers from lunar and Martian soil simulants. *Advances in Space Research* 59 (1): 490–495.
- Bish, David L., et al. 2013. X-ray diffraction results from Mars science laboratory: Mineralogy of rocknest at Gale crater. *Science* 341:6153
- Blake, David F., et al. 2013. Curiosity at Gale crater, Mars: Characterization and analysis of the Rocknest sand shadow. *Science* 341:6153
- Carr, M.H., and J.W. III Head. 2003. Oceans on Mars: An assessment of the observational evidence and possible fate. *Journal of Geophysical Research: Planets* 108(E5).
- Carrier III, W.D., G.R. Olhoeft, and W. Mendell. (1991). Physical properties of the lunar surface. In *Lunar sourcebook*, 475–594.
- Castelein, Sofie M., et al. 2020. Iron can be microbially extracted from Lunar and Martian regolith simulants and 3D printed into tough structural materials. bioRxiv.
- Cockell, Charles S., et al. 2020. Space station biomineral experiment demonstrates rare earth element extraction in microgravity and Mars gravity. *Nature Communications* 11(1):1–11.
- Davidovits, Joseph, ed. 2005. *Geopolymer, green chemistry and sustainable development solutions: proceedings of the world congress geopolymer 2005*. Geopolymer Institute.
- Davidovits, Joseph. 2013. Geopolymer cement. A review. Geopolymer Institute, Technical papers 21:1–11.
- Diaz Lankenau, G., K. Skonieczny, et al. 2012. Effect of bucket-wheel scale no excavation forces and soil motion. *Journal of Terramechanics* 49 (6): 341–348.
- Dreyer, C.B., O. Walton, E.P. Riedel. 2012. Centrifugal Sieve for size-segregation and beneficiation of Regolith. In *Earth and space 2012 engineering, science, construction, and operations in challenging environments*, 31–35.
- Dunn, Tasha L., et al. 2010. Analysis of ordinary chondrites using powder X-ray diffraction: 1. Modal mineral abundances. *Meteoritics & Planetary Science* 45(1):123–134.
- Fujiwara, Akira, et al. 2006. The rubble-pile asteroid Itokawa as observed by Hayabusa. *Science* 312(5778):1330–1334.
- Gertsch, R.E. 1983. A method for mining lunar soil, AAS83–234, space manufacturing. In *Advances in the astronautical sciences*, vol 53, eds. J.D. Burke and A.S. Whitt, 337–346.
- Goldstein, B., and Shotwell, R. 2008. Phoenix—the first mars scout mission. In *IEEE Aerospace Conference 2008*, March 1–8, Big Sky, MT.
- Grugel, Richard N., and Houssam Toutanji. 2008. Sulfur “concrete” for lunar applications—sublimation concerns. *Advances in Space Research* 41 (1): 103–112.
- Hartmann, William K. 1973. Ancient lunar mega-regolith and subsurface structure. *Icarus* 18 (4): 634–636.
- Heiken, Grant H., David T. Vaniman, and Bevan M. French. 1991. *Lunar Sourcebook, a user's guide to the Moon*.
- Hintze, P.E. 2010. Building a vertical take off and landing pad using in situ materials. *Space Manufacturing* 14: 29–31.
- Howard, K.T., et al. 2011. Modal mineralogy of CM chondrites by X-ray diffraction (PSD-XRD): Part 2. Degree, nature and settings of aqueous alteration. *Geochimica Et Cosmochimica Acta* 75 (10): 2735–2751.

- Imhof, B., D. Urbina, P. Weissc, and M. Sperld. 2017. Advancing solar sintering for building a base on the Moon. In *68th international astronomical congress (IAC)*, 25–29, Adelaide, Australia.
- Jarosewich, Eugene. (1990). Chemical analyses of meteorites: A compilation of stony and iron meteorite analyses. *Meteoritics* 25(4): 323–337. <https://doi.org/10.1111/j.1945-5100.1990.tb00717.x>.
- Just, G.H., K. Smith, K.H. Joy, and M.J. Roy. 2020. Parametric review of existing regolith excavation techniques for lunar In Situ Resource Utilisation (ISRU) and recommendations for future excavation experiments. *Planetary and Space Science* 180: 104746.
- Khoshnevis, Behrokh, et al. 2016. Construction by Contour Crafting using sulfur concrete with planetary applications. *Rapid Prototyping Journal*
- Langseth, M.G., S.J. Keihm, and J.L. Chute Jr. 1973. Heat-flow experiment. In *Apollo 17 Preliminary Science Report*, pp. 9-1 to 9-24. NASA SP-330.
- Malin, Michael C., and Kenneth S. Edgett. 2000. Sedimentary rocks of early Mars. *Science* 290 (5498): 1927–1937.
- Martin, R.T., J.L. Winkler, S.W. Johnson, and W.D. Carrier. 1973. Measurement of conductance of Apollo 12 lunar simulant taken in the molecular flow range for helium, argon, and krypton gases. Unpublished report. In *Lunar Sourcebook* 1991. Cambridge University Press.
- Metzger, Philip T., et al. 2015. Results of the 2015 Workshop on Asteroid Simulants. In *Earth and space 2016: engineering for extreme environments*, 94–104. Reston, VA, American Society of Civil Engineers.
- Meurisse, Alexandre, et al. 2018. Solar 3D printing of lunar regolith. *Acta Astronautica* 152:800–810.
- Meurisse, Alexandre, et al. 2020. Neutron radiation shielding with sintered lunar regolith. *Radiation Measurements* 132:106247.
- Montes, Carlos, et al. 2015. Evaluation of lunar regolith geopolymer binder as a radioactive shielding material for space exploration applications. *Advances in Space Research* 56(6):1212–1221.
- Mueller, R.P., N.J. Gelino, J.D. Smith, B.C. Buckles, T. Lippitt, J.M. Schuler, and I.I. Townsend. 2018. Zero launch mass three-dimensional print head. In *Earth and space 2018: Engineering for extreme environments*, 219–232. Reston, VA: American Society of Civil Engineers.
- Mueller, R.P., and R.H. King. 2008. Trade study of excavation tools and equipment for lunar outpost development and ISRU. In *AIP Conference Proceedings*, vol. 969, No. 1, 237–244. American Institute of Physics.
- Mueller, R.P., L. Sibille, P.E. Hintze, T.C. Lippitt, J.G. Mantovani, M.W. Nugent, and I.I. Townsend. 2014. Additive construction using basalt regolith fines. *Earth and Space* 394–403.
- Mueller, R.P., et al. 2021. NASA lunabotics robotic mining competition 10th anniversary (2010–2019): Taxonomy and technology review. In *Earth and space 2021: engineering for extreme environments conference*, Reston, VA, American Society of Civil Engineers.
- Mueller, Robert P., et al. 2016. Automated additive construction (AAC) for Earth and space using in-situ resources. In *Proceedings of the fifteenth biennial ASCE aerospace division international conference on engineering, science, construction, and operations in challenging environments (Earth & Space 2016)*. American Society of Civil Engineers.
- Omar, Husam A., and Mohsen Issa. 1994. Production of lunar concrete using molten sulfur. In *Proceedings: 4th Conf. on Space* vol 94.
- Papike, J.J., S.B. Simon, and J.C. Laul. 1982. The lunar regolith: Chemistry, mineralogy, and petrology. *Reviews of Geophysics* 20(4): 761. <https://doi.org/10.1029/RG020i004p00761>.
- Pilehvar, Shima et al. 2020. Utilization of urea as an accessible superplasticizer on the moon for lunar geopolymer mixtures. *Journal of Cleaner Production* 247:119177.
- Rasera, J.N., et al. 2020. The beneficiation of lunar regolith for space resource utilisation: A review. *Planetary and Space Science* 186: 104879.
- Richardson, James E., and Oleg Abramov. 2020. Modeling the formation of the lunar upper megaregolith layer. *The Planetary Science Journal* 1 (1): 2.
- Roberts, B.C. 2015. SLS-SPEC-159 cross-program design specification for natural environments (DSNE) revision D.

- Roedel, H., M.D. Lepech, and D.J. Loftus. 2014. Protein-regolith composites for space construction. *Earth and Space* 2014: 291–300.
- Romo Rodrigo. 2018. Planetary Lego: designing a construction block from a regolith derived feed-stock for in situ robotic manufacturing. In *Earth and space, engineering for extreme environments*, 289–296. Reston, VA, American Society of Civil Engineers.
- Schlüter, Lukas, and Aidan Cowley. 2020. Review of techniques for In-Situ oxygen extraction on the moon. *Planetary and Space Science* 181: 104753.
- Schwandt, Carsten, et al. 2012. The production of oxygen and metal from lunar regolith. *Planetary and Space Science* 74(1):49–56
- Skonieczny, K., D.S. Wettergreen, and W.R. Whittaker. 2016. Advantages of continuous excavation in lightweight planetary robotic operations. *The International Journal of Robotics Research* 35 (9): 1121–1139.
- Skonieczny, Krzysztof, et al. 2014. Productive lightweight robotic excavation for the moon and Mars. *Journal of Aerospace Engineering* 27(4):04014002.
- Smith, J.D., et al. 2019. RASSOR Excavator for ISRU Lunar Mining. In *Proceedings of the space resources roundtable (SRR)/planetary terrestrial mining sciences symposium (PTMSS)*, June 2019, Golden, Colorado.
- Song, Lei, et al. 2019. Vacuum sintered lunar regolith simulant: Pore-forming and thermal conductivity. *Ceramics International* 45(3):3627–3633.
- Sowers, George F., and Christopher B. Dreyer. 2019. Ice mining in lunar permanently shadowed regions. *New Space* 7 (4): 235–244.
- Taylor, S. Ross, and Scott McLennan. 2009. *Planetary crusts: their composition, origin and evolution*, vol. 10. Cambridge University Press.
- Taylor, Lawrence A., and Thomas T. Meek. 2005. Microwave sintering of lunar soil: Properties, theory, and practice. *Journal of Aerospace Engineering* 18 (3): 188–196.
- Trigwell, Steve, et al. 2013. Electrostatic beneficiation of lunar regolith: applications in in situ resource utilization. *Journal of Aerospace Engineering* 26(1):30–36.
- Volger, R., et al. 2020. Mining moon & mars with microbes: Biological approaches to extract iron from Lunar and Martian regolith. *Planetary and Space Science* 184: 104850.
- Walsh, Kevin J. 2018. Rubble pile asteroids. *Annual Review of Astronomy and Astrophysics* 56:593–624.
- Walton, Otis, et al. 2015. *Size-separation of dry granular materials*. U.S. Patent US20160082478A1.
- Williams, Hunter, and Evan Butler-Jones. 2019. Additive manufacturing standards for space resource utilization. *Additive Manufacturing* 28: 676–681.
- Wilson, T.L., and K.B. Wilson. 2005. Regolith sintering: A solution to lunar dust mitigation?. In *36th Annual Lunar and Planetary Science Conference*.
- Younse, P., A. Stroupe, T. Huntsberger, M. Garrett, J.L. Eigenbrode, L.G. Benning, and A. Steele. 2009. Sample acquisition and caching using detachable scoops for Mars sample return. In *2009 IEEE Aerospace Conference*, 1–12, IEEE.
- Zacny, K., R.P. Mueller, J. Craft, J. Wilson, M. Hedlund, and J. Cohen. 2010. Five-step parametric prediction and optimization tool for lunar surface systems excavation tasks. In *Earth and space 2010: Engineering, science, construction, and operations in challenging environments*, 1128–1151.

Chapter 12

Sintering: A Method for Construction of Off-Earth Infrastructure from Off-Earth Materials



Liz Scott and Thao Nguyen

Abstract The establishment of a self-sustaining human presence on planetary bodies depends on the capability to build infrastructure. Infrastructure, such as roads, habitations, garages, and landing pads, will be necessary to protect humans and robots from harsh environmental conditions, including radiation, micrometeorite bombardment, extreme temperature changes, and dust. Due to their large mass and physical size, infrastructure components would be very expensive to launch from Earth. This makes infrastructure an ideal use case for the concept of in-situ resource utilization, where local space resources provide material for use in space, rather than launching them from Earth. The most abundant physical resource on rocky bodies is regolith, the rocks, soil, and dust that comprise the bodies' surfaces. Many different methods of producing infrastructure from regolith are under investigation. Sintering, a thermal process whereby heat is applied to fuse powder particles into a solid material, is a promising technique to produce infrastructure from regolith and is the focus of this chapter. The chapter begins with a discussion of the physics of sintering and current terrestrial sintering techniques. Next is a summary of the properties of lunar and Martian regolith, followed by a summary of the work done so far in applying conventional, microwave, laser, and concentrated solar sintering technologies to extraterrestrial construction. The chapter concludes with a discussion of applications of sintering to off-world construction, including two- and three-dimensional structures and robotic constructors capable of using sintering to build them.

L. Scott (✉) · T. Nguyen
Colorado School of Mines—Center for Space Resources, 1310 Maple St., GRL 234, Golden,
CO 80401, USA
e-mail: eascott@mines.edu

T. Nguyen
e-mail: thaonguyen@mines.edu

12.1 Off-World Construction

The establishment of a self-sustaining human presence on planetary bodies other than Earth depends on the capability to build infrastructure. Planetary bodies such as the Moon and Mars are harsh environments for human and robotic explorers, and infrastructure will be required to mitigate dust, ease surface travel, and provide protection from extreme temperatures, micrometeorites, and radiation. Off-Earth infrastructure will resemble terrestrial infrastructure in many ways. Roads will serve the same purpose on other planets as they do on Earth: to make driving faster, safer, easier, and less dusty. Landing pads will provide safe, dust-free surfaces for incoming and outgoing spacecraft, reducing risk to both the lander and the surrounding environment. Structures such as houses and garages on Earth protect humans and equipment from the elements; on other planets, such structures will provide protection from extreme temperatures, vacuum, radiation, and micrometeorites. Infrastructure is heavy and large on Earth and the same will be true off Earth. The physical size of roads, radiation shielding, landing pads, and other structures make them prohibitively expensive to launch from Earth because high mass translates directly to increased launch cost. Fortunately, there is another option.

In-situ resource utilization (ISRU) is defined as “the collection, processing, storing and use of materials encountered in the course of human or robotic space exploration that replace materials that would otherwise be brought from Earth to accomplish a mission critical need at reduced overall cost and risk” (Sacksteder and Sanders 2007). ISRU is, essentially, the concept of living off the land by producing materials from local resources rather than launching them from Earth. Construction is a perfect application of ISRU. Construction materials are heavy, which makes them expensive to launch from Earth. Producing infrastructure from local materials frees up launch mass and cost for use on items that cannot yet be produced in space, such as electronics, power generation hardware, and humans.

The most abundant physical resource on rocky bodies is the native material that comprises the body’s surface: rocks, soil, and dust, referred to collectively in this chapter as regolith. Researchers have begun to study how regolith might be used to construct landing and launch pads, berms, roads, radiation shielding for habitats, and other infrastructure. Although the functions of infrastructure on other planetary bodies are like those on Earth, the unique constraints of construction in space have led to the consideration of many different construction methods. The first constraint is limited mass of construction equipment due to high launch costs. Terrestrial construction relies on heavy equipment such as cranes and excavators that would be prohibitively expensive to launch to the Moon or Mars. The second constraint is limited power. Terrestrial production of construction materials like concrete requires an enormous amount of power, which cannot be reasonably generated in space yet. The third constraint is limited materials. For example, water, which is an integral component of terrestrial concrete, is limited to certain locations on the Moon and will likely be too valuable as a propellant or for life support to use on building material. The fourth constraint is the unique space environment. The finished construction

material must be able to withstand vacuum or very low surface pressure, impacts from micrometeorites, extreme temperature swings, and radiation.

Many different techniques for off-Earth construction are under consideration, including: the use of binder additives which produce a solid material when sprayed onto or mixed with regolith; processing regolith to produce metals which could be used for various types of construction; and the production of cement and concrete from regolith materials. Each of these methods has benefits, but also significant drawbacks: binder additives require bringing large quantities of material from Earth, and producing metals or concrete from regolith requires tremendous power. Another construction technique that addresses these concerns is sintering.

This chapter will summarize principles of sintering, sintering techniques developed terrestrially, the application of terrestrial sintering techniques to off-world materials, and the integration of sintering devices into the robotic systems that may one day build off-world infrastructure. This chapter is intended as an introduction to the concept of sintering as a method for planetary construction for technologists who are new to regolith sintering, and to introduce research that was performed prior to 2021. The field of planetary construction is evolving rapidly and many technologies that showed initial promise may not, under further testing and characterization, live up to expectations.

12.2 Sintering Fundamentals

Sintering can be defined as “a thermal process used to bond contacting particles into a solid object” (German 2014). This simple definition can be expanded to provide additional clarification. First, sintering involves the *application of heat* to powders/particulates. Heat can be applied in many different ways. Application of pressure may also be part of the sintering process, particularly in compacting the particles before the application of heat. Second, *contacting particles are bonded* together during heating. Sintering is differentiated from complete melting—the mass of material need not reach its melting point or become completely fluid prior to solidification. The powder/particulate can be metallic, ceramic, plastic, or made up of other materials. Third, the loose powders/particulate material is sintered into a *solid object*, although some porosity may remain.

While many other extraterrestrial construction and manufacturing methods have been proposed, sintering of regolith is attractive because it may be possible under the following conditions:

- Relatively little surface preparation or material processing
- No existing infrastructure (such as plants for crushing and grinding rock, power generation, water production, et cetera)
- No resupply from Earth required
- Sintering temperature is below the melting temperature of the material, making the process less energy intensive than many other processes such as casting.

The primary difference between different sintering techniques is the method in which the thermal energy or heat is applied to the particles of regolith to sinter them together. This chapter will cover various heat application techniques including concentrated solar heating, microwave heating, and laser-based sintering; an emphasis will be placed on the research being conducted to apply these methods to extraterrestrial infrastructure.

12.2.1 *Physics of Sintering*

The sintering process bonds particles together by application of thermal energy. This bonding often occurs at the atomic scale through simultaneous changes to the pore and grain structures (German 2014). There are two major categories of sintering processes: solid-state sintering (SSS) and liquid-phase sintering (LPS). In SSS, the entirety of the material remains in the solid state throughout the entire process—no melting occurs at the sintering temperature. Figure 12.1 shows a graphical representation of the stages of the SSS process. Initially, the material is made up of loose powder. The spaces between particles are referred to as pores, and pore size depends on the size and shape of the particles and whether the powder was compacted prior to sintering. As the material is heated, the areas of contact between the particles, called the grain boundaries, grow to form ‘necks’ of material. It should be noted that atomic motion can occur by different diffusion mechanisms, which is partially dependent on the heating mechanism employed.

As the necks grow, they overlap with each other, which rounds and shrinks the pores, densifying the material. The extent of densification is carefully balanced with grain growth, which progressively decreases the strength of the final product. Shrinkage of the overall part also typically occurs as pore size decreases. Scanning electron microscope (SEM) images of the sintering process (densification phase) are presented in Fig. 12.2.

LPS refers to the scenario in which a liquid phase is created within the powder compact during sintering; this melted phase wets the remaining solid particles. The difference between liquid-phase and solid-state sintering mechanics is illustrated in Fig. 12.3. The shaded cell and area represent the component or portion of the

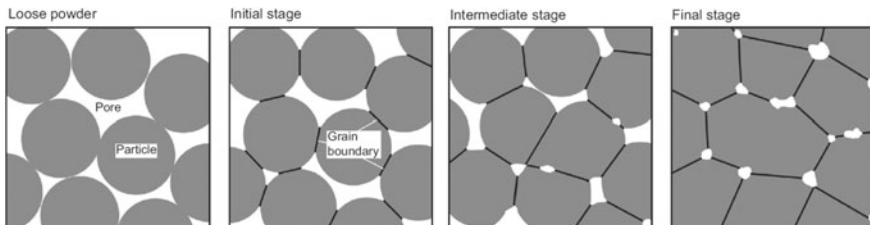


Fig. 12.1 Graphical representation of sintering stages (German 2014)

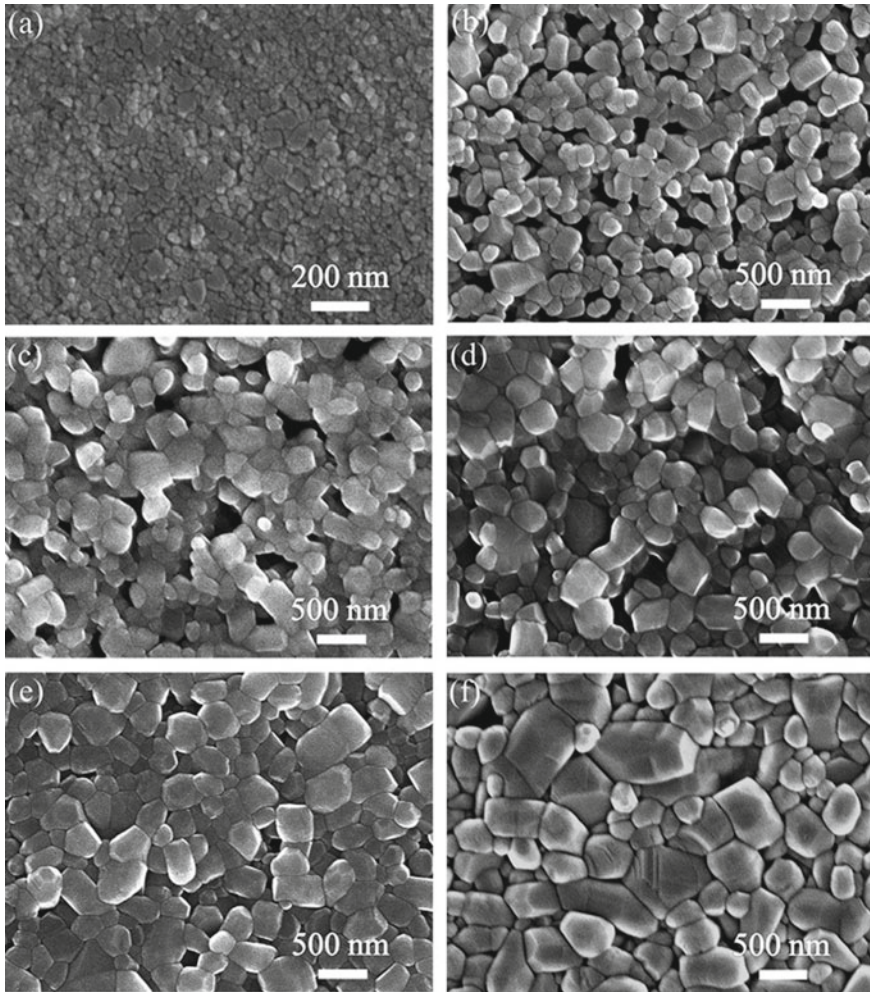


Fig. 12.2 SEM images of ZnO nanoparticles through sintering stages (Chen et al. 2013)

starting material that melts at the sintering temperature and wets the solid-state (unshaded/white) particles. The occurrence of SSS or LPS can be predicted based on the material composition and the chosen sintering temperature—the phase diagram shown in Fig. 12.4 illustrates the transition from SSS to LPS as sintering temperature increases for a given material composition.

Key parameters of the starting material that will affect the final properties and microstructure of the sintered material include:

- *Particle size and shape.* Particle shape impacts the packing of the powder prior to sintering. Particle size affects the size and number of contact points between particles, which impacts neck formation. Smaller particles are generally preferred

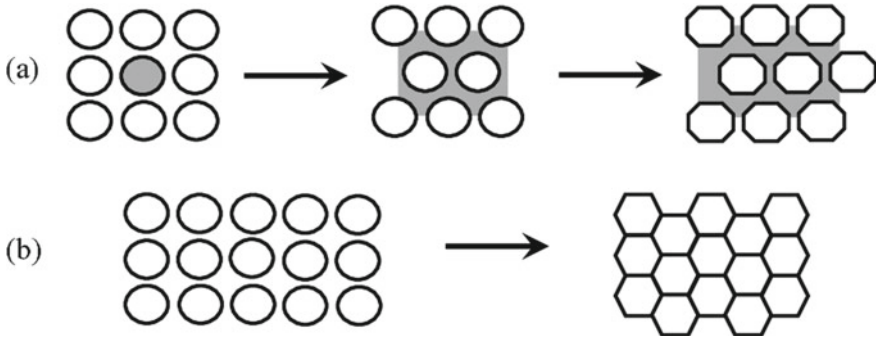


Fig. 12.3 a Liquid-phase sintering. b Solid-state sintering (Halim 2018)

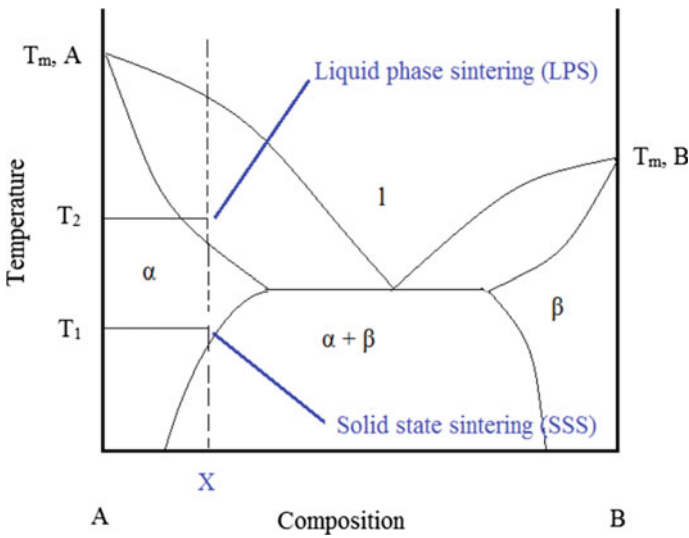


Fig. 12.4 Phase diagram depicting solid-state sintering and liquid-phase sintering temperatures for a given composition (Sipola 2015)

to improve the sintering response. Particle size distribution will also affect the final product. A range of particle sizes can improve the initial powder density—small particles fill the pores between large particles, which can also reduce shrinkage. Large differences between the largest and smallest particle sizes can increase the risk of cracking (German 2014).

- *Composition of the particles.* The composition of the particles is a key factor in determining the strength of the final sintered material. The chemical and mineralogical composition of the particles, the presence of contaminants, volatiles in the mixture, and the homogeneity of the powder will all influence how the material sinters.

- *Density prior to sintering.* Increasing the density of the powder prior to sintering decreases the porosity and increases the number of contact points between particles. This results in a denser, stronger sintered material, and can reduce sintering times.

Key parameters of the sintering process that will also affect the final properties and microstructure of the sintered material include peak temperature, heating and cooling rates, pressure, and composition of the atmosphere.

Industrial sintering is performed in controlled conditions. The parameters listed above are carefully balanced to achieve a product with desirable properties. Powders are of a deliberate size distribution. Particle shapes and composition are typically well known or characterized. A standardized compaction protocol may be used, and sintering occurs under highly controlled temperature and pressure conditions.

A typical goal of sintering is to achieve material properties as close as possible to those of the theoretical, fully dense solid of the same composition; sintering can also be used to purposefully adjust properties (enhance some and diminish others) to better meet the application's specific requirements. For example, by not sintering to the extent of reaching full consolidation, the material maintains some porosity, leading to a more lightweight structure with lowered strength. Besides manipulation of microstructure and material properties, other goals of recent sintering research include reducing sintering time and reducing energy consumption.

Natural regolith is an incredibly complex material that is very different from the homogeneous, controlled feedstocks used in terrestrial sintering. The heterogeneity of regolith can present numerous challenges for sintering. A wide distribution of particle sizes within a given volume of regolith can improve the packing density of a green body, as smaller particles fill the voids between larger particles; however, mechanical interlocking of particles, especially rough, irregular lunar regolith particles, can inhibit packing density. Variable mineralogical composition can also present a challenge, since different minerals melt at different temperatures. This variability will also impact the properties of the sintered material and can create uneven mechanical properties throughout the material. Quantifying these effects on natural regolith will be nearly impossible without significant testing to determine the regolith's composition prior to sintering.

12.2.2 Terrestrial Sintering Techniques

In terrestrial applications, many steps must occur before the sintering process begins. The raw, particulate material may be shaped by dry pressing or other means. The raw material may be turned into a clay by mixing with solvents, lubricants, or other additives. The clay may then be shaped by slip casting, pressure casting, injection molding, extrusion, or 3D printing to achieve the desired geometry. After shaping occurs, the part is termed a green body or green part.



Fig 12.5 (left) A typical furnace used for sintering ceramics (Source David Harvey, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons); (right) A conveyor furnace for continuous sintering of many parts

The green body can then be heated to decompose or evaporate any additives and sinter particles together using one of the following heating methods (list is not exhaustive).

12.2.2.1 Conventional Sintering (Pressureless)

A kiln or furnace is used to apply heat for conventional sintering of parts. Radiative and convective heat transfer occur at the outer surface of the part, and heat is conducted through to the core of the part. Figure 12.5 shows typical equipment used in conventional sintering.

12.2.2.2 Microwave Sintering

Utilization of microwaves for high-temperature processing of ceramics is a relatively recent innovation. Heating occurs by a coupling or absorption of the microwave field by the material as a whole; volumetric heating occurs within the material, rather than heating the exterior surface alone as is the case with conventional sintering (Agrawal 2006).

Relative to conventional sintering, microwave sintering allows for reduced processing time and energy consumption and increased heating rates. Microwave heating also differs from conventional heating in the diffusion mechanisms that occur at the atomic level—this difference, while not discussed here, allows for finer microstructures with enhanced mechanical properties (Oghbaei and Mirzaee 2010). Figure 12.6 presents key differences between conventional and microwave sintering.

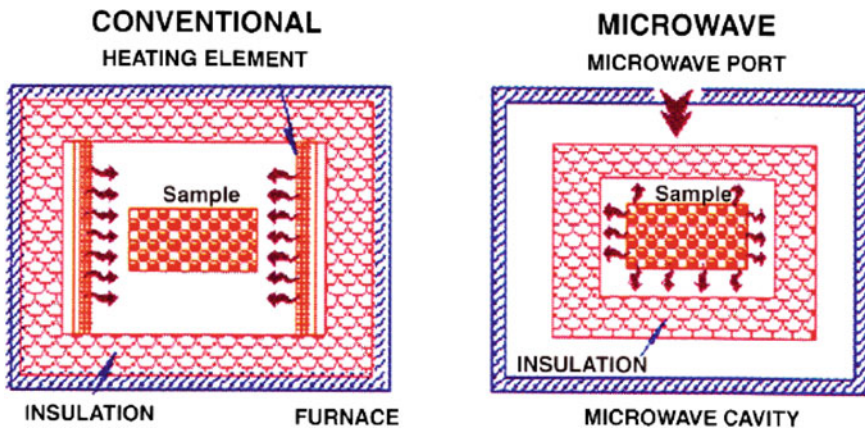


Fig. 12.6 Comparison of conventional and microwave heating set-ups (Matli et al. 2016)

The particulate to be sintered is typically surrounded by a material that is particularly good at absorbing and transferring the energy of microwave radiation, called a microwave susceptor. A resonance cavity may also be used to concentrate microwaves as standing waves.

12.2.2.3 Laser Sintering

One of the more common applications of laser sintering is in direct metal laser sintering (DMLS) 3D printers. This device incrementally adds thin layers of metal particulate to a powder bed. After the deposition of each layer, a high-wattage laser beam scans over the surface, precisely sintering a cross section of the desired shape. After the completion of all layers, the consolidated part is then removed from the loose powder bed and postprocessed. This system is shown in Fig. 12.7.

This concept is being adapted for ceramic materials that are similar in composition to lunar and Martian regolith. Ceramics typically have a higher melting point and require more powerful lasers which can lead to overheating and decomposition of the ceramics. Ceramics also have lower thermal conductivity, causing an increased melt pool area around the laser focus point. More sophisticated laser scanning strategies must be employed to scan the desired cross section and avoid thermal stress and dimensional distortions (Qian and Shen 2013). The resultant solid material can contain large glass phases. A laser-sintered glass specimen is pictured in Fig. 12.8.

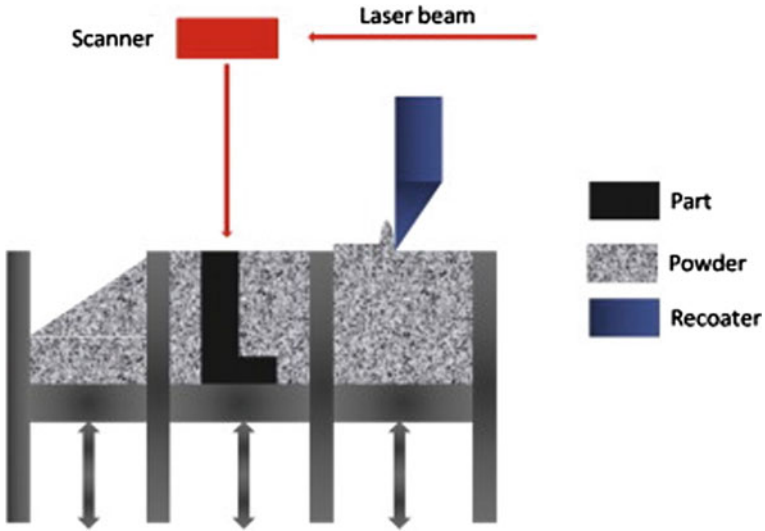
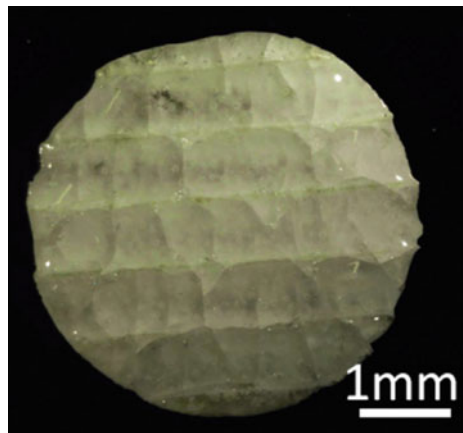


Fig. 12.7 Basic concept for a selective laser sintering system (Qian and Shen 2013)

Fig. 12.8 Laser-sintered glass sample made from crystalline oxides (Qian and Shen 2013)



12.3 Application of Sintering to Off-World Materials

Researchers have already begun studying the use of sintering to build infrastructure for future space missions. The primary focus of this research has been for missions to Earth’s Moon, but some work has also been done in applying sintering on Mars and asteroids. In this section, we will establish key differences between the well characterized and refined terrestrial materials and the extraterrestrial materials available for sintering.

12.3.1 Regolith

The surfaces of Earth's Moon and Mars are covered in regolith, an unconsolidated, heterogeneous rocky material covering bedrock. It is composed of rock, mineral, and glass fragments and ranges in diameter from nanometers through pebbles and rocks. The crusts of the terrestrial planets are composed primarily of silicate minerals such as olivine, pyroxene, and plagioclase feldspars. Particle shape varies greatly as well, from weathered grains on Mars to rough agglutinates and glass beads on the Moon.

12.3.1.1 Lunar Regolith

Lunar regolith, especially at the surface, is very loose, fine, and powdery. The median particle size is 40–130 μm (Heiken et al. 1991). This is good for sintering because smaller powders can sinter to a higher density than large particles. Fine particles are already present in the lunar regolith and would not need to be produced by grinding or crushing. Furthermore, the lunar regolith is poorly sorted; larger particles can be found below the first layers of the lunar surface. Larger particles could be added to the regolith fines to enhance packing and decrease shrinkage. Lunar regolith grain size distribution is shown in Fig. 12.9.

The particle shape of lunar regolith makes sintering more challenging. Lunar regolith particles are highly irregular due to the lack of weathering effects from water or wind. Intragranular porosity is high, which reduces the true packing density and may contribute to shrinkage of sintered parts.

The mineralogical composition of lunar regolith may be beneficial for sintering. Lunar regolith is mostly composed of aluminosilicates and other oxides (Meurisse

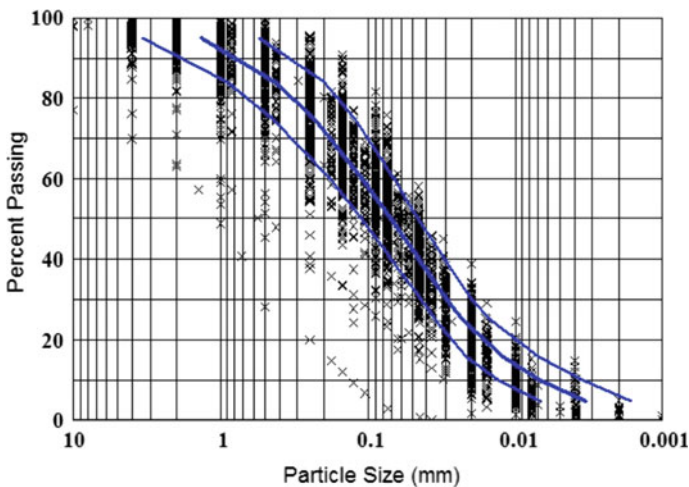


Fig. 12.9 Grain size distribution of lunar regolith from multiple Apollo samples (Carrier 2005)

Table 12.1 Chemical composition of lunar soils obtained by XRF (Sarantos et al. 2012)

Element	Low-Ti Mare Soils	High-Ti Mare Soils	Highland Soils
O	60.26	60.30	60.82
Si	17.30	15.86	16.31
Al	5.56	5.70	10.66
Mg	5.53	5.70	3.84
Ca	4.44	4.60	5.92
Fe	5.85	5.29	1.9
Ti	0.66	2.01	0.17
Na	0.26	0.31	0.29
K	0.06	0.05	0.05
Mn	0.08	0.07	0.03

et al. 2017). Chemical composition is provided in Table 12.1. The rock of the lunar highlands is primarily anorthositic, while the rock of the lunar maria is primarily basaltic. Terrestrial basalts that are believed to be of similar composition have shown promise as feedstock material for sintering (Edison et al. 2019). Sintering temperatures are dependent on the exact material composition and sintering technique employed but are similar to those used terrestrially for ceramic particulates.

Another similarity between lunar regolith and terrestrial ceramic particulates is their thermally insulative nature. Based on data from the Lunar Reconnaissance Orbiter, the thermal conductivity of lunar regolith is between 0.0001 and 0.03 W/mK (Yu and Fa 2016).

12.3.1.2 Martian Regolith

Less is known about the regolith that covers the Martian surface; to date, no Martian regolith samples have been returned to Earth. Unlike the Moon, the Martian surface is subjected to weathering and sedimentation processes, and there exists a greater amount of water, hydrated minerals, and other organic volatiles within the regolith. Particle size distribution is expected to vary more based on location, but Martian regolith appears to be sandy as compared to the dusty lunar surface, with average particles sizes of 0.17 mm (Carrier 2005). The chemical composition of Martian regolith is shown in Table 12.2; it is basaltic

12.3.1.3 Regolith Simulants

Very little lunar regolith was returned from the Apollo missions, and Martian regolith has yet to be returned to Earth, so sintering research must employ the use of regolith

Table 12.2 Martian soil composition (Litvak et al. 2016)

Element	Average abundance in Martial soil, %
O	43.886
Si	20.274
Al	4.968
Mg	5.112
Ca	4.306
Fe	12.748
Ti	0.578
Na	1.912
K	0.384
Mn	0.231
P	0.537
S	3.777
Cl	1.000
Cr	0.200
Zn	0.027
Br	0.017

simulants made from terrestrial materials. Simulants aim to replicate the chemistry, mineralogy, petrology, and physical characteristics of regolith, such as particle size distribution, shape, composition, or electromagnetic properties, but no single simulant is a perfect surrogate. It is therefore important to choose a simulant that accurately replicates properties that are relevant to the material process under investigation, including particle size distribution, shape, and elemental/mineral composition. A detailed discussion of simulants can be found in Chap. 3.

12.3.2 Extraterrestrial Application of Sintering Techniques

This section will summarize work done prior to 2021 in applying conventional, microwave, and laser sintering techniques to regolith simulants. Concentrated solar sintering, a novel technique not previously developed for terrestrial use, will also be presented. Non-sintering methods of utilizing in-situ materials for construction, such as casting and fiber production, are not considered here, but can be found in work done by Farries et al. (2021). This section will focus on laboratory-scale demonstrations that establish proof of concept and begin to characterize the process and final products. Integrations of sintering devices into full-scale robotic systems can be found in the following section.

Early sintering systems for extraterrestrial infrastructure will most likely need to be able to operate using unprocessed or minimally processed regolith. On Earth, the

machines and processes used to produce high-fidelity particulate feedstock are heavy and energy intensive. This constraint is reflected in the following sections.

12.3.2.1 Conventional Sintering

Sintering by heating in an oven is less technically challenging and less complex than other methods. It is a good way to establish the sinterability of regolith and simulants, although launch and operation of heavy and energy-intensive ovens is not ideal (substantial modification would be needed).

In one of the first investigations of this traditional sintering method, regolith simulant samples were successfully sintered in a Sircar furnace at a constant temperature of 1200 °C for 20 min (Gualtieri and Bandyopadhyay 2015). A mixture of JSC-1A, JSC-1AF, and JSC-1AC were sieved in order to make samples of two levels of porosity. A high-porosity sample was made by using particles >212 μm. A low-porosity sample was made by using particles between 25 and 212 μm—inclusion of fine particles can increase packing efficiency leading to a low-porosity product, as described previously. Based on the regolith collected from multiple sites during the Apollo 12 and 14 missions, these size distributions should be achievable by sieving alone. Both samples were compressed up to 145 MPa, increasing the green density and allowing the sample to be removed from the die to a furnace.

Sintering was performed in an air environment (not vacuum). The resultant densities for the high- and low-porosity samples were 92 and ~99%, respectively. X-ray diffraction (XRD) and scanning electron microscopy (SEM) revealed changes to the phases present and revealed that the powder underwent liquid-phase sintering. Sintering in air can produce lower density materials than vacuum sintering due to gases retained in the pores between grains (German 2014); as such, sintering in the near vacuum of the lunar environment could result in higher densities than those produced in this investigation.

Compression and hardness testing data are presented in Table 12.3. Hardness of both samples were both high, comparable to zirconia. As expected, a lower porosity is associated with increased compressive failure stress and modulus of elasticity (E). The demonstrated failure stress and modulus indicate that sintered regolith has potential to be used for structural components. For reference, Portland cement concrete has compressive strength of 20–40 MPa, and a modulus of elasticity of 14–41 GPa, depending on mixture.

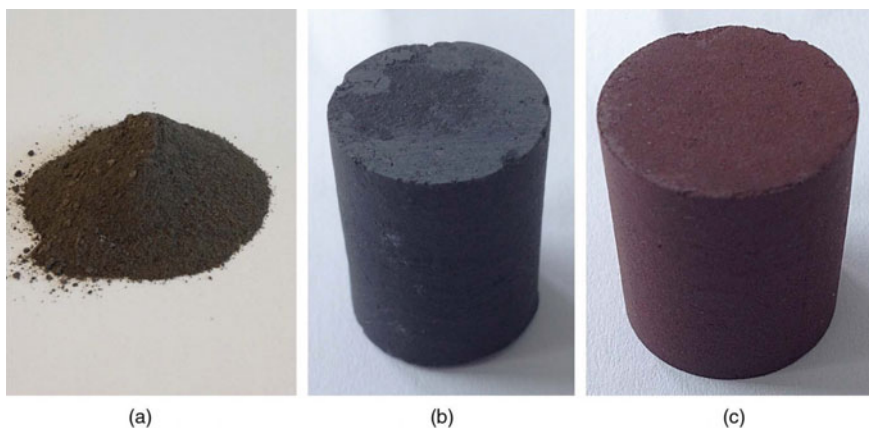
Table 12.3 Hardness and compression test results for conventionally sintered JSC-1A at 1200 °C for 20 min (Gualtieri and Bandyopadhyay 2015)

	Apparent porosity (%)	Failure stress (MPa)	E (GPa)	Hardness (HV _{0.1})
99.0% Dense (± 0.5%) Avg.	0.41	232	10.9	1027
Standard deviation	0.24	43.7	1.89	98.41
92% Dense (± 2%) Avg.	8.44	103.2	5.98	"
Standard deviation	1.44	26.7	0.71	"

A similar study was conducted with higher compaction pressures (255 MPa), a sintering temperature ramp/cooling rate (400 °C/h), and lower sintering temperatures (1,070–1,125 °C—minimal liquid phase sintering was observed) (Meurisse et al. 2017). This study aimed to elucidate the effects of sintering in vacuum versus air and the effects of mineral composition on sintering.

The effects of sintering in air are made immediately obvious in Fig. 12.10. The red color of the sample sintered in air is a result of oxidation. SEM images also reveal the formation of oxidized species hematite and magnesia. Both compressive stress and Young's modulus were improved by sintering in vacuum. This is good news for sintering in the lunar environment.

These studies of the application of conventional sintering to lunar regolith simulant establish the feasibility of sintering regolith grains into usable, consolidated parts that could be used in lunar construction. There will be less emphasis on sintering temperatures, pressures, and particle size distributions hereafter.

**Fig. 12.10** a JSC-1A powder b sintered in vacuum c sintered in air (Meurisse et al. 2017)

12.3.2.2 Microwave Sintering

Another terrestrial sintering technique that has been investigated for lunar applications is microwave sintering. Microwave energy is applied to the regolith to sinter the regolith, either in situ on the lunar surface or inside a resonance chamber. The microwave energy can, potentially, couple with the lunar regolith, which could lead to energy-efficient heating and sintering (Taylor and Meek 2005). The extent of coupling is dependent on the material's dielectric properties. Most testing prior to 2021 was done using microwaves with a frequency around 2.45 GHz and power of 700–1000 W, which are the typical specifications of a standard kitchen microwave. In space applications, 700–1000 W is a high-power draw application; however, the coupling effect between the microwaves and the lunar regolith could produce much more energy-efficient heating than other forms of thermal heating (apart from solar heating). The most significant benefit of microwave sintering is that it does not rely on conductive transfer of thermal energy through regolith; as such, it is theoretically possible to sinter in-situ lunar regolith down to several tens of centimeters in a single pass. Microwave sintering has also been demonstrated on actual lunar regolith (Taylor and Meek 2005).

While microwave heating could be promising for use in constructing both 2D (in-situ) and 3D (molten material extrusion from a resonance chamber) structures, there are several drawbacks. The first is that it is not precisely known why microwave energy couples with lunar regolith to produce the observed thermal effects. It was originally thought that nanophase iron rims, deposited onto individual grains via micrometeorite impacts, coupled with microwave energy and melted to sinter the grains of regolith together (Taylor and Meek 2005). More recent assessments, however, point to grain morphology as the reason for microwave coupling (Barmatz et al. 2013). Under this theory, it is the sharp, pointed grain structure of glasses within the regolith that respond to the microwave energy; the sharp points vibrate and melt, sintering the material.

The uncertainty about why microwave energy couples with lunar regolith causes several issues for testing microwave sintering in terrestrial labs. If it is unknown why microwave energy causes sintering in lunar regolith, it is difficult to replicate in regolith simulants and, indeed, there has been little success in microwave sintering of standard lunar regolith simulants. Susceptors, such as silicon carbide, can be added to the regolith simulant to improve sintering (Allan et al. 2013), but adding susceptors reduces the simulant's fidelity to actual lunar regolith, and thus reduces the usefulness of testing results. Figure 12.11 shows comparative results between regolith simulant sintered without (left) and with (right) silicon carbide added. The simulant sintered with added susceptors shows a much more homogeneous internal structure than the regolith simulant sintered without susceptors.

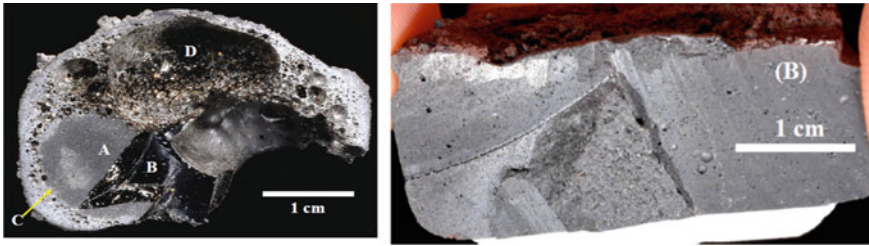


Fig. 12.11 Microwave sintered regolith simulant without (left) and with (right) added susceptors (Allan et al. 2013)

This uncertainty also creates challenges in optimizing microwave frequencies. 2.45 GHz is a common frequency to test at simply because that is the frequency that commercial microwave ovens operate at. This frequency was chosen because it couples well with water and is therefore good at heating food; this does not mean it is especially effective at heating lunar regolith. Without a good regolith simulant to perform tests on, it is difficult to determine if other frequencies might produce better results.

Work has been done on building microwave heating systems specifically designed for sintering lunar regolith, as opposed to heating water. Such a system would be capable of applying multiple frequencies, concentrating microwave energy to a single hotspot, and conducting experiments in vacuum. The system shown in Fig. 12.12 is in development to accomplish these goals (Lim et al. 2019).

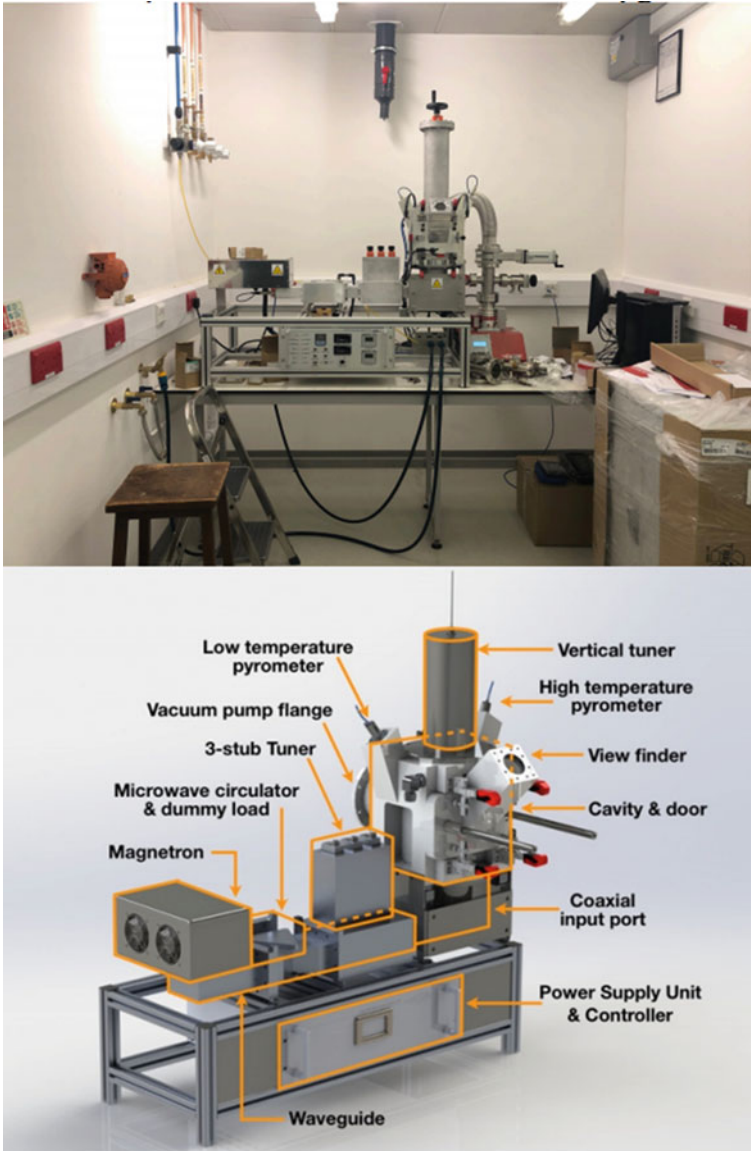


Fig. 12.12 Microwave heating apparatus for lunar simulant heating experiments (Lim et al. 2019)

12.3.2.3 Laser Sintering

Laser sintering requires that the wavelength of the laser be tuned to the absorption of the regolith to maximize the energy absorption rate. So far, this method has been primarily investigated for use in constructing small parts, although it could potentially

scale up for use in consolidating regolith either in situ (direct heating of the lunar surface) or inside a chamber. This method relies on conduction of energy through the regolith. As previously mentioned, since the thermal energy of the laser is focused in a small area, transient and residual stresses can arise within the regolith (Balla et al. 2012). To mitigate these stresses, a layered approach has been employed to sinter JSC-1A (Fateri and Gebhardt 2015). The goal is to carefully control the area of fully melted simulant interspersed with areas of sintered regolith. Figure 12.13 shows a small cube manufactured using laser sintering; the darker lines represent fully melted regolith simulant, while the light areas are sintered. The third panel (c) depicts the layering pattern employed to achieve a lower-stress profile.

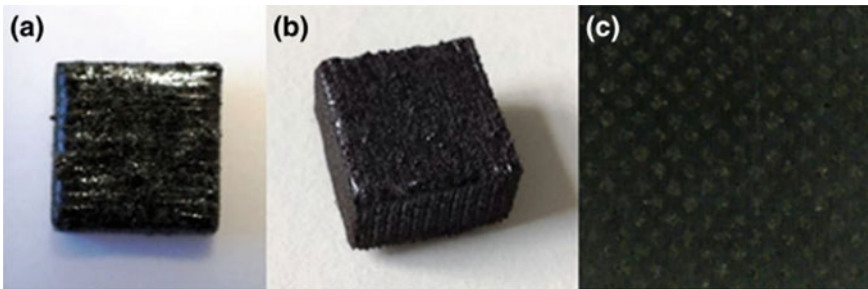


Fig. 12.13 JSC-1A consolidated cube manufactured with laser sintering (Fateri and Gebhardt 2015)

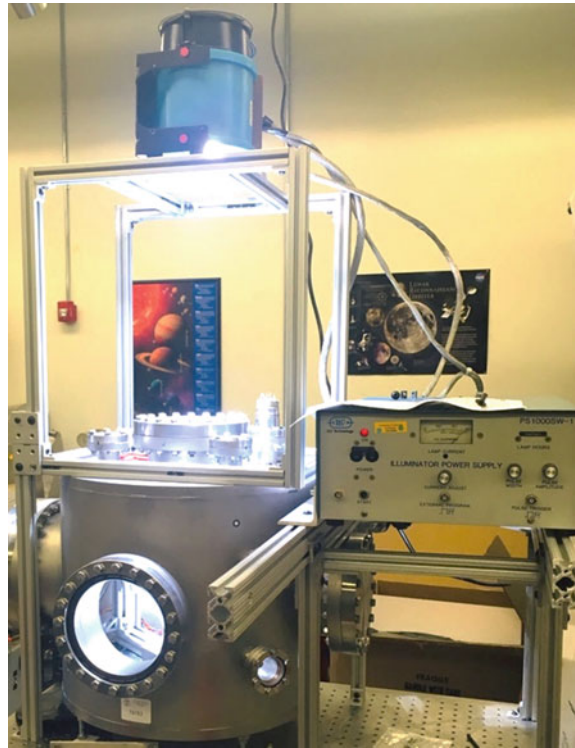
This cube was manufactured using a selective laser melting (SLM) printer equipped with a 100 W Yb:YAG fiber laser. The cube was sintered at a setting of 50 W. Layers of 100–300 μm thickness were incrementally sintered to construct the 10 mm \times 10 mm \times 3 mm cube.

12.3.2.4 Concentrated Solar Sintering

Sintering via concentrated solar energy is an attractive option for off-Earth construction because the power required to heat the regolith comes directly from the Sun and requires no conversion to another form of energy. Sunlight enters an aperture, reflects off a concentrator, and then is focused using a lens. This focused beam of energy is then directed at the lunar regolith, either in situ or inside a chamber, to heat it up to the appropriate temperature. As with laser sintering, concentrated solar sintering relies on conduction of energy through the regolith. If the material is heated in situ, it will radiatively cool into a solid material once the beam of concentrated sunlight moves away from it. Regolith can also be heated in a chamber that can be used to thermally insulate the material during heating, allowing for melting. Once melted, the molten regolith can be extruded through a nozzle or cast in a mold to make bricks. Figure 12.14 shows an experimental set-up for concentrated solar sintering, using a

xenon arc lamp to simulate solar illumination, of lunar simulant in a vacuum chamber (Williams and Butler-Jones 2019).

Fig. 12.14 Solar sintering test set-up using a xenon arc lamp (Williams and Butler-Jones 2019)



Although the low- or zero-power requirements for solar sintering are desirable from a system design perspective, there are challenges with this approach. The first major challenge with solar sintering (and laser sintering) is the incredibly poor thermal conductivity of lunar regolith under vacuum. As a result, concentrated solar energy applied to in-situ regolith would sinter (or melt, depending on the dwell time of the solar energy) the very top surface of the regolith without affecting the regolith below the surface. This would result in a thin, cracker-like crust of sintered regolith, which would be unsuitable for load-bearing applications such as landing pads or roads.

Some work has been done to demonstrate concentrated solar heating of regolith in layers, similar to the method employed for laser sintering (Urbina et al. 2017). In this method, solar energy is applied to sinter a thin layer of regolith. Then a layer of fresh regolith (approximately 0.1 mm) is placed over the sintered layer and heat is applied to sinter the fresh material and bond it to the layer below. A system for feeding and spreading regolith in this manner is shown in Fig. 12.15. A sample of sintered JSC-2A was produced with this distribution system using a Fresnel lens

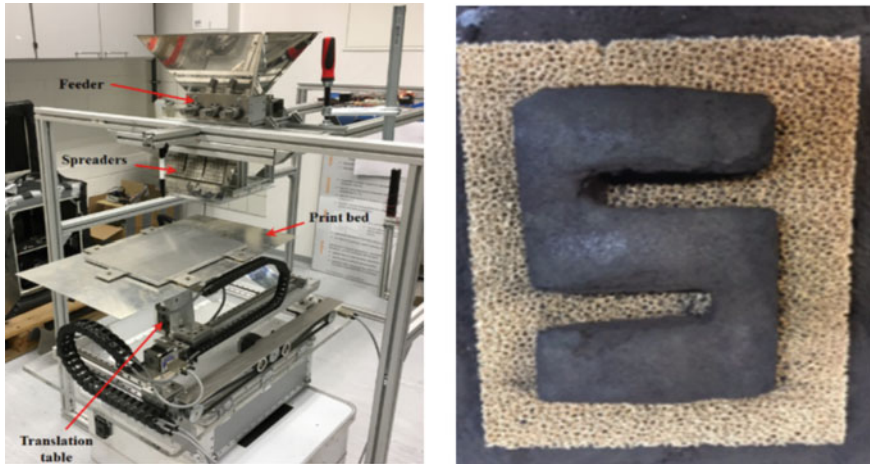


Fig. 12.15 (left) Simulant distribution system and (right) a solar sintered sample (Urbina et al. 2017)

to concentrate solar light. The results of these initial studies are promising for the creation of small components like bricks but could be very challenging to apply in situ due to the requirement to apply a very thin, even layer of regolith over a large area (see discussion in Sect. 12.4.2).

Beyond simulant distribution systems, entire solar oven test beds have been created to sinter regolith with actual sunlight, depicted in Fig. 12.16 (Meurisse et al. 2018). This system uses a mobile 52 m² mirror called a heliostat to track the Sun and reflect light into a stationary concentrator. The concentrator reflects a focused beam of light into a laboratory and onto a testbed with a distribution system to create even simulant layers. On the surface of Earth, this set-up is hindered by the flux density variations caused by fluctuations in the atmosphere, but this is a non-issue on the lunar surface.

Parts were produced using actual concentrated sunlight and simulated sunlight created by a xenon lamp. While the solar 3D printing concept has been demonstrated, the parts produced by both sunlight and simulated sunlight possessed poor mechanical properties; further refinement of this technique is ongoing.

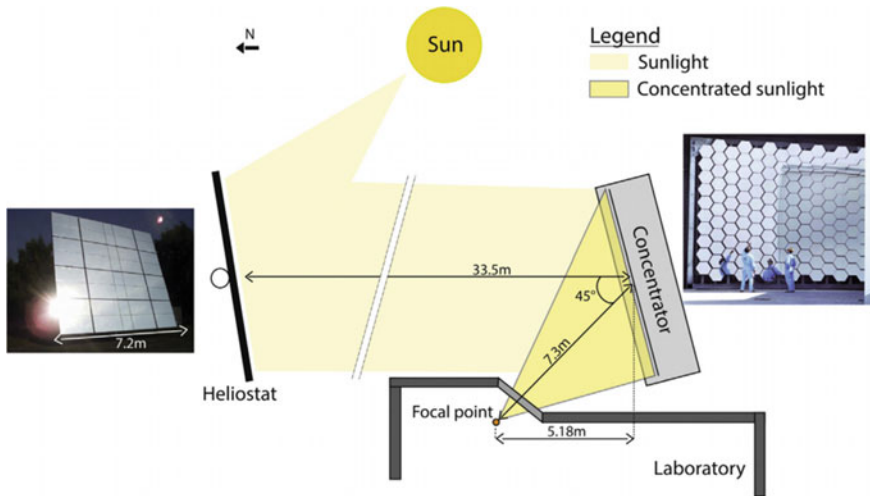


Fig. 12.16 Solar oven layout (Meurisse et al. 2018)

12.4 Applications of Sintering to Off-World Construction

Different types of structures will be required to protect landing vehicles, emplaced assets, robotic functionality, and human health. These structures can be broken into two broad categories: two-dimensional structures and three-dimensional structures.

Two-Dimensional Structures

Two-dimensional (2D) structures include landing pads, roads, walkways, and dust mitigation areas. Low 3D structures, such as berms, can also be included in the 2D category because their construction may be more like 2D structures than taller 3D structures. 2D structures are critical for both human and robotic operations for two reasons: they mitigate dust and improve trafficability. These structures will likely be the first to be constructed at an off-Earth human or robotic base of operations.

Within the next decade, commercial, civil space, and international activities on the lunar surface will increase. One significant difference between this era of exploration and previous missions is that the landing areas will be relatively closely spaced. Some areas are of particular scientific or commercial interest, such as the lunar South Pole around ‘peaks of eternal light’ or the permanently shadowed regions of craters. Multiple civil, commercial, and international entities may land missions relatively near one another to independently explore or extract resources from these regions. Some commercial enterprises, such as a water-propellant extraction and processing system, may require multiple missions to land near one another to create an efficient base of operations. Any human outposts, including research stations, tourist destinations, or manufacturing facilities, will also require landing Earth-launched

hardware in close proximity. The problem with increased landing activities, especially very close together, is that landing spacecraft kick up a lot of regolith dust. This ejecta consists of very high-velocity, fine, abrasive particles that can cause significant damage to nearby hardware and infrastructure. The blasted regolith can disturb resources, historic sites such as the Apollo landing sites, or scientific sites, even kilometers away from the landing spacecraft. There is also evidence that the regolith can even be ejected into lunar orbit, especially by large spacecraft such as Apollo or the proposed Artemis lander (Metzger 2020). This ejected regolith could cause damage to other incoming landers or even orbiting spacecraft such as the Lunar Gateway, increasing risk for all future lunar activities.

While dust can be a nuisance anywhere, it is a particularly profound problem on the Moon. Lunar dust particles are frequently jagged since there are no erosion effects to smooth the grains. These jagged particles cause rapid degradation of mechanical joints, wear on wheels, and can abrade soft materials on astronauts' spacesuits. Lunar regolith also contains very fine particles, approximately half of which are finer than the human eye can see, that are electrostatically charged and thus adhere to anything that touches the lunar surface (Heiken et al. 1991). This is especially problematic for human explorers, as the Apollo astronauts discovered when their suits became coated up to the shoulder in dust. Lunar dust is primarily comprised of silicates which, when inhaled, can cause silicosis and can be fatal. The jagged nature of the particles can also damage the soft tissue of the respiratory tract. Since the dust is finer than the human eye can resolve, jagged, and electrostatically 'clingly,' cleaning spacesuits prior to entry into habitats is unlikely to be very effective; dust mitigation, in the form of walkways, roads, and dust-suppressed work areas, will be required for safe operations for humans and robots.

In addition to dust mitigation, 2D structures also improve trafficability. Trafficability is defined as "the capacity of a soil to support a vehicle and to provide sufficient traction for movement" (Heiken et al. 1991). Roads provide many benefits for wheeled rover mobility: they reduce energy needed to move the vehicle, allow for higher speeds, reduce wear on vehicle wheels, and provide a pre-planned route to traverse around hazards such as craters, areas of soft dirt, and boulders. Humans may also find it easier to traverse the lunar surface on walkways rather than on loose regolith.

The definition of "trafficability" could be expanded to include soil support for landing and launching spacecraft. Landing pads provide a safe, smooth, flat surface for incoming spacecraft, which reduces overall mission risk. Current landing craft, especially on Mars, are equipped with automated landing zone assessment sensors and software that must autonomously survey potential landing zones during descent and adjust to avoid hazards such as boulder fields or craters, with failure resulting in a loss of the mission. Landing pads would limit the need for such complex systems. Beacons or targets could be placed on landing pads to guide incoming vehicles to the pad. The pad itself, in addition to mitigating dust, serves as a safe location for spacecraft to set down, free of the risk of impacting boulders, partially sinking into unstable dirt, or landing at an angle due to uneven ground.

Three-Dimensional Structures

Three-dimensional (3D) structures include protective structures, such as habitats for humans or thermal shelters for robotic equipment, and structural elements, such as supports for scientific instruments or vehicles on a launch pad.

Protective structures will allow human and robotic explorers to survive harsh off-world environments. Habitats will provide radiation shielding, thermal protection, and atmospheric control to humans. There are many concepts for habitats, but the most likely first-use scenario is a regolith shell surrounding a pressurized bladder supplied from Earth. The regolith shell provides protection from radiation and micrometeorites, which are especially prevalent on the Moon due to the lack of atmosphere, while the bladder maintains a human-appropriate pressure and temperature environment. Figure 12.17 shows a concept called SinterHab. In this concept, inflatable, pressurized membrane structures are launched from Earth and a shell of sintered regolith is built around them to provide radiation protection and micrometeorite shielding (Rousek et al. 2012).

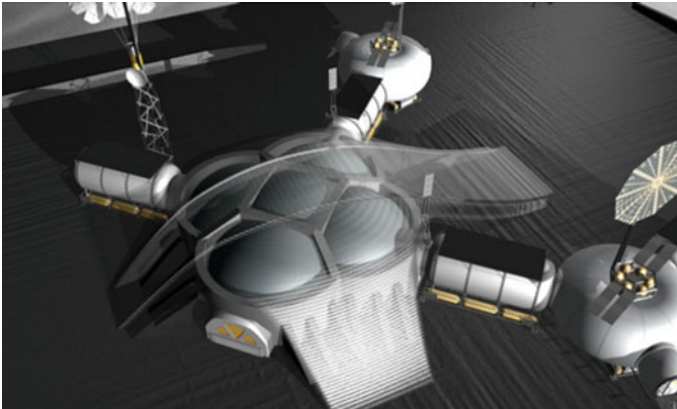


Fig. 12.17 SinterHab 3D habitat concept (Rousek et al. 2012)

Robotic equipment will also need protection from thermal changes and micrometeorites on the Moon and dust storms on Mars. The temperature swing from night to day on the lunar surface is about 280 K (Heiken et al. 1991) and it is very challenging to design rovers and equipment that can survive through a lunar night. A garage could be built to house the equipment and provide thermal protection through the night.

Structural elements will allow the construction of equipment that is too large to be launched from Earth. This could include supports for large deep-space antennas or supports and service towers for launch pads. The launch pads themselves may also require 3D elements, such as stands, to raise the launch vehicle above the pad to reduce thermal effects on the launch pad.

12.4.1 *Two-Dimensional Construction*

Sintering can, theoretically, be used to produce 2D structural material in situ. This is primarily applicable to structures such as roads, landing pads, walkways, and dust mitigation areas, as well as low 3D structures such as berms. To improve the quality of the sintered regolith, it may be necessary to prepare the ground by removing rocks or pebbles and smoothing or compacting the regolith. These minimal surface preparation steps will result in a more uniform sintered mass, better contact between grains to improve the quality of the sintering, and a smoother top surface. Once the regolith is prepared, the sintering energy source is passed over the regolith on the ground (in situ), which sinters it in place into a solid surface. For construction of berms and other low 3D structures, regolith will be piled in place and then sintered. This method could be performed in layers, where an initial substrate is created directly on the surface and subsequent layers are applied over the top and sintered into place. Layering may be necessary for sintering methods such as solar sintering, which relies on conductive transfer of heat through the regolith.

In-situ sintering is advantageous because it is, theoretically, very simple; an autonomous or semi-autonomous robotic constructor ambles along the surface, sintering the regolith into smooth, glassy surfaces that mitigate dust from incoming landers and robotic or human surface activities. Ground preparation, such as scooping out rocks and smoothing or compacting the soil, could be performed by the robotic constructor without the need for additional earthmoving equipment. There is no need for the robotic constructor to pick up or process the regolith, which would add complexity to the system. Several in-situ sintering approaches have been proposed.

Robotic design for an in-situ constructor is relatively simple. Figure 12.18 shows a concept for a simple Lunar Road-Paving Wagon that uses microwave sintering (Meek et al. 1986). The blade at the front of the rover smooths the surface for the magnetrons in the back to sinter. The parabolic metal reflector directs the microwaves down onto the surface and prevents electromagnetic interference with the rover's electronics. The sketch is overly simplified—it does not show onboard electronics, communications hardware, or a power supply—but it demonstrates the relative simplicity of a microwave sintering rover.

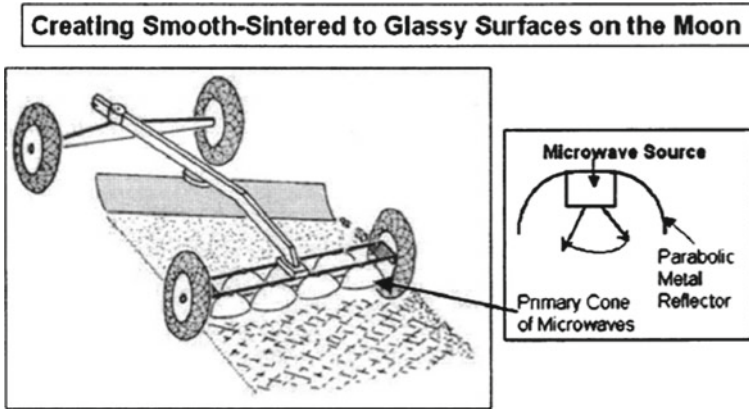


Fig. 12.18 Sintering lunar road-paving wagon (Meek et al. 1986)

Design of a solar sintering robotic constructor is simpler still. Since solar sintering does not require generation of electrical energy—sunlight is simply focused to produce thermal energy for sintering—the power requirements for the vehicle are limited only to what is needed to move and operate the vehicle. Figure 12.19 shows a mock-up of a solar sintering vehicle, which uses a Fresnel lens to concentrate sunlight onto the surface (Cardiff and Hall 2008). The mock-up vehicle does not include additional processing hardware such as a smoothing blade or compaction wheel, which would likely be necessary to improve the properties of the sintered surface.



Fig. 12.19 Solar sintering dust mitigation vehicle (Cardiff and Hall 2008)

In practice, however, in-situ sintering may be very challenging. Microwave sintering of in-situ regolith may be impossible, or the resulting sintered material may be unacceptable for load-bearing applications without the addition of susceptors. Mixing susceptors into the regolith could require large quantities of materials from Earth and would significantly increase the complexity and energy required to sinter. Sintering methods that rely on thermal conductivity, such as concentrated solar and laser sintering, will require tremendous amounts of energy to overcome regolith's natural insulative properties; this is especially true in the vacuum of the lunar surface. Given the variable mineralogical composition of the regolith, full melting of the regolith may produce better results than sintering.

Sintering can also be used in the creation of bricks, tiles, or pavers, which can be used to create either 2D or 3D structures, depending on how they are shaped and assembled. To create bricks, regolith is collected and placed into molds. Energy is applied to the material in the mold to sinter the regolith into bricks, which are then emplaced to form a 2D or 3D structure. These bricks can include interlocking features to hold them together, or they can be placed and then sintered together to form a complete structure. Figure 12.20 shows the PICSES Rover Helelani (with a robotic arm supplied by Honeybee Robotics) assembling interlocking pavers made with Hawaiian basalt for use in a landing pad (Romo et al. 2018). Bricks or pavers have one significant benefit over monolithic 2D structures, especially on the Moon. Temperature swings between day and night will cause thermal expansion and contraction, which could cause monolithic structures such as roads or landing pads to crack. Landing pads have additional thermal stressors due to heat from incoming or launching spacecraft. Cracking could cause large chunks of the structure to break off, especially during launch or landing of a spacecraft, which would be extremely dangerous. Structures built with bricks and pavers could be designed to accommodate thermal expansion and contraction.



Fig. 12.20 PISCES Rover Helelani assembling interlocking pavers made with sintered Hawaiian Basalt (Romo et al. 2018)

12.4.2 Three-Dimensional Construction

Sintering methods can also be used to construct 3D shapes. This method can be accomplished in three ways: by layering regolith and sintering it in place (similar to the in-situ method discussed earlier but with the goal of creating 3D structures instead of flat structures); by creating bricks and building 3D structures; or by heating regolith into a molten state and extruding it in layers.

The first method of 3D construction is by layering regolith and sintering it into place to form 3D structures. This technique is straightforward; the first layer of regolith is sintered in situ to anchor it to the surface and then subsequent layers are placed on top and sintered to form a 3D shape. This method can be achieved with either microwave or solar sintering. One robotic chassis that has been investigated for use in 3D construction is the All-Terrain Hex-Limbed Extra-Terrestrial Explorer (ATHLETE) robot, developed by the Jet Propulsion Laboratory (JPL) (Wilcox et al. 2007). ATHLETE is a six-limbed, high-mobility rover that has a wheel on each leg. Each leg can move independently with a high degree of precision. This allows the ATHLETE to roll swiftly over open terrain while switching to walking mode to navigate difficult terrain. With its high-precision movement, each leg can also be equipped with different tools and instruments. The chassis itself can carry many different payloads, including solar panels, batteries, regolith hoppers, and melting chambers. The long reach provided by the arms allows the rover to construct large 3D structures. Figure 12.21 shows a microwave sintering head integrated onto an ATHLETE rover (Howe et al. 2013). Figure 12.22 shows an ATHLETE rover with solar concentrators mounted to the center of the chassis, with incoming solar energy focused on a fiber optic cable that applies it to the regolith (Howe et al. 2013). Note

that both ATHLETE configurations can be used to construct 2D (in-situ) structures as well.

The second method of 3D construction is creating bricks and stacking them to form 3D structures. This method is similar to the construction of 2D structures, as discussed in the previous section.

The third method of 3D construction is heating the regolith to a molten state and extruding it in layers to form the structure, similar to material extrusion in 3D printing done on Earth. First, the regolith is collected and heated in a chamber. The thermal energy could come from sunlight or from microwave energy; this method works well with both types of thermal energy inputs. Solar energy works with this method because the chamber can be pressurized, increasing thermal conductivity through the regolith, or agitated to ensure even heating. Microwave energy input benefits

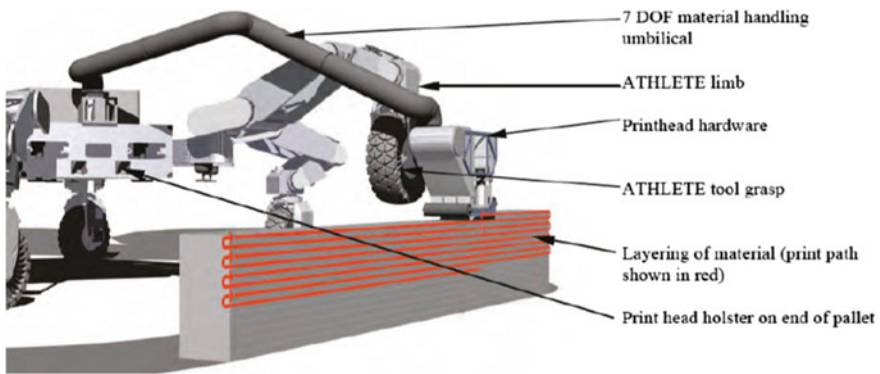


Fig. 12.21 Microwave sintering using the ATHLETE chassis (Howe et al. 2013)

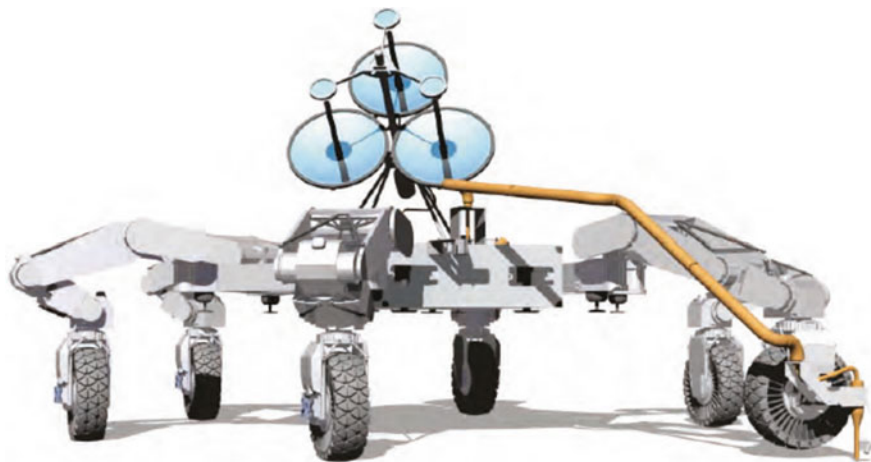


Fig. 12.22 Solar sintering using the ATHLETE chassis (Howe et al. 2013)

from this method because the chamber can be a resonance cavity, similar to what is created in a standard kitchen microwave oven; this increases the energy efficiency of the applied microwaves. Once the regolith is melted, it is transported to a print head, which applies the molten material in layers. These layers cool and harden into a solid material. While not technically a sintering method, since the regolith is heated to the melting point before extrusion, it is included here because the methods of applying thermal energy to the regolith are the same as for sintering. This method is suitable for building 3D structures when the extruding head is integrated into a construction robot or gantry system. The ATHLETE rover, as described earlier, is well suited to this method.

12.5 Summary

Humanity's ability to expand into the solar system will depend on the capability to build infrastructure to lower the risk to human and robotic explorers by providing protection and dust mitigation. Construction is an ideal application of ISRU concepts since the materials are expensive to launch from Earth and abundant on extraterrestrial surfaces in the form of regolith. Sintering, in which thermal energy is applied to particles to bond them into a solid material, is a promising technique for 2D (in-situ) and 3D off-Earth construction. Theoretically, sintering has several benefits over other extraterrestrial construction methods; it requires no resupply of materials from Earth, uses less energy than melting the material completely, and requires little to no handling or processing of regolith.

Researchers are investigating several methods for sintering regolith for construction, each with different benefits and drawbacks. The primary difference between these methods is how heat is applied to regolith particles. These methods include: conventional sintering, where heat is applied to the particles inside an oven; microwave sintering, where thermal energy is applied via microwaves; laser sintering, where thermal energy comes from a tuned laser; and concentrated solar, where sunlight is collected and focused onto regolith to sinter it. Investigations into each sintering method are ongoing using regolith simulants, which mimic certain properties of lunar and Martian regolith.

Once off-Earth sintering construction methods are refined in terrestrial labs, these techniques will be integrated onto robotic constructors and tested on the Moon and Mars. Robotic constructors, ranging from simple wagons to complex machines such as the ATHLETE rover, will be able to build 2D structures, such as roads and landing pads, and 3D structures, such as habitats and structural supports. Sintering could be used to build structures directly onto a planetary surface or to produce bricks or pavers for emplacement in 2D or 3D structures.

Sintering has been used by humans for thousands of years and has the potential to enable humanity's expansion to extraterrestrial bodies through the use of regolith, their most abundant natural resource.

Acknowledgements The authors gratefully acknowledge the support of the Center for Space Resources at the Colorado School of Mines and the guidance of Dr Christopher Dreyer. The authors also thank Kyla Edison of the Colorado School of Mines for her review and advice.

References

- Agrawal, D. 2006. Microwave sintering of ceramics, composites and metallic materials, and melting of glasses. *Transactions of the Indian Ceramic Society* 65 (3): 129–144. <https://doi.org/10.1080/0371750X.2006.11012292>.
- Allan, S.M., B.J. Merritt, B.F. Griffin, P.E. Hintze, and H.S. Shulman. 2013. High-temperature microwave dielectric properties and processing of JSC-1AC lunar simulant. *Journal of Aerospace Engineering* 26 (4): 874–881. [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0000179](https://doi.org/10.1061/(ASCE)AS.1943-5525.0000179).
- Balla, V.K., L.B. Roberson, G.W. O'Connor, S. Trigwell, S. Bose, and A. Bandyopadhyay. 2012. First demonstration on direct laser fabrication of lunar regolith parts. *Rapid Prototyping Journal* 18 (6): 451–457. <https://doi.org/10.1108/13552541211271992>.
- Barmatz, M. B., D. E. Steinfeld, D. Winterhalter, D. L. Rickman, and M. A. Weinstein. 2013. Microwave heating studies and instrumentation for processing lunar regolith and simulants. In *44th lunar and planetary science conference*, vol. 51, 1223.
- Cardiff, E. H., and B. C. Hall. 2008. A dust mitigation vehicle utilizing direct solar heating. In *Joint annual meeting of lunar exploration analysis group-international conf. on exploration and utilization of the moon-space resources roundtable*.
- Carrier, D. 2005. The four things you need to know about the geotechnical properties of lunar soil. Lunar Geotechnical Institute, September, 23. http://www.lpi.usra.edu/lunar/surface/carrier_lunar_soils.pdf?origin=publication_detail.
- Chen, B., Z. Xia, and K. Lu. 2013. Understanding sintering characteristics of ZnO nanoparticles by FIB-SEM three-dimensional analysis. *Journal of the European Ceramic Society* 33 (13–14): 2499–2507. <https://doi.org/10.1016/j.jeurceramsoc.2013.04.026>.
- Edison, K., C. Andersen, K. Harford, K. Higaki, and R. Romo. 2019. Hawaiian basalt characterization and the effects of chemical composition variances on the sintering process; Potential implications for Lunar/Mars ISRU applications. In *International astronomical conference*.
- Farries, K.W., P. Visintin, S.T. Smith, and P. van Eyk. 2021. Sintered or melted regolith for lunar construction: State-of-the-art review and future research directions. *Construction and Building Materials* 296: 123627. <https://doi.org/10.1016/j.conbuildmat.2021.123627>.
- Fateri, M., and A. Gebhardt. 2015. Process parameters development of selective laser melting of lunar regolith for on-site manufacturing applications. *International Journal of Applied Ceramic Technology* 12 (1): 46–52. <https://doi.org/10.1111/ijac.12326>.
- German, R. 2014. *Sintering: From empirical observations to scientific principles*. Elsevier Inc. <https://doi.org/10.1016/C2012-0-00717-X>
- Gualtieri, T., and A. Bandyopadhyay. 2015. Compressive deformation of porous lunar regolith. *Materials Letters* 143: 276–278. <https://doi.org/10.1016/j.matlet.2014.11.153>.
- Halim, J. 2018. Synthesis and transport properties of 2D transition metal carbides (Issue October). <https://doi.org/10.13140/RG.2.2.20355.55849>
- Heiken, G. H., D. T. Vaniman, and B. M. French. 1991. *Lunar sourcebook*, 778. Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.004>
- Howe, S. A., B. H. Wilcox, C. McQuin, J. Townsend, R. R. Rieber, M. Barmatz, and J. Leichty. 2013. Faxing structures to the moon: Freeform additive construction system (FACS). In *AIAA SPACE 2013 conference and exposition*, vol. 2, no. 2, 54–65. <https://doi.org/10.2514/6.2013-5437>

- Lim, S., Y. Jiang, A. D. Morse, M. Anand, J. Bowen, and A. Holland. 2019. Microwave heating experiment of lunar simulant (JSC-1A) using a bespoke industrial microwave apparatus. In *European lunar symposium (ELS2019)*, 1–3.
- Litvak, M.L., D.V. Golovin, I. Jun, A.S. Kozyrev, I.G. Mitrofanov, A.B. Sanin, V.N. Shvetsov, G.N. Timoshenko, and A. Zontikov. 2016. Implementation of gamma-ray instrumentation for solid solar system bodies using neutron activation method. *Nuclear Instruments and Methods in Physics Research, Section a: Accelerators, Spectrometers, Detectors and Associated Equipment* 822 (June): 112–124. <https://doi.org/10.1016/j.nima.2016.03.087>.
- Matli, P. R., R. A. Shakoor, A. M. A. Mohamed, and M. Gupta. 2016. Microwave rapid sintering of al-metal matrix composites: A review on the effect of reinforcements, microstructure and mechanical properties. *Metals* 6 (7). <https://doi.org/10.3390/met6070143>
- Meek, T. T., D. T. Vaniman, R. D. Blake, and F. H. Cocks. 1986. Electromagnetic energy applied to and gained from lunar materials. In *Symposium '86: The first lunar development symposium*.
- Metzger, P. 2020. Dust transport and its effects due to landing spacecraft. *Lunar Dust 2020*(2141).
- Meurisse, A., J.C. Beltzung, M. Kolbe, A. Cowley, and M. Sperl. 2017. Influence of mineral composition on sintering lunar regolith. *Journal of Aerospace Engineering* 30 (4): 04017014–04017023. [https://doi.org/10.1061/\(asce\)as.1943-5525.0000721](https://doi.org/10.1061/(asce)as.1943-5525.0000721).
- Meurisse, A., A. Makaya, C. Willsch, and M. Sperl. 2018. Solar 3D printing of lunar regolith. *Acta Astronautica* 152 (September): 800–810. <https://doi.org/10.1016/j.actaastro.2018.06.063>.
- Oghbaei, M., and O. Mirzaee. 2010. Microwave versus conventional sintering: A review of fundamentals, advantages and applications. *Journal of Alloys and Compounds* 494 (1–2): 175–189. <https://doi.org/10.1016/j.jallcom.2010.01.068>.
- Qian, B., and Z. Shen. 2013. Laser sintering of ceramics. *Journal of Asian Ceramic Societies* 1 (4): 315–321. <https://doi.org/10.1016/j.jascer.2013.08.004>.
- Romo, R., C. Andersen, K. Defore, K. Zacny, M. Thangavelu, and T. Lippitt. 2018. Planetary lego: Designing a construction block from a regolith derived feedstock for in situ robotic manufacturing. *Earth and Space* 2018: 289–296. <https://doi.org/10.1061/9780784481899.029>.
- Rousek, T., K. Eriksson, and O. Doule. 2012. SinterHab. *Acta Astronautica* 74: 98–111. <https://doi.org/10.1016/j.actaastro.2011.10.009>.
- Sacksteder, K. R., and G. B. Sanders. 2007. In-situ resource utilization for lunar and mars exploration. In *Collection of technical papers—45th AIAA aerospace sciences meeting, 6*(January 2007), 4232–4237. <https://doi.org/10.2514/6.2007-345>
- Sarantos, M., R. M. Killen, D. A. Glenar, M. Benna, and T. J. Stubbs. 2012. Metallic species, oxygen and silicon in the lunar exosphere: Upper limits and prospects for LADEE measurements. *Journal of Geophysical Research: Space Physics* 117(3). <https://doi.org/10.1029/2011JA017044>
- Sipola, J. 2015. Quality management during sintering of cemented carbides and cermets. Uppsala Universitet.
- Taylor, L.A., and T.T. Meek. 2005. Microwave sintering of lunar soil: Properties, theory, and practice. *Journal of Aerospace Engineering* 18 (3): 188–196. [https://doi.org/10.1061/\(ASCE\)0893-1321\(2005\)18:3\(188\)](https://doi.org/10.1061/(ASCE)0893-1321(2005)18:3(188)).
- Urbina, D.A., H.K. Madakashira, J. Salini, S. Govindaraj, R. Bjoerstad, J. Gancet, M. Sperl, A. Meurisse, M. Fateri, B. Imhof, W. Hoheneder, P. Weiss, M.M. Peer, and E. Prodeka. 2017. Robotic prototypes for the solar sintering of regolith on the lunar surface developed within the Regolith project. *Proceedings of the International Astronautical Congress, IAC* 4 (November): 2632–2641.
- Wilcox, B.H., T. Litwin, J. Biesiadecki, J. Matthews, M. Heverly, J. Morrison, J. Townsend, N. Ahmad, A. Sirota, and B. Cooper. 2007. Athlete: A cargo handling and manipulation robot for the moon. *Journal of Field Robotics* 24 (5): 421–434. <https://doi.org/10.1002/rob.20193>.

- Williams, H., and E. Butler-Jones. 2019. Additive manufacturing standards for space resource utilization. *Additive Manufacturing* 28 (June 2019): 676–681. <https://doi.org/10.1016/j.addma.2019.06.007>
- Yu, S., and W. Fa. 2016. Thermal conductivity of surficial lunar regolith estimated from Lunar Reconnaissance Orbiter Diviner Radiometer data. *Planetary and Space Science* 124: 48–61. <https://doi.org/10.1016/j.pss.2016.02.001>.

Chapter 13

The Effects of Mineral Variations on the Basalt Sintering Process and Implications for In-Situ Resource Utilization (ISRU)



Kyla P. Edison, G. Jeffery Taylor, Christian B. Andersen, and Rodrigo F. V. Romo

Abstract There is a significant interest in lunar and Mars regolith as a source for construction materials, sintering being one proposed method to convert it from loose unconsolidated rock material into durable solid forms. Hawaii's basalt has chemical properties like those of lunar and Martian regolith. The Pacific International Space Center for Exploration Systems (PISCES) has developed sintered materials under four different thermal profiles (1,120, 1,135, 1,150 and 1,180 °C) using Hawaiian basalt and Hawaiian basalt scoria. The structural properties of some of these materials exceed the strength of both residential and specialty concrete, showing promise for sintered materials for in-situ resource utilization application. PISCES has researched and characterized basalt samples from multiple locations on Hawai'i Island and has shown that chemical composition and mineral abundances significantly impact the durability and outcome of a sintered basalt material. Identifying desirable basalt characteristics for sintering will aid in the search for locations and sources of optimal basalt regolith for sintering on the Moon and Mars. New sintering technologies in particular may also want to be considered, moving forward, to optimize the feasibility of sintering in space.

K. P. Edison (✉)

Colorado School of Mines, 1310 Maple St., GRL 234, Golden, CO 80401, USA

e-mail: Kedison@mines.edu

G. Jeffery Taylor

7012 E Calle Morera, Tucson, AZ 85750, USA

e-mail: gjtaylor@higp.hawaii.edu

C. B. Andersen

PISCES, 200 W. Kawili St., Hilo, HI 96720, USA

e-mail: canderse@hawaii.edu

R. F. V. Romo

1445 Waiianuenue Ave, Hilo, HI 96720, USA

e-mail: rromo@hawaii.edu

© Springer Nature Switzerland AG 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,

https://doi.org/10.1007/978-3-030-97913-3_13

13.1 Introduction to Sintering

In-Situ Resource Utilization (ISRU) refers to the ability to utilize local resources found in an area of interest. In space exploration and settlement, ISRU aims to utilize resources found on the Moon and Mars to build infrastructure, thereby reducing the need for materials from Earth. Roadways, sidewalks, foundations, and other structural components on Earth are typically made with bitumen, cement concretes, ceramics, and metals. Building upon centuries of knowledge and evolving methods, these materials are mined, refined, and processed. Terrestrially, their acquisition and use are reliant on accessibility and abundance in concentrated locales, energy availability, and the capacity to extract and refine the desired components from raw feedstock.

As mankind embarks on long-term exploration and habitation on the Moon and then Mars, many of the terrestrial processes for material production will not be readily adapted to these novel environs. On the Moon and Mars, we do not yet know of ore deposits, nor do we expect that early expeditions will have the equipment and energy needed for their extraction and refinement. The scarcity of hydrated minerals and water on the Moon makes cement products unfeasible. Water ice has been identified in lunar polar regions, but it is too valuable as propellant (Kornuta et al. 2018) and life support to use in construction. The lack of hydrocarbons on the Moon and their unlikely prevalence on Mars means that bitumen is a non-starter on the Moon, and at best, a distant possibility on Mars.

The one resource that is known to be abundant on both the Moon and Mars is basalt. Basalt is a mafic (fine-grained <1 mm) extrusive (volcanic) igneous rock containing 42–53% SiO₂, comprised mostly of plagioclase feldspar and pyroxenes, and may contain olivine (Tietz and Büchner 2018). Basalt is formed by the rapid cooling of magnesium/iron-rich lavas exposed on the surface of terrestrial planets and is a common rock type found on Earth, the Moon, and Mars.

Basalt has a terrestrial history as a structural material. In ancient times, basalt was used to build rock walls, buildings, and foundations. In the early twentieth century manufacturers in Europe developed methods to cast basalt that resulted in materials that were strong, nonporous, and chemically inert. Basalt casting should be feasible on the Moon and Mars; however, it is extremely energy intensive. An alternative to basalt casting that can enable early construction and part production is basalt sintering.

Sintering is a process that consolidates a collection of particles into an agglomerate through heating. There are many variables that contribute to the strength, uniformity, and porosity of the finished product. Particle sizes, grain orientations, particle shape, chemical and physical homogeneity are all factors that can affect material strength, densification, porosity, and shrinkage. Most sintering is either done as a solid-state or a liquid-phase process. Solid-state sintering utilizes atomic diffusion between particles, whereas in liquid-phase sintering a portion of the particles enter a liquid phase and bond to solid particles forming a liquid–solid matrix that is strengthened further by accelerated diffusion.

Sintered materials are of particular interest for sustained habitation and exploration on the Moon and Mars. This interest in sintered materials is due to potential use of abundant and readily available in-situ resources. Rather than a reliance on more complicated chemical processes, regolith can be used with minimal preparation once it has been mined, sieved, and sorted. In the same way that early human civilizations first sintered available resources and materials to make pottery, early lunar and Martian pioneers will sinter regolith to make infrastructure and early structural elements. For early terrestrial civilizations, the need to use in-situ resources stemmed from a limited access to resources and a developing understanding of mining, refining, and extraction of ores and chemical binding methods. The early lunar and Martian explorers will likewise be bound by limited payload capacities and energy restrictions that will prevent or impede their capacity to bring terrestrial building materials or to replicate their production in situ. It is these constraints that make regolith sintering so desirable in the early habitation and exploration of the Moon. With a larger variety of available minerals and water-facilitated weathering, Mars offers a larger variation in the methods of production and the types of materials that can be produced. Nonetheless, with the current plans to explore and develop technologies and habitats on the Moon as the foundation and testbed for Mars exploration, it is likely that early Martian construction will rely on lessons and technologies developed for the Moon.

The Moon's surface varies in composition (see Sect. 13.4), ranging from the highlands with an abundance of plagioclase that averages about 80 wt.% to the maria basalts which have only 30–35 wt.% plagioclase. The regolith also contains a significant amount of glass produced by micrometeorite impacts, often accompanied by glass spherules produced by volcanic eruptions billions of years ago. The lunar regolith, unlike Mars, lacks hydrated minerals, except for small quantities of OH in apatite. Grain-size samples of lunar regolith from the Apollo and Luna missions had medians that ranged from 42 to 800 μm (Heiken et al. 1991). Because sintering strengths tend to be optimized with small initial grain sizes with minimal size variation, lunar regolith will need some particle sorting or sieving prior to sintering.

The heterogeneous nature of lunar regolith (varied minerals and amorphous glass) means that different components within constituent particles have various melting points. Due to the range of component melting points in lunar regolith, in-situ liquid-phase sintering can produce a composite with low pre-processing requirements and lower energy inputs than casting from completely melted regolith.

Liquid-phase sintering of lunar basalts occurs when lower melting point constituents liquify and fill the voids between the particles that remain solid. This action has the advantage of reducing voids and increasing contact through the liquid–solid interface and thus increasing diffusion by 2–3 orders of magnitude (German 2014). The accelerated diffusion and the network of liquid constituents forms a low-porosity high-strength material upon cooling.

High-strength sintered basalt offers a feasible solution for mobility surfaces, landing and launch pads, thermal wadis, and habitats for early exploration and habitation. Sintering of the lunar regolith has been studied before (e.g., Allen et al. 1994; Taylor and Meek 2005; Indyk and Benaroya 2017; Fischer 2018; Meurisse et al. 2018;

Song et al. 2020; Zocca et al. 2020). In this chapter we highlight the value of doing experiments with basalts from Hawai'i, which are readily available from commercial quarries, summarize the main compositional features of lunar and Martian regolith, and present results from sintering experiments on Hawaiian basalts.

13.2 Background on Sintering of Hawaiian Basalt

Between the fall of 2015 and the spring of 2016, the Pacific International Space Center for Exploration Systems (PISCES), a Hawai'i state agency, in collaboration with a team from NASA's Kennedy Space Center Swamp Works, Honeybee Robotics, and the County of Hawai'i R&D Department, joined forces to build a full-scale vertical takeoff/vertical landing (VTVL) pad in Kea'au, Hawai'i. The Additive Construction with Mobile Emplacement (ACME) project, In-Situ Vertical Takeoff/Vertical Landing (VTVL) Pad task was a joint venture between NASA's Space Technology Mission Directorate (STMD) Game Changing Development (GCD) Program, NASA Kennedy Space Center, and the Pacific International Space Center for Exploration Systems (PISCES).

The project goal was to design, telerobotically build, and test a landing pad using local resources of similar characteristics/composition as lunar regolith. The site selected for the construction of the VTVL was in Kea'au on the island of Hawai'i. The pad was built with interlocking tiles made with Hawaiian sintered basalt. The basalt used to make the tiles was procured from a local commercial quarry (19°42'35.66"N, 155° 3'19.00"W). Tiles had been sintered at 1,150 °C (Fig. 13.1).

13.3 Basalt Paver Manufacturing

Pavers were designed for the 3-m-by-3-m VTVL surface such that they would interlock and be thick enough to provide adequate compression strength. The pavers were constructed from sub-150 μm basalt fines purchased from the quarry as a waste material, with no additives, obviating the need to sieve. Chemical composition of the basalt used in the project analytical details are given in Sect. 13.5.1 (Fig. 13.2).

Paver production was a complicated task, turning out to be more of an "art" in defining the proper thermal profile, particle sizes, and mold design for thoroughly sintering the basalt without creating cracks/breakages in the pavers. The molding material used was Cotronics RESCOR 750 castable silica ceramic. It was selected because it can withstand up to ~4,900 °C, has low thermal conductivity and low thermal expansion. The spring/summer 2015 timeframe saw a steep learning curve in designing a thermal profile that would create a suitable basalt paver. The resultant thermal profile was an approximately 30-h run time in a high-temperature kiln, with a maximum sintering temperature of 1,150 °C. Early paver prototypes consistently failed in the same manner and revealed lateral stress areas where the pavers were



Fig. 13.1 Paver Deployment Mechanism (PDM) on PISCES planetary rover

pulling against the corners of the molds upon thermal contraction of the paver. These failures were resolved by making part of the molds “float” such that the molds would slip upon paver contraction. This was achieved by carving notches at the corners of the molds.



Fig. 13.2 Interlocking pavers made with Hawaiian sintered basalt

Before making the molds with the ability to slip, only about 10% of the pavers were intact by the end of the run time. The post-modification intact rates improved to 50% but revealed secondary stresses that accounted for the remaining failures. These secondary failures were due to vertical stress from the pavers contracting and pulling against the ledges of the mold. Due to the geometric constraints of the molds, additional floating modifications were not feasible. To reduce vertical contraction, the sub-150 μm basalt fines were mixed with #4 basalt mortar sand, which has a grain size of ≤ 4.75 mm. The mortar sand was sourced from the same quarry as the sub-150 μm fines and was assumed to have similar chemical and mineral properties. Because sintering primarily melts and acts on the margins of particles, and larger particles have a higher volume-to-surface ratio, the addition of larger particles reduced the overall contractions.

Various mix-ratios were tried; a 50% sand/fine ratio was found to offer the best performance with minimal shrinkage, and overall failure rates were $<10\%$. The overall profile was finalized in September 2015, paver production started in October 2015 and was completed in December 2015.

13.3.1 Basalt Paver Prototype Structural Analysis

One of the pavers manufactured by PISCES was sent to NASA's Engineering Directorate Laboratories and Test Facilities Division at Kennedy Space Center, Florida, to undergo flexural and compressive test analysis.

The flexural strength (modulus of rupture) test was conducted per ASTM C133, Standard Test Methods for Cold Crushing Strength and Modulus of Rupture of Refractories. Flexural strength specimens were prepared to be approximately 6.0 by 1.0 by 1.0 inches. The samples were then cut to shape using a Buehler Delta Manual Abrasive Cutter circular saw with a diamond-tipped cut-off wheel.

Figure 13.3 shows the test set-up of the flexural stress test. The top roller is placed at the halfway point of the specimen, and the bottom rollers have a span (L) of 5.0 inches. The specimen is preloaded to about 10 lbf before starting the test. Once the test begins, the load is applied with a loading rate of 174 pounds of force per minute (lbf/min).

The cold crushing strength, or compressive strength, test was also conducted in accordance with ASTM C133. The specimens used in the flexural stress tests were cut into cubes that were about 1.0 by 1.0 by 1.0 inches. The samples were preloaded to 250 lbf before the test started. The specimens were compressed with a loading rate of 1750 pounds of force per minute (lbf/min).

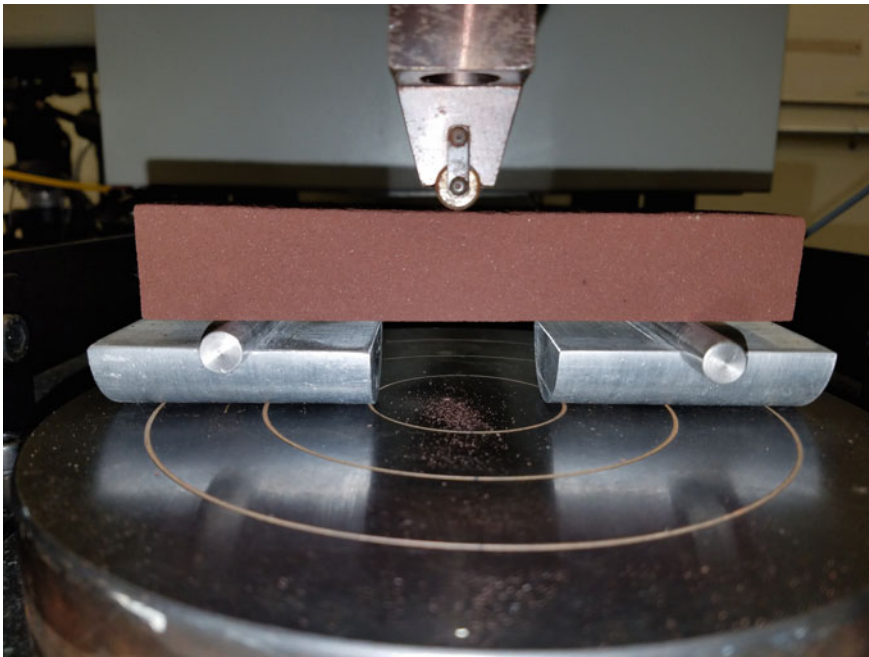


Fig. 13.3 Flexural testing of sintered basalt sample

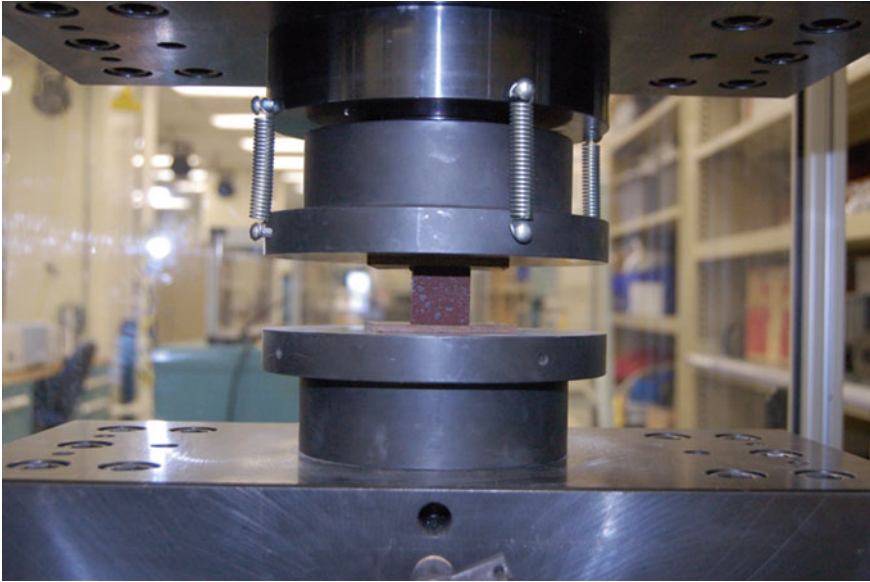


Fig. 13.4 Compressive testing set-up for Hawaiian sintered basalt

Figure 13.3 is a picture of the compressive stress test set-up using a Hawaiian paver specimen. Masonite compressed fiberboard was used between the specimen and the compression platforms to prevent premature failure from point loading of specimen irregularities (Fig. 13.4).

13.3.2 Basalt Paver Structural Analysis

According to “Concrete in Practice” (2014), document 35 from the National Ready Mixed Concrete Association (NRMCA), the compressive strength of concrete can vary from 17.2 MPa (2500 psi) for residential concrete to 27.6 MPa (4000 psi) for commercial structures. Some applications can achieve or exceed 69 MPa (10,000 psi). According to “Concrete in Practice” (2016) document 16 from NRMCA, the flexural strength of concrete is about 10–20% of compressive strength. This is highly dependent on the type, size, and volume of the coarse aggregate used in the mixture. This suggests that the maximum flexural strength of concrete can vary from 3.4 MPa (500 psi) for residential concrete to 5.5 MPa (800 psi) for commercial structures and 17.2 MPa (2000 psi) for special applications. Since concrete compression testing is performed per ASTM C39 (not ASTM C133), any comparison of the data should only be generalized. Figure 13.5 shows the first-generation Hawai’i basalt strength testing results. The structural results indicate that the 1,150 °C material is not almost on a par with commercial concrete. This was a surprising result and was the reason

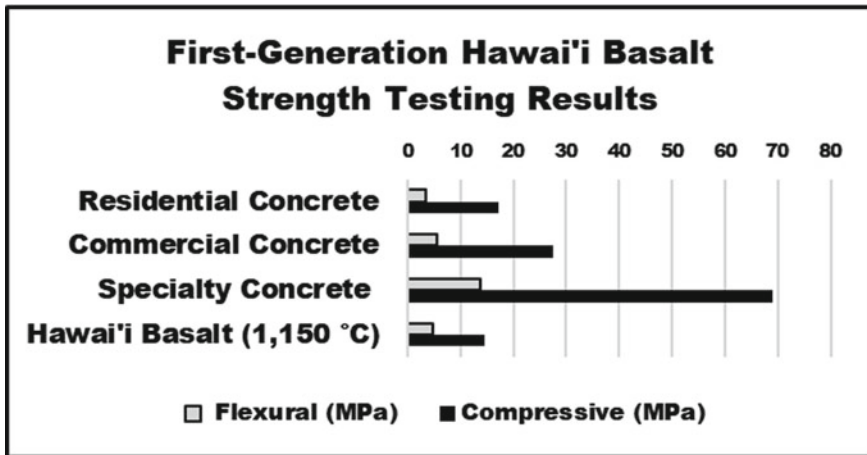


Fig. 13.5 Comparison of Hawaiian basalt with flexural and compressive strength of concrete

for the eventual collection and characterization of basalts from various flows and volcanoes on the island.

13.3.3 Higher-Density Sintered Material

As with many unintended scientific discoveries, during the ACME project, to produce a sintered sample in a shorter period, the sintering temperature was increased by 30–1,180 °C. This slight increase in temperature produced a material with different coloration, higher density, and improved structural properties. The new material also presented new challenges in its manufacturing process due to shrinkage, adhesion to the molds, and severe cracking during the cooling process. Figure 13.6 displays the second-generation material being tested for flexural and compressive strength. Figure 13.7 compares the 1,180 °C Hawaiian basalt sinter with flexural and compressive stress with concrete. The accompanying Table 13.1 displays the data and compares both the 1,150 °C material and the 1,180 °C materials to residential, commercial and specialty concretes and shows that the 1,180 °C material has ~3 × greater compressive strength than specialty concrete.

13.3.4 Comparing Basalts from Different Locations and Sources

After the successful sintering of the basalt from a local commercial quarry designated as Quarry 1, the ACME project, PISCES and Honeybee Robotics secured funding

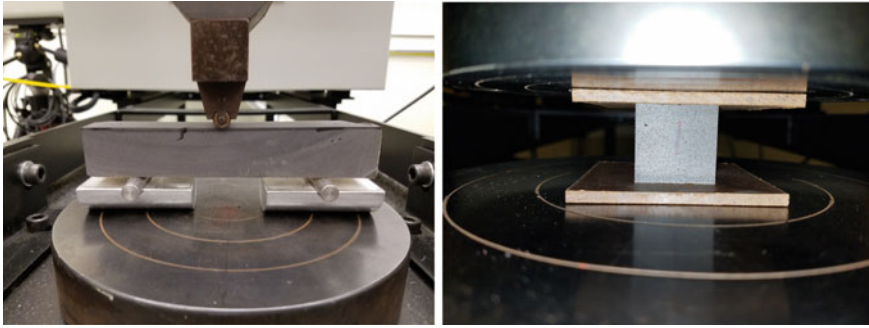


Fig. 13.6 Displays the second-generation material going through a flexural and compressive test

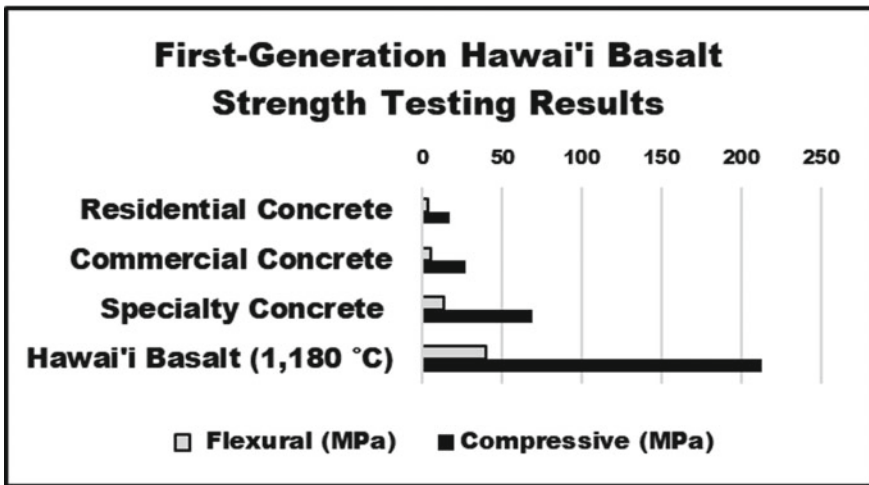


Fig. 13.7 Compares the 1,180 °C Hawaiian basalt sinter with flexural and compressive stress with concrete

Table 13.1 Comparison of the first- and second-generation materials to the flexural and compressive strengths of concrete as per ASTM C133

Test	Residential concrete	Hawai'i 1,150 °C	Commercial concrete	Specialty concrete	Hawai'i 1,180 °C
Flexural stress (Mpa)	3.44	4.93	5.51	13.79	40.5
Compressive stress (Mpa)	17.24	21.48	27.58	68.95	212.5
Density (g/cm ³)		1.7			2.64

through a 2017 NASA STTR project to further develop the sintering technology to allow for vertical construction. Part of the objectives of this project included evaluating basalt from other sources on the island to determine whether they could provide the same structural properties when sintered.

For this project, samples from five different locations on the island of Hawai'i were selected. The sample locations included the same commercial quarry that provided the material for the ACME project, a commercial quarry located in Kea'au, a commercial/military quarry located on the saddle area between Mauna Kea and Mauna Loa, a private quarry located in Kona, and an analog test site on Mauna Kea previously used for ISRU field tests. Immediately after sintering the samples from the various locations at 1,150 °C the differences in the sintered material properties were obvious and significant. Sintered samples ranged from compact high-density material to a highly loose, unsuitable, and easy-to-crumble material. Details of the structural properties as well as the mineralogy of the samples and the relation to the sintering product are discussed in detail in this chapter.

13.4 Background on Lunar and Martian Regolith

Regolith is an old geological term for the layer of fragmental and unconsolidated rock material, whether residual or transported and of highly varied character, that nearly everywhere forms the surface of the Earth. The Moon and Mars are also blanketed with fragmental layers. The chemical compositions and mineral abundances in planetary regoliths vary as they are affected by the composition of the underlying bedrock and by the environments at their surfaces.

13.4.1 Properties of Lunar Regolith

The lunar regolith was generated from the underlying bedrock by meteorite impacts over billions of years. Its depth ranges from 2 to 30 m, with most areas in the range 5–10 m. It contains rock and mineral fragments and impact-produced glass. It also contains porous particles called agglutinates, which are glass-bonded aggregates of rock and mineral fragments. Agglutinates are produced by micrometeorite impacts into the lunar regolith and their abundance increases with the amount of surface processing a given patch of regolith has experienced. The regolith is a complex mixture of these components. Single mineral grains are rare; almost all are either coated with agglutinitic glass or are attached to other minerals in rock fragments. The extensive reworking by impacts has rendered the original bedrock into a powder with rock fragments. Its mean grain size ranges from 40 to 800 μm (McKay et al. 1991). Inferring from diagrams in McKay et al. (1991), about 20 wt. % of the regolith in the <1 mm size fraction is composed of particles smaller than 20 μm ; about 10 wt. % is smaller than 10 μm .

Regolith properties vary with the amount of reworking experienced by a given patch of regolith. With increasing time of exposure to the space environment, the regolith matures. As maturity increases, the abundance of agglutinates increases, the mean grain size decreases, and the abundance of solar wind implanted volatiles (H, He, C, N) increases. A particularly important product is the formation of nanometer-sized metallic iron grains in silica-rich glassy rims of minerals. The metal grains form by reduction of FeO in minerals by solar wind hydrogen during micrometeorite impacts. In mature lunar soils, most mineral grains contain nanometer metallic iron grains on their surfaces. Microwaves couple readily with the nanophase iron, causing rapid heating (Meek 1987).

The chemical composition and mineral abundances of the lunar regolith reflects the composition of the underlying bedrock. We have assembled some representative average compositions of the regolith sampled at all six Apollo sites (Table 13.2). The chemical data are taken from Lucey et al. (2006) and are averages from sampling locations representative of geologic features or are typical for the entire site. Mineral abundances are from quantitative X-ray diffraction measurements of 118 Apollo soils (Taylor et al. 2019). Each value averages the same samples as used for the chemical analyses. Mare sites Apollo 11, 12, 15, and the valley floor at Apollo 17 are high in FeO and low in Al₂O₃. The mare sites typically have higher TiO₂ than other sites, especially at Apollo 11 and 17. The feldspathic highlands (Apollo 16) are low in FeO and high in Al₂O₃. The Apennine Front at Apollo 15 and South Massif at Apollo 17 are intermediate in FeO and Al₂O₃, as is the Apollo 14 site. Mineral abundances reflect these bulk chemical variations: high TiO₂ samples have higher ilmenite abundances and high Al₂O₃ reflects a high abundance of plagioclase feldspar. Glass abundances are high, indicating that these are averages of mature soils. Glass compositions (not shown) are roughly the same as the soil they are in, though there are subtle variations in soil compositions with size fraction.

To plan sintering experiments, we need to know the melting temperatures of the lunar regolith. Few melting experiments have been done on regolith compositions. Nevertheless, numerous experiments have been done on appropriate chemical systems and on lunar basalt compositions that allow us to understand the melting of materials with compositions like those of the lunar regolith. Walker et al. (1972) determined the solidus and liquidus temperatures of Apollo 14 soil 14,259. The experiments indicated that the solidus temperature at 1 bar pressure for this composition is ~1,150 °C and the liquidus is between 1,242 and 1,252 °C. This soil is like the typical Apollo 14 soil (Table 13.2). Other Apollo soil samples fall in the region of the olivine-plagioclase-SiO₂ phase diagram, as does soil 14,259, suggesting that the solidus for all lunar soils is around 1,150 °C. Thus, to do liquid-phase sintering, temperatures would need to be >1,150 °C. If total melting is required for producing a product from the lunar regolith, we need to know the liquidus temperatures of the regolith. Experiments on mare basalt compositions allow us to estimate the likely liquidus temperatures of lunar basaltic regolith. Typical liquidus temperatures for lunar mare basalts range from ~ 1280 °C for low-Ti mare basalt to 1,170 °C for high-Ti mare basalts. Highland regolith, especially the feldspathic highlands, probably have higher liquidus temperatures, possibly as high as 1,500 °C.

Table 13.2 Chemical compositions (Lucey et al. 2006) and mineral abundances (Taylor et al. 2019) of representative lunar regolith

	A-11 typical	A-12 typical	A-14 typical	A-15 mare	A-15 Apen. front	A-16 Plains	A-17 valley floor	A-17 South Massif
<i>Chemistry (wt.%)</i>								
SiO ₂	42	46.1	47.7	46.2	46.5	44.9	39.9	45.1
TiO ₂	7.5	2.7	1.7	2	1.4	0.59	9.6	1.3
Al ₂ O ₃	13.5	12.6	17.4	10.4	16.5	26.7	10.9	21.3
Cr ₂ O ₃	0.3	0.38	0.2	0.53	0.34	0.11	0.46	0.22
FeO	15.8	16.5	10.5	19.8	12.3	5.44	17.7	8.3
MnO	0.21	0.21	0.14	0.25	0.16	0.07	0.24	0.11
MgO	7.9	10.2	9.4	11.1	10.9	6	9.5	9.8
CaO	12	10.3	10.9	9.6	11	15.3	10.7	12.9
Na ₂ O	0.44	0.46	0.7	0.3	0.47	0.48	0.38	0.43
K ₂ O	0.14	0.24	0.52	0.094	0.18	0.12	0.78	0.14
P ₂ O ₅	0.1	0.3	0.49	0.11	0.18	0.12	0.07	0.13
Total	99.89	99.99	99.65	100.38	99.93	99.83	100.23	99.73
<i>Mineralogy (wt.%)</i>								
Plagioclase	18.8	20.5	28.7	18.8	26.6	54.7	19.2	45.3
Olivine	0.3	5.7	3.4	0.3	6.4	5.3	4.9	10.6
Augite	22.6	15.2	8.4	16.9	5.3	1.5	15.9	5
OPX	0	2.7	4.9	2.1	4.1	0.4	0.6	5.1
Pigeonite	9.2	18.8	7.5	24	10.8	5.9	12.8	7.1
Ilmenite	12	4.5	5	1.4	3	0.2	11.5	2.8
Chromite	0.4	1.4	0.4	1.3	0.7	0	0.8	0.8
Quartz	0	0.8	1.1	0.4	0.5	0	0.3	0.5
Cristobalite	0.8	0.6	0.3	0.8	0.6	0	1.2	0.5
Whitlockite	0.1	1	1.6	0.2	0.6	0	0.1	0
Glass	35.9	28.6	38.8	33.8	41.3	32	32.7	22.3
Total	100	100	100	100	100	100	100	100

The exceptionally low water content in the Moon and in magmas that were produced inside the Moon, e.g., Robinson and Taylor (2014), means that hydrous minerals are rare in Apollo regolith samples. The most conspicuous one is apatite, and even that contains $\ll 1$ wt.% H₂O, usually < 0.01 wt.%. Liu et al. (2012) show that the glass in agglutinates contains OH, but < 500 ppm (< 0.05 wt.% H₂O equivalent). Common hydrous minerals such as phyllosilicates (clay minerals) or oxy/hydroxides such as goethite, are absent in Apollo samples. It is possible that such minerals are present in polar regions where H₂O ice is present in places in permanent shadow, if the ice can be mobilized as liquid water (Stopar et al. 2018), but these have not been

detected yet. For designing sintering processes, it is safe to assume that the water content of the lunar regolith is trivial.

13.4.2 Properties of the Martian Regolith

Martian regolith, as examined by orbital and landed spacecraft, differs dramatically from lunar regolith. It is predominantly composed of basaltic rocks and the regolith formed from them, as shown by the global coverage provided by the Mars Odyssey Gamma-Ray spectrometer (GRS) and analyses from the Pathfinder rover (Sojourner), Mars Exploration rovers (Spirit and Opportunity), and the Curiosity rover of the Mars Science Laboratory. The GRS average for the Martian surface (McSween et al. 2009) falls unambiguously in the basalt field on a plot of $\text{Na}_2\text{O} + \text{K}_2\text{O}$ versus SiO_2 (Fig. 13.8). Three of the average soils (unconsolidated fine-grained materials at the landing sites) also plot in the basalt field, but the one from Gale Crater is slightly to the low-silica side of the basalt boundary. The average compositions of the soils from the rover missions are given in Table 13.3. Their similarity in composition suggests that aeolian (windblown) processes may homogenize the dusty part of the regolith, with modifications from alteration of local rocks (O'Connell-Cooper et al. 2017).

Some of the same processes that operate on the Moon also process the Martian regolith: comminution by meteorite impacts, irradiation by solar flare particles and cosmic rays. However, the most significant difference between regolith on the Moon and Mars is that flowing water shapes the landscape and reacts with primary igneous rocks to produce weathering products. Mars has a much higher concentration of water than does the Moon. GRS data (Boynton et al. 2008) show that the mean H_2O concentration in the upper meter of the Martian surface is 3.9 ± 0.3 wt.%. This is in large part accounted for by phyllosilicates, hydrated Ca- and Mg-sulfate minerals, and minor hydrous iron oxides. Hematite (Fe_2O_3) and magnetite are also present. Besides higher water abundances compared to the Moon, the Martian crust also contains enrichments in sulfur (oxidized) and chlorine. Aqueous alteration of primary igneous minerals is common at the landing sites. (See thorough review of the mineralogy at the Gale Crater site by Rampe et al. 2020). Thus, during sintering these volatiles are likely to outgas from the regolith and almost certainly react with minerals.

13.5 Variation in Chemical Composition/Mineral Abundances of Hawaiian Basalt

Hawaiian basalts tend to be in several respects similar in composition both chemically (Fig. 13.8) and mineralogically to regolith on the lunar maria and the Martian surface, allowing Hawai'i to be used as an analog site and proof-of-concept testing

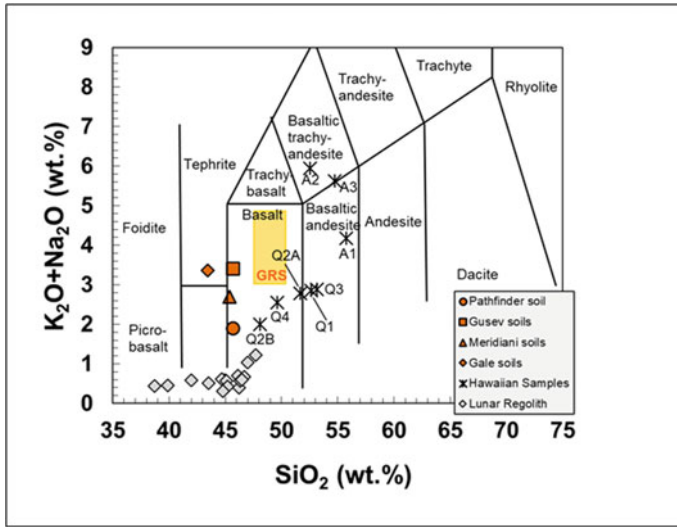


Fig. 13.8 Total alkalis versus silica diagram that is used for classification of terrestrial volcanic rocks. GRS data rectangle represents the global average abundance plus/minus 2-sigma standard deviation of the mean for the Martian crust details in McSween et al. (2009). Sources for Mars soil means are given in Table 13.3 Hawaiian samples (our data) are described below. The points are labeled according to the numbering system shown in Table 13.3. Q stands for quarry; A stands for analog. The lunar regolith tends to be low in alkalis, but this should not affect sintering behavior significantly. Also plotted are average lunar regolith samples from different sampling sites data from Lucey et al. (2006). Lunar samples are near the basaltic range, but it is important to note that samples from the feldspar-rich lunar highlands are unlike basalts when all elements are considered

area for basalt sintering and material development that may be implemented for future lunar/Mars ISRU applications. On Earth there are two major magma series or classifications of basalt based on chemical composition: tholeiitic (sub-alkaline) and alkaline. Tholeiitic basalts are characterized as being saturated or slightly over-saturated with silica compared to alkalis and are rich in iron and largely made up of calcic-plagioclase and low-calcium pyroxene. Alkaline basalts are characterized by higher amounts of Na₂O and K₂O and contain olivine and titanium-rich augite in groundmass as phenocrysts (Winter 2013, p. 159). Due to basalts varying in chemical composition and mineral abundances, it is important to know if all basalts are suitable for sintering to make durable construction materials regardless of composition, and if not, what basalts are most suitable?

The first step in determining if all basalts can be sintered into a useful product was to collect quarried aggregate samples. A geologic map overlay in Google Earth Pro (Fig. 13.9) was used to characterize basalts based on the volcano of origin, location of flow, age, and magma series (chemical composition of originating magma). The map colors indicate the lava flows, their origins and year in which they erupted.

Figure 13.9 shows the locations of the collected samples and location of volcanoes.

Table 13.3 Chemical compositions of Martian soils from three rover missions: Ares Vallis (Pathfinder), Gusev Crater (MER-A), Meridiani Planum (MER-B), and Gale Crater (MSL). Data for Pathfinder is from Brückner et al. (2008) and the others are from O’Connell-Cooper et al. (2017)

	Ares Vallis	Gussev	Meridiani	Gale
SiO ₂	41.5	45.9	45.4	43.5
TiO ₂	0.8	0.9	1.1	1
Al ₂ O ₃	10.2	10.1	9.1	9.4
Cr ₂ O ₃	0.3	0.3	0.4	0.4
FeO	20.2	16	18.9	18.7
MnO	0.3	0.3	0.4	0.4
MgO	7.4	8.8	7.3	8.4
CaO	6	6.3	7	7
Na ₂ O	2.8	3	2.2	2.8
K ₂ O	0.6	0.4	0.5	0.6
P ₂ O ₅	0.8	0.9	0.9	0.9
SO ₃	6.1	6.2	6	6
Cl	0.9	0.7	0.7	0.8
Total	97.9	99.8	99.9	99.9

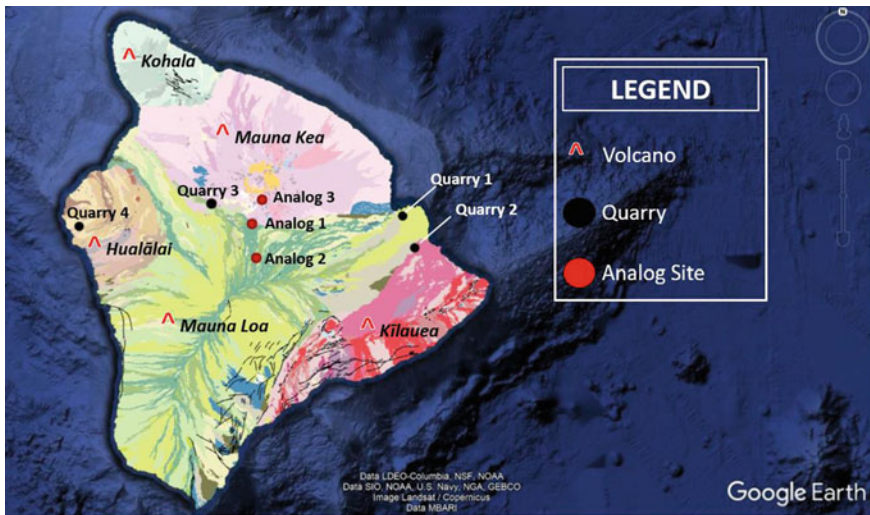


Fig. 13.9 Geologic map of the island of Hawai'i showing the locations of the collected samples and their volcano of origin

13.5.2 Thin-Section and X-ray Diffraction (XRD) Analysis of Basaltic Parent Rock Samples

Whole-rock samples of each aggregate were collected for thin-section analysis and X-ray diffraction analysis; with XRD it is possible to determine the weight percentage of minerals in a sample. We used the same analytical equipment and data reduction procedures (Rietveld refinement) used by Taylor et al. (2019). It is important to know what minerals make up the matrix of a sample, then later to track changes of those minerals through the sintering process. This will help determine the appropriate rock composition that is best for sintering. Thin sections were also used to validate the findings of the XRD. Samples Quarry 1, 3, 4, and Analogs 1 and 2 are the only samples that were analyzed via thin section and XRD. Quarry 2 samples were not analyzed because of limited supply and can no longer be obtained.

Quarry 1. According to the XRD analysis, samples from Quarries 1 and 3 are similar and when sintered produce a material that is identical. Due to ease of access, Quarry 1 has been used exclusively to produce material and therefore will be the representative parent rock for this group. The thin section shows a plagioclase- and pyroxene-dominated matrix with interstitial opaque minerals. Two large phenocrysts of olivine are displayed in the image showing a clear conchoidal fracture and 3rd-order birefringence. The XRD analysis reflects minerals as seen in the thin section in that there is a high abundance of calcic-plagioclase and moderate abundance of clinopyroxene. It does not, however, show the presence of olivine. Despite its prominence in the photomicrograph, olivine was not detectable (<0.1 wt.%) in our XRD analyses. Figure 13.10 displays the Quarry 1 XRD data and compares it to its thin section.

Quarry 4. The matrix is dominated by the presence of olivine and plagioclase with interstitial opaque minerals. This sample's olivine is not isolated to phenocrysts like

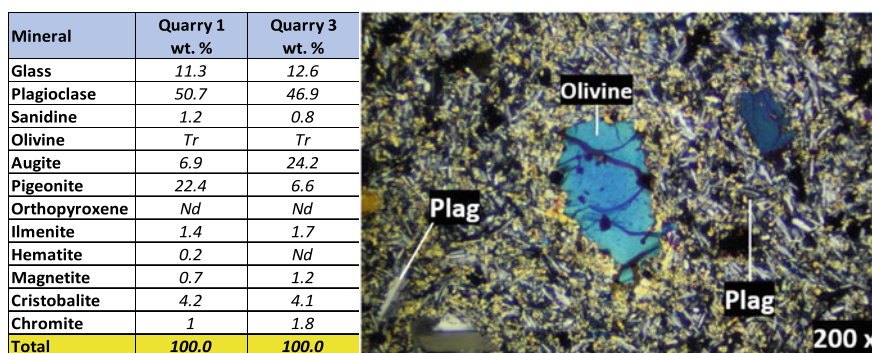


Fig. 13.10 Quarries 1 and 3 XRD data accompanied with thin section of Quarry 1 displaying a dominant amount of plagioclase and a moderate amount of pyroxene and the two phenocrysts of olivine. Despite its prominence in the photomicrograph, olivine was not detectable (<0.1wt. %) in our XRD analyses. Nd means “Not detected”

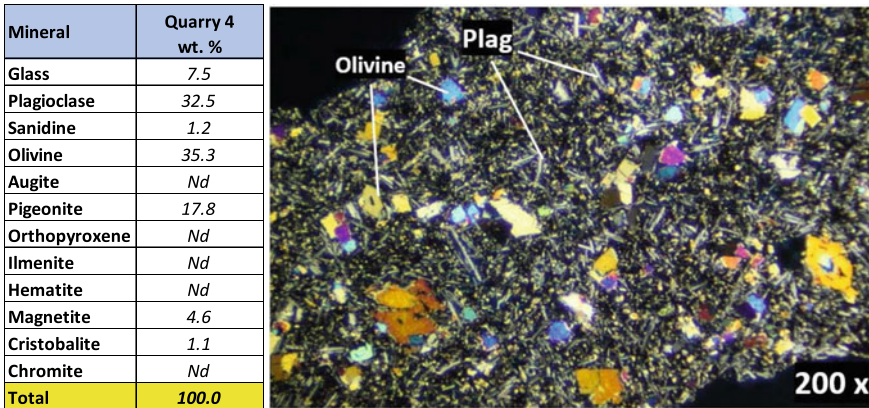


Fig. 13.11 Quarry 4 XRD data with accompanying thin section displaying a large amount of olivine crystals as characterized by the conchoidal fracturing and 3rd-order birefringence, and a matrix of plagioclase feldspar

Quarries 1 and 3, rather is found as a part of the sample’s matrix. This is confirmed by the XRD analysis with 33.6% of forsterite (magnesium-rich) olivine and calcium-rich plagioclase. Figure 13.11 displays the Quarry 4 XRD data and compares it to its thin section displaying a large percentage of olivine crystals throughout the sample.

Analog 1. Analog 1 was the only basalt analog analysis done due to its ease of access and due to its use as the dominant sample of its kind for sintering. This sample is not a true basalt but rather basalt scoria. Scoria forms when magma containing abundant dissolved gas is violently blown out during an eruption; the pressure upon it is reduced and the dissolved gas starts to escape in the form of bubbles. Scoria is predominantly made up of a large percentage of glass, vesicles, and fractured crystals. The thin section shows the matrix is predominantly made up of opaque glass and large vesicles. Small amounts of fractured plagioclase are also present. The XRD analysis confirms this and shows that there is a large amount of volcanic glass, calcic-plagioclase and a fair amount of pyroxene which cannot easily be identified by the thin section. Figure 13.12 displays the XRD data and thin section of Analog 1 displaying a plagioclase agglomerate in the center surrounded by glass.

The above XRD results give a general idea of what minerals each parent rock contains. A much more accurate mineral identification is possible by using an SEM fitted with EDS detectors, but these capabilities were not available. More testing is needed in the future.

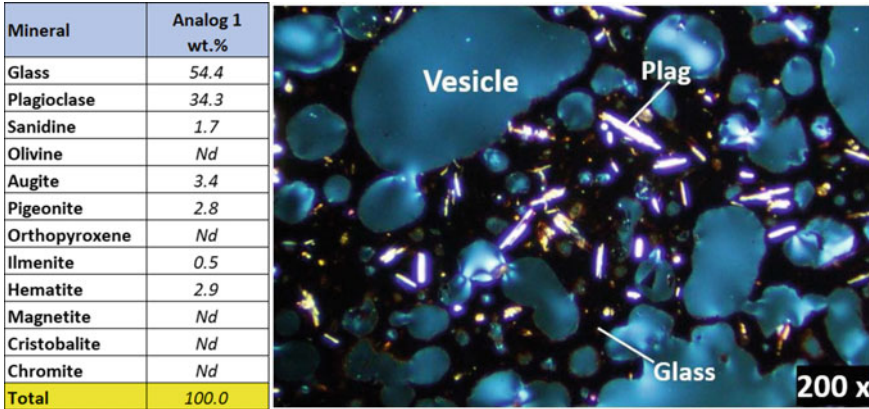


Fig. 13.12 Analog 1 XRD data accompanied by its thin section displaying a large abundance of vesicles (blue), opaque glass, and a patch made up of predominantly of plagioclase

13.5.3 Thin-Section Analysis of Sintered Materials

The basalt and basalt scoria samples were sintered at 1,150 and 1,180 °C. This created two different materials of varying strengths and durability. Both types were cut into thin sections to fully understand the bonding and sintering effects of the material.

13.5.3.1 Quarry 1

1,150 °C: The sample is densely vesiculated, with a scattering of black minerals which are assumed to be olivine and possibly pyroxene, with partially visible plagioclase. There is a noticeable cohesion between grains illustrating a change from the parent rock to the new sintered material. Figure 13.13 displays the Quarry 1 sintered material.

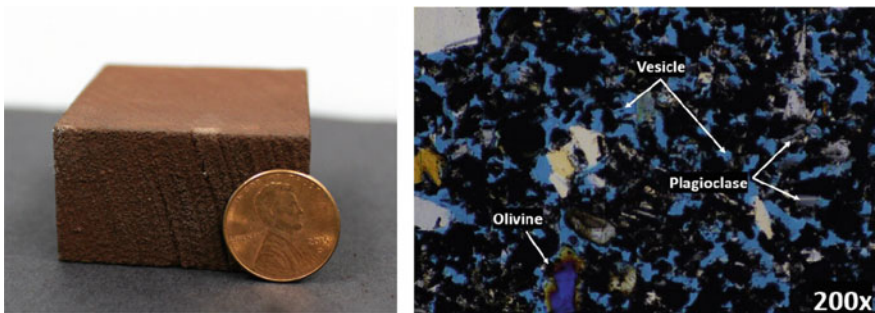


Fig. 13.13 Quarry 1 low-temperature sintered material (left) and thin section (right)

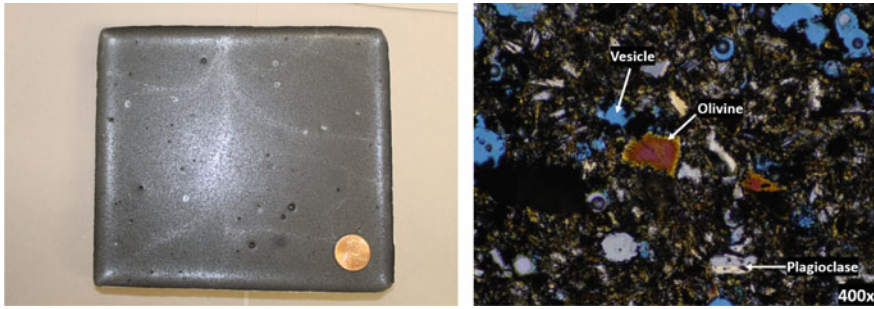


Fig. 13.14 Quarry 1 high-temperature sintered material (left) and thin section (right)

1,180 °C: Material is much less vesiculated than the 1,150 °C material and has a denser mineral/grain cohesion. Plagioclase is easily identifiable and packed tightly to surrounding minerals. Quarry 1 parent rock displays a dense amount of plagioclase within its matrix. When sintered this material becomes highly vesiculated at lower temperatures, indicating poor grain cohesion, whereas at higher temperatures there is a denser grain cohesion and fewer vesicles. It appears that high temperature and longer time allows for the grains/minerals to bond tightly, creating a stronger, more durable material. The actual materials are sintered well and are durable. The lower-temperature material is more porous than the higher-temperature material which has unique structural qualities. Figure 13.14 displays the high-temperature Quarry 1 sintered material, showing a noticeable change from the low temperature and the parent rock thin sections shown in Fig. 13.10.

13.5.3.2 Quarry 4

1,150 °C: Highly vesiculated material with little to no cohesion of grains. The thin section displays noticeably altered plagioclase and a higher volume of what is assumed to be olivine, given certain clues by the thin sections. The thin sections show a few olivine crystals that have not completely altered to black which still display some of their third-order birefringence and conchoidal fracture. Due to the large amount of olivine that was observed in the parent rock, it is likely that most of the blackened minerals present in the sintered thin section are the olivine. Figure 13.15 displays this idea.

1,180 °C: Highly vesiculated material sintered at this higher temperature has more cohesion of grains than the 1,150 °C material; however it is still weak compared to the Quarry 1 sample. This sample contains grains of the yet to be determined black mineral. Quarry 4 parent material has a large volume of olivine and it is assumed that the scattered black minerals found in the sintered material are the olivine. It does not seem to have been altered other than the color change. In both sintered samples there is little to no cohesion of minerals and both are highly vesiculated. The actual sintered

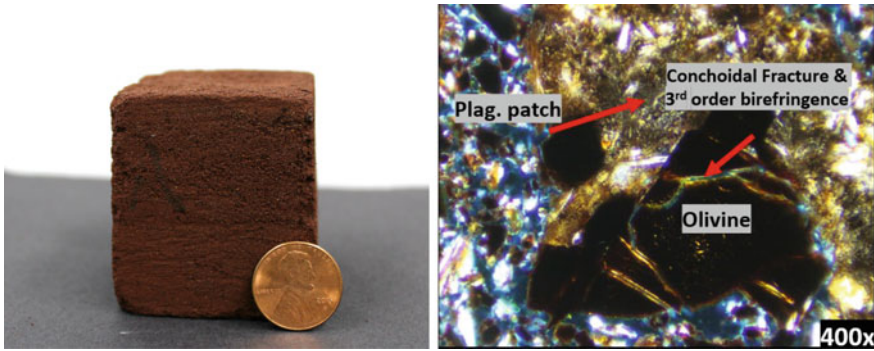


Fig. 13.15 Quarry 4 sintered material (left) and sintered thin section (right)

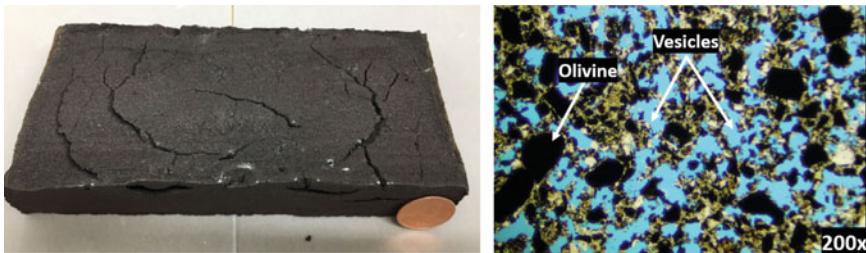


Fig. 13.16 Quarry 4 high-temperature sintered material (left) and thin section (right)

material is very weak structurally and breaks apart easily. Figure 13.16 displays the 1,180 °C sintered material and the highly vesiculated thin section.

13.5.3.3 Analog Site

1,150 °C: Highly vitrified material displaying rounded vesicles and scattered plagioclase occur throughout. The glassy appearance could be due to the parent rock being basalt scoria which contains a high percentage of volcanic glass, possibly giving the sample a vitrified look. Figure 13.17 displays the sintered material and thin section.

1,120 °C: Loosely sintered material almost indistinguishable from the parent rock thin section, displaying angular vesicles and a scattering of plagioclase. The hand sample shows a loose cohesion of particles which slough off easily from the mass of the material. Figure 13.18 shows the sintered material accompanied by its thin section.

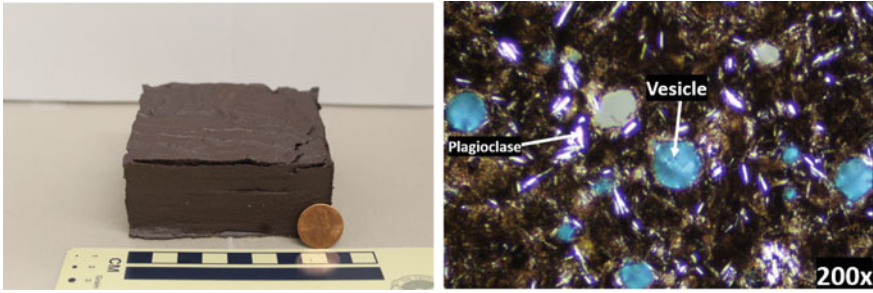


Fig. 13.17 1,150 °C Basalt scoria sintered material (left) and thin section (right)

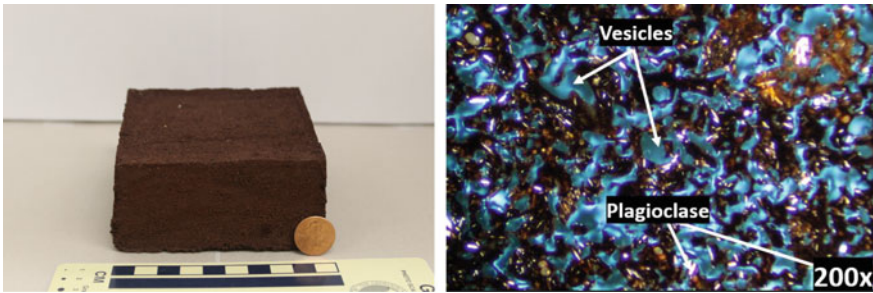


Fig. 13.18 1,120 °C basalt scoria sintered material (left) and thin section (right)

1,135 °C: Tightly sintered particles showing angular vesicles and slightly more vitrification than the 1,120 °C material. The hand sample displays better welding of particles than the 1,120 °C material and exhibits no sloughing of particles. Figure 13.19 displays the sintered material and its corresponding thin section.

Compared to the parent rock, the basalt scoria sintered at 1,150 °C is much less vesiculated and highly vitrified. Due to the vitrified nature of the 1,150 °C material, the temperature was lowered to 1,120 °C and 1,135 °C to determine if basalt scoria

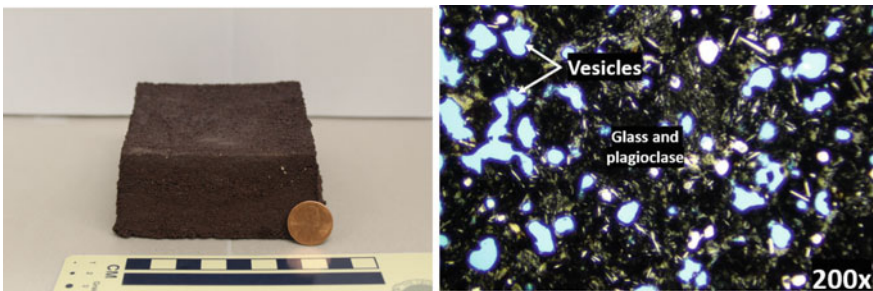


Fig. 13.19 1,135 °C basalt scoria sintered material (left) and thin section (right)

could be sintered at lower temperatures to create a similar material to Quarry 1. The material sintered at 1,120 °C had sufficient cohesion of particles and little to no vitrification, unlike the 1,150 °C material. The 1,135 °C material is much less friable than the 1,120 °C material. Both the 1,120 °C and 1,135 °C are reminiscent of the Quarry 1, 1,150 °C material, showing that basalt scoria could be sintered at lower temperatures and still produce an acceptable material. Sintering observations show that not all basalt feedstocks produced the same sintered material. This confirms that not all basalts are suitable for sintering to produce a durable material. There are several grades of materials from poor to exceptional sintering depending on chemical and mineral composition. Quarry 1 materials have consistently shown that they sinter to a durable material at both thermal profiles. Quarry 4 sinters a poor material at both thermal profiles, whereas the analog site material creates a unique and glassy material that can be sintered at lower temperatures.

13.6 Structural Properties of Sintered Basalt Versus Chemical Composition/Mineral Abundances

Our data suggest that a slight chemical variation can affect the physical properties of the end material. A possible reason may be linked to the mineral abundances in a feedstock. The most notable difference in the chemical make-up of the samples is the spike in MgO % for the sample from Quarry 4. The higher percentage of MgO may be caused by a high percentage of forsterite olivine in the feedstock. As the temperature increases during melting of a basalt, minerals sequentially dissolve, often leaving only olivine as the remaining solid phase. Thus, a portion of the original olivine (perhaps most of it) may not be sintering or bonding with surrounding particles, creating gaps between grains. These observations show that different basalts create different grades of materials, some being better suited for construction than others, although there may be other uses for lower grade materials. There is a defining balance of minerals that constitute the durability of a sintered material. From the data and observations, a higher percentage of plagioclase, moderate glass, and lower percentage of olivine creates the best materials. It was shown by observation that a higher percentage of glass brought the sintering temperature down but created a material that was glassy and not in the same grade category as the Quarry 1 material; although a good material it may not be suitable for launch pad or shielding applications. The Quarry 4 material which had a high percentage of olivine, and moderate amounts of plagioclase and amorphous glass, created a material that broke up on touch and failed significantly on flexural and compressive strength testing.

Chemical composition and mineral abundances play a key role in the final sintered product because basalt is made up of varying abundances of minerals. Basalt is a heterogeneous mixture made up of plagioclase and mafic minerals (most commonly

pyroxene and olivine) and some amount of amorphous volcanic glass. Each mineral has its own defining characteristics such as chemical make-up, density, hardness, color, melting point, etc. This fact must be taken into consideration when sintering basalt, as it defines the temperature profiles at which to fire, and the overall durability of the final product.

13.7 Conclusions

Much consideration has been given by many in the space industry to sintering as a suitable process for ISRU construction, on the assumption that regolith both on the lunar and Martian surfaces is plentiful and easy to collect. However, as the research work described in this chapter shows, even slight variations in the chemical and/or mineralogical composition of the regolith can have significant effects on the quality of the sintered material produced. For this reason, mineralogical and a chemical composition analysis will be required for any potential site that may be considered for a permanent settlement for which sintering is to be considered for ISRU construction or manufacturing. Because of time constraints on the XRD machine which was on loan, testing of the sintered material was not conducted but will need to be done to identify any changes in the mineral composition, and to identify the mysterious black mineral thought to be olivine.

Moving forward, more consideration of sintering technologies and additive manufacturing will also need to be considered. As this research has shown, there are many factors that must be considered when sintering basalts, specifically chemical and mineral compositions. This means there will need to be a way to locate the appropriate composition of basaltic regolith to truly create an optimal material. Although this may prove difficult, if other sintering technologies or additive manufacturing technologies are explored further, the problem of composition/ sintering temperatures may be overcome.

Acknowledgements The authors wish to thank the State of Hawai'i and the Department of Business, Economic Development of Tourism (DBEDT) for their continued support of PISCES.

References

- Allen, C.C., J.C. Graf, D.S. McKay. 1994. Sintering bricks on the Moon. In *Engineering, construction and operations in space (Space'94)*, ed. R.G. Galloway, and S. Lokaj, 1220–1229.
- Boynton, W.V., G.J. Taylor, S. Karunatillake, R.C. Reedy, and J.M. Keller. 2008. Elemental abundances determined by Mars Odyssey GRS.
- Brückner, J., G. Dreibus, S.W. Gellert, W.W. Squyres, H. Wänke, A. Yen, and J. Zipfel. 2008. Mars Exploration rovers: Chemical composition by the APXS. In *The Martian surface: Composition, mineralogy, and physical properties*, ed. J.F. Bell, 59–101. Cambridge University Press.

- Fischer, H.R. 2018. In-situ resource utilization—feasibility of the use of lunar soil to create structures on the Moon via sintering based additive manufacturing technology. *Aeronautics and Aerospace Open Access Journal* 2: 243–248.
- German, R. 2014. *Sintering: From empirical observations to scientific principles*. Butterworth-Heinemann.
- Heiken, G. H., D.T. Vaniman, and B.M. French. 1991. *Lunar Sourcebook, a user's guide to the Moon*.
- Indyk, S.J., and H. Benaroya. 2017. A structural assessment of unrefined sintered lunar regolith simulant. *Acta Astronautica* 140: 517–536.
- Kornuta, D.A. 2018. *Commercial Lunar propellant architecture: A collaborative study of lunar propellant production*, 171. United Launch Alliance
- Liu, Y., Y. Guan, Y. Zhang, G.R. Rossman, J.M. Eiler, and L.A. Taylor. 2012. Direct measurement of hydroxyl in the lunar regolith and the origin of lunar surface water. *Nature Geosci.* 5: 779–782.
- Lucey, P., R.L. Korotev, J.J. Gillis, L.A. Taylor, D. Lawrence, B.A. Campbell, ... and S. Maurice. 2006. Understanding the lunar surface and space-Moon interactions. *Reviews in Mineralogy and Geochemistry* 60 (1): 83–219.
- McKay, D.S., G. Heiken, A. Basu, G. Blanford, S. Simon, R. Reedy, B.M. French, and J. Papike. 1991. The lunar regolith. In Chapter 7, *Lunar sourcebook*, ed. G. Heiken, D. Vaniman, and B. French, 285–356. Cambridge University Press.
- McSween, H.Y., Jr., G.J. Taylor, and M.B. Wyatt. 2009. Elemental composition of the Martian crust. *Science* 324: 736–749. <https://doi.org/10.1126/science.1165871>.
- Meek, T.T. 1987. A proposed model for the sintering of a dielectric in a microwave field. *Journal of Material Science Letters* 6: 638–640.
- Meurisse, A., Makaya, A., Willsch, C., & Sperl, M. (2018). Solar 3D printing of lunar regolith. *Acta Astronautica*, 152, 800–810.
- National Ready Mixed Concrete Association. 2003a. CIP 16-flexural strength of concrete. concrete in practice: What, why and how. <https://www.nrmca.org/wp-content/uploads/2020/04/16pr.pdf>.
- National Ready Mixed Concrete Association. 2003b. CIP 35-testing compressive strength of concrete. Concrete in practice: What, why and how. <https://www.nrmca.org/wp-content/uploads/2020/04/35pr.pdf>.
- O'Connell-Cooper, C.D., J.G. Spray, L.M. Thompson, R. Gellert, J.A. Berger, J.I. Boyd, E.D. Desouza, G.M. Perrett, M. Schmidt, and S.J. VanBommel. 2017. APXS-derived chemistry of the Bagnold dune sands: Comparisons with Gale Crater soils and the global Martian average. *Journal of Geophysical Research: Planets* 122: 2623–2643. <https://doi.org/10.1002/2017JE005268>.
- Oyedotun, T.D.T. 2018. X-ray fluorescence (XRF) in the investigation of the composition of earth materials: A review and an overview. *Geology, Ecology, and Landscapes* 2 (2): 148–154.
- Rampe, E.B., et al. 2020. Mineralogy and geochemistry of sedimentary rocks and eolian sediments in Gale crater, Mars: A review after six Earth years of exploration with *Curiosity*. *Geochemistry* 80: 125605. <https://doi.org/10.1016/j.chemer.2020.125605>.
- Robinson, K.L., and G.J. Taylor. 2014. Heterogeneous distribution of water in the Moon. *Nature Geoscience* 7: 401–408. <https://doi.org/10.1038/NGEO2173>.
- Song, L., J. Xu, H. Tang, Jiquan Liu, Jianzhong Liu, X. Li, and S. Fan. 2020 Vacuum sintering behavior and magnetic transformation for high-Ti type basalt simulated lunar regolith. *Icarus* 347. <https://doi.org/10.1016/j.icarus.2020.113810>.
- Stopar, J.D., B.L. Jolliff, E.J. Speyerer, and E.I. Asphaug. 2018. Potential impact-induced water-solid reactions on the Moon. *Planetary and Space Science* 162: 157–169. <https://doi.org/10.1016/j.pss.2017.05.010>.
- Taylor, L. A., and T.T. Meek. 2005. Microwave sintering of lunar soil: Properties, theory, and practice. *Journal of Aerospace Engineering*, July 2005: 188–196.
- Taylor, G.J., L.M.V. Martel, P.G. Lucey, J.J. Gillis-Davis, D.F. Blake, and P. Sarrazin. 2019. Modal analyses of lunar soils by quantitative x-ray diffraction analysis. *Geochimica Et Cosmochimica Acta*. <https://doi.org/10.1016/j.gca.2019.07.046>.

- Tietz, O., and J. Büchner. 2018. The origin of the term 'basalt.' *Journal of Geosciences* 63 (4): 295–298.
- Walker, D., J. Longhi, and J.F. Hays. 1972. Experimental petrology and origin of fra Mauro rocks and soil. In: *Proceedings of the Third Lunar Science Conference*, 797–817.
- Winter, J.D. 2013. *Principles of igneous and metamorphic petrology*. Pearson Education.
- Zocca, A., M. Fateri, D. Al-Sabbagh, and J. Günster. 2020. Investigation of the sintering and melting of JSC-2A lunar regolith simulant. *Ceramics International* 46: 1409714104. <https://doi.org/10.1016/j.ceramint.2020.02.212>.

Chapter 14

Rocket Mining for Lunar and Mars ISRU



Matthew Kuhns, Roger J. Kuhns, Forrest Meyan, Philip Metzger, Hunter Williams, and Kris Zacny

Abstract The recovery of water ice from icy regolith in lunar permanently shadowed regions is of fundamental importance in developing lunar and cis-lunar economies for establishment of outposts and conducting exploration. Masten Space Systems has innovated *rocketM*, a system utilizing small rocket engines to excavate icy regolith using economically feasible amounts of energy in a durable, low-maintenance rover-mounted mining and processing system with a landed mass of 818 kg capable of operating in the lunar environment. Using water ice grades in regolith of 5–10%, the *rocketM* system is modeled to produce in excess of 426,000 kg of water ice per year. The system overcomes many of the maintenance and weight challenges of mining systems based on drilling or mechanical excavation, and the high energy use challenges of thermal mining systems.

M. Kuhns (✉)
Masten Space Systems, 1570 Sabovich St., Mojave, CA 93501, USA
e-mail: mmkuhns@gmail.com

R. J. Kuhns
7 Bradley Lane, Mystic, CT 06355, USA
e-mail: rogerjameskuhns@gmail.com

F. Meyan
Lunar Outpost Inc, 17700 S Golden Rd, Unit 102, Golden, CO 80401, USA
e-mail: forrest@lunaroutpost.com

P. Metzger
University of Central Florida, Partnership I, Research Pkwy, Orlando, FL 32826, USA
e-mail: philip.metzger@ucf.edu

H. Williams · K. Zacny
Honeybee Robotics, 2408 Lincoln Ave, Altadena, CA 91001, USA
e-mail: hjwilliams@honeybeerobotics.com

K. Zacny
e-mail: kazacny@honeybeerobotics.com

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_14

14.1 Introduction

Masten Space Systems (Masten), Honeybee Robotics (Honeybee), and Lunar Outpost have partnered to develop Resource Ore Concentrator using Kinetic Energy Targeted Mining (*rocketM*), a new method of extraction and harvesting lunar water in support of NASA's solar system exploration goals. The *rocketM* system adapts the deep cratering phenomenon, a risk for landing large rockets on unprepared surfaces, into a benefit via a new and not previously imagined space use. It provides an affordable solution at tens of millions of dollars versus billions of dollars to the costly In-Situ Resource Utilization (ISRU) endeavors on the Moon and Mars to support a permanent presence in space and provide the foundation of a thriving cislunar economy.

14.1.1 Concept of Operations

The rover and mining-beneficiation hardware are delivered to the lunar South Pole using a Masten Lunar Lander or comparable lander. The rover is able to make up to four 8-km round trips daily (24-h Earth day) from the landing site to an excavation site (Figs. 14.1 and 14.2). This site characterization is based on a model developed by NASA for their 2021 Break The Ice Challenge Water may be delivered to customers as ice or liquid, depending on their needs and the power available for the melt heating elements.

After landing, the delivery vehicle and the rover perform health and system checks to verify a safe landing. Then, within two hours of landing, the lander deploys ramps and the rover descends to the surface. The 2-h limit is for thermal and power considerations in case the lander lands at an off-nominal angle or location; this provides proper time to accommodate operations. Control of the rover is provided by remote teleoperation and through automation capabilities provided by Lunar Outpost. The *rocketM* system then performs another short health check on the surface, and begins

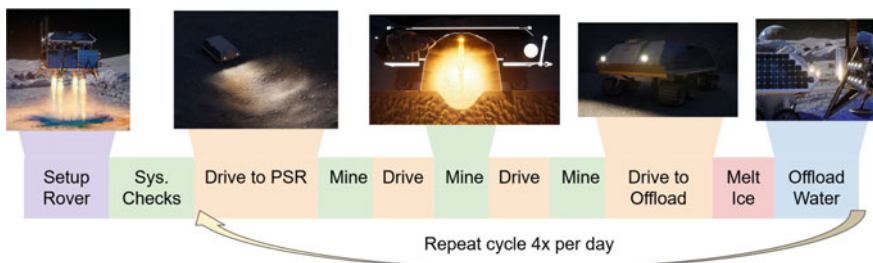


Fig. 14.1 *rocketM* concept of operations and daily process cycle

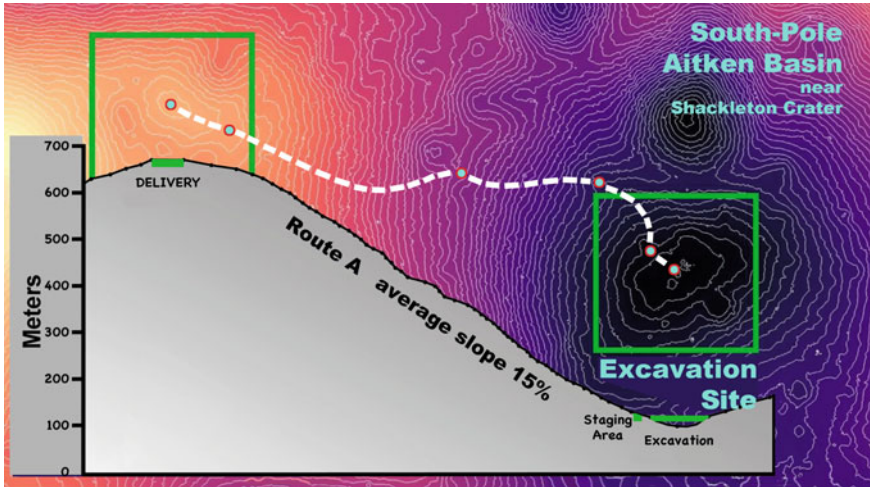


Fig. 14.2 Reference mission map and slope for ice mining from a lunar PSR

its first operation, a commissioning phase. Operations then proceed through the processes described above and illustrated in Fig. 14.1.

The operations are modelled on a NASA proposed delivery site located at an altitude above the excavation site, located within a PSR of a crater (Fig. 14.2). This identifies slope variations that the *rocketM* rover-mounted system will navigate.

14.1.2 Innovation and Impact

As the tempo of missions to the Moon and Mars increase there are many engineering challenges that must be overcome to ensure successful missions and an affordable long-term presence (NASA 2017). One key mission element is the duration of stay and activities on the lunar surface, with longer durations requiring more supplies and support. One approach for providing supplies is ISRU, where local water is harvested and used to sustain both the orbiting and surface lunar habitats and surface operations, including fuel depots, environmental systems, astro-agriculture, manufacturing, and other uses.

The Masten-led team has developed a novel rocket mining method called *rocketM* to efficiently, reliably, and—ultimately—economically extract lunar water ice and frozen volatiles locked in regolith at the Moon’s poles.

The *rocketM* system is an icy regolith excavation/cratering methodology. It draws upon deep cratering excavation using a small rocket thruster mounted inside a dome that encapsulates the exhaust gasses and fluidized icy regolith. This game changing technology can produce in excess of 426,000 kg water ice per year using a single rover-mounted mining and processing system with a landed mass of 818 kg, minimal

infrastructure requirements, and at economically competitive costs. Size and number of rover-mounted systems, as modules, can be scaled up to increase production as needed. Each module can be developed and deployed within the economic model to ensure economic viability is maintained. The system is designed for optimal lunar use including high reliability, the ability to bootstrap its power and propellant requirements to minimize landed mass, low cost, and its endurance in Permanently Shadowed Regions (PSRs).

The extracted ice can be electrolyzed into oxygen and hydrogen, turning it into rocket propellant. A small fraction can be set aside to continue powering the rocket engine for further excavation and mining. Based on projected ore reserves at the lunar South Pole, the system has an efficiency of 15 g of ice excavated for every 1 g of propellant used. This enables the system to bootstrap the initial activity with a very small quantity of propellant.

14.1.3 Current Technology Gaps and Solutions

Mining in-situ resources on the Moon requires innovative mining and recovery systems that are versatile and cost effective. Water ice is critical to expansion into space, especially as related to a cislunar economy, but several technology gaps currently exist as roadblocks. NASA has identified the following Key Technology Gaps related to the recovery of ice from lunar regolith: (1) excavation of large quantities of icy regolith; (2) delivery of large quantities of water; (3) hardware and equipment mass and efficiency; and (4) hardware and equipment reliability and durability in extreme lunar environmental conditions. The *rocketM* system addresses these Key Technology Gaps.

The development of a cislunar economy, if it is to be viable, must have economically feasible and profitable ventures to support the work of NASA and private space enterprises. The key *rocketM* breakthrough is the ability to excavate, transport, and sort large quantities of water ice bearing regolith for very low amounts of energy input by using a rocket engine plume as a drill to take advantage of a phenomenon called deep cratering, and do so under the environmental and geomorphic conditions present at the south lunar pole.

14.2 Hardware

All of the hardware necessary to achieve mission success of excavating, recovering, and delivering water ice to customers is contained in the *rocketM* system. The system has 4 major components: (1) Masten's rocket engine excavator/cratering plume and dome enclosure; (2) Honeybee Robotics' PlanetVac fluidized regolith pneumatic conveyance system; (3) the modified Aqua Factorem water ice beneficiation system; and (4) Lunar Outpost's rover (Fig. 14.3).

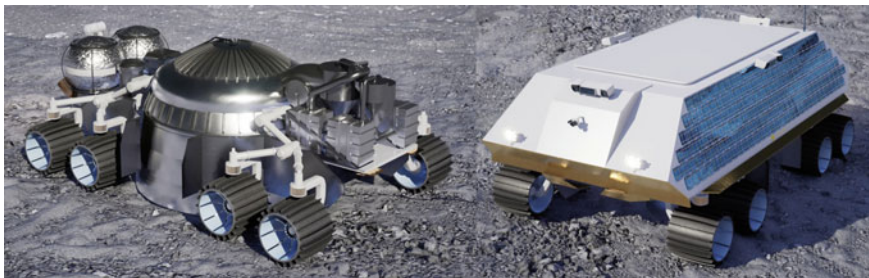


Fig. 14.3 The *rocketM* system and rover are shown here without protective coverings (left) and with coverings (right). The dome houses the rocket for crater mining. The PlanetVac and modified Aqua Factorem beneficiation system is mounted on the front (right side) of the rover shown at left

14.2.1 Hardware Feasibility

The feasibility of the rocket mining approach has been proven through tests of the method by Masten and others. Masten demonstrated deep cratering and particle transport during Phase I of its Deep Cratering and Plume Surface Effects SBIR in 2020 (Kuhns et al. 2021a) and successful ice excavation in 2021. That work characterized cratering rates based on plume and regolith conditions (see below). This enables the design of a rocket engine to achieve the desired cratering effects needed for mining while demonstrating that the core of the *rocketM* technology is sound. We are not the first to use rockets as drills. Work performed in Russia in the 1970 and 1980s demonstrated successful hard rock and soil drilling using a rocket engine plume. It achieved excavation rates of 100 m per minute in soil and permafrost and 5 m per minute through hard rock. Several companies worked on this technology including Mikhail Tsiferov and MBB (Underground Rocket 2021).

Not only does the *rocketM* system build on the proven feasibility of the rocket mining approach, but the integrated system incorporates a host of proven technologies (Fig. 14.4). This foundation of demonstrated capability at higher TRLs increases *rocketM*'s reliability and capability.

- The Masten's Resource Ore Concentrator using Kinetic Energy Targeted Mining (*rocketM*) utilizes its rocket engine as an excavator; the engine has been tested in several hundred firings in extreme thermal and regolith ejecta conditions. While the majority of water is sold to customers, a small portion will be electrolyzed into oxygen and hydrogen gas to feed back into the rocket mining system and power the excavation thruster. This enables the system to bootstrap itself with low launch mass, while still achieving high productivity on the Moon.
- The Honeybee Robotics' World Is Not Enough (W.I.N.E.) technology demonstrator matured the volatile transport technology and cold plate capture. It was tested in a lunar vacuum chamber where it extracted frozen regolith with a drill, captured the water vapor, and then used that to power a steam rocket (HoneybeeRobotics Ltd. 2019).



Fig. 14.4 Main hardware components of the *rocketM* rover-mounted system

- The Honeybee PlanetVac system is being used for particle transport, and it has been successfully demonstrated on flight missions.
- The modified Aqua Factorem system has successfully extracted ice from icy regolith simulant during bench tests at the University of Central Florida (UCF).
- Lunar Outpost’s Mobile Autonomous Prospecting Platform (MAPP) rover has successfully demonstrated terrestrial operation and is soon heading to the Moon. The Lunar Outpost Mars Oxygen In-Situ Resource Utilization Experiment (MOXIE) system is currently in operation on Mars on the Perseverance rover breaking new ground in ISRU capabilities.

14.2.2 Hardware Reliability and Durability

Reliability of the rocket engine when exposed to plume ejecta and repeated operational cycles is an engineering challenge. As a leader in reusable rocketry, Masten has some of the most-flown engines in the world. The 440 N thruster selected for this design has been fired hundreds of times as part of development tests studying new engine designs as well as plume surface interactions. The 440 N thruster has been repeatedly impacted by regolith ejecta and exhaust gasses while successfully completing rapid turnaround tests and up to a dozen Plume Surface Interaction (PSI) tests per day with no maintenance. The igniter is the critical hardware component on the thruster, and the item most prone to failure. To account for this, triple redundant igniters will be designed into the injector along with high reliability long-life spark generators. Based on operations planning, the engine will experience 159,000 cycles during a 5-year mission. The current record holder for operational cycles of a small

thruster has a cycle count of over 1 million starts in a vacuum test chamber, so the rocket mining requirement is achievable based on the current state of the art.

By using a rocket pulse to fluidize the regolith and break apart and separate ice particles, no mechanical excavation is required. This saves *rocketM* considerable amounts of equipment weight, reduces wear-and-tear of components, significantly reduces energy needs, minimizes external dust, offers an extremely low maintenance and high reliability system—many components of which have been or are in testing—and ensures a system that can operate in the extreme environmental conditions of the south lunar pole. Compared to the state of the art, the *rocketM* system has fewer failure modes with valves and beneficiation drums as the only moving parts, higher efficiency, long life, faster production speed, and low cost of entry. Vehicle platforms for the rocket mining device draw upon existing technologies matured by Lunar Outpost, which are Earth tested and robust enough to support mission lengths of 5 years or more. The rover design makes use of components and manufactures with significant flight heritage including use on NASA Mars rovers or deployed for multiple years in GEO. The only moving parts in the system are the propellant valves, dome lowering mechanism, and volatile control valves. This makes the system incredibly reliable with virtually no maintenance needs, and provides a system not subject to the mass, wear, and deterioration of mechanical excavators.

The interface between rocket mining and the icy regolith is the exposed surface area of the dome and engine. These surfaces will experience sandblast-style peening effects from the liberated regolith while also being exposed to the hot exhaust gasses. The dome pressure of <2 psi and short thruster firings of ½ s will minimize radiation and convective heat transfer to the hardware. The rocket injector will be kept clear of debris as it is in a zone of high pressure, which has been demonstrated through repeated testing by Masten. Debris does not backflow into the injector so long as proper engine conditions are maintained, including proper shut down purges. The material that is exposed to the ejecta shall be coated in abrasion resistant materials. For aluminum parts this will be a Type III hard anodization layer at least 0.012 mm thick. For non-aluminum parts this may be hard ceramic coatings such as alumina. Additionally, the additively manufactured components will be made from extremely hard metal matrix composites that Masten developed with Elementum3D especially for rocket use and demanding thermal environments.

14.2.3 Hardware Mass and Efficiency

The rocket pulse provides the pressure and mass flow to transport regolith and extract the ice, the exhaust gases are then condensed on cold plates before the tailings are deposited back on the Moon. This capability for a 1 kg thruster to do the work of an 80 kg excavator, based on system mass for moving similar volumes of regolith and mentioned here for comparison purposes because other proposed systems rely on mechanical excavators, makes the system extremely mass efficient. The captured propellants mean the system has a low loss rate and therefore can bootstrap itself

and operate nearly indefinitely on the surface so long as there is sunlight for battery recharging. This further reduces the landed mass and maximizes efficiency. The rate of extraction and refinement is also a game changing differentiator, with *rocketM* capable of extracting $87 \times$ the quantities per day of water compared to one other state-of-the-art system (Kornuta et al. 2018). Heat generated by the half-second rocket plume and the rover's electric motors is managed on-board and does not significantly phase change the ice. Any gaseous phases are captured on cold plates at the end of the beneficiation cycle.

14.3 Rocket Mining Excavation System

The rocket mining portion of the *rocketM* system architecture provides the methodology for excavation and capture of the water into the collection system (Fig. 14.5). The area of excavation is enclosed under a strong, impact resistant, deployable and transportable containment dome that is built into the rover. This dome enables deep cratering excavation to be contained, and allows mixing of regolith, and capture of released water ice particles. A small 440 N thrust oxygen/hydrogen rocket, mounted at the top of the dome, is used in place of a drill or mechanical excavator. This eliminates components that typically exhibit highest wear in a mining operation.

The hot plume of expanding gas acts to rapidly disaggregate, fluidize, and transport the icy regolith into the beneficiation system. This process has the advantage of

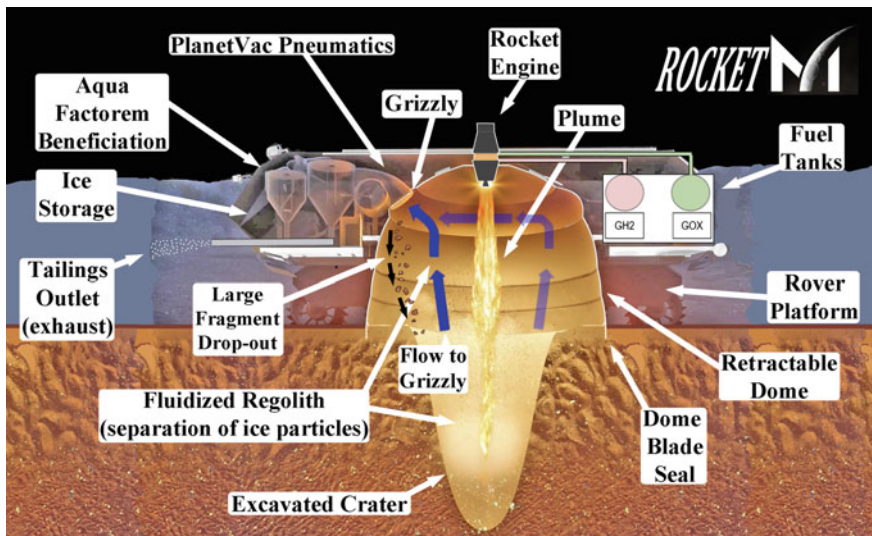


Fig. 14.5 Masten frozen regolith excavation testing using a small rocket

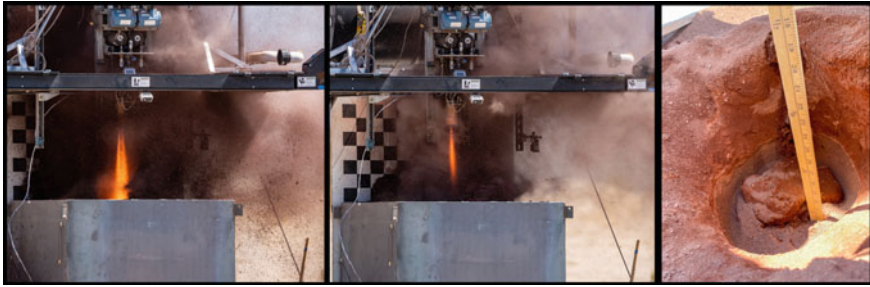


Fig. 14.6 Firing tests of Masten's small 440 N thrust oxygen/hydrogen rocket and the resulting crater in frozen icy regolith simulant. The rocket is mounted at the top of the dome in the *rocketM* system

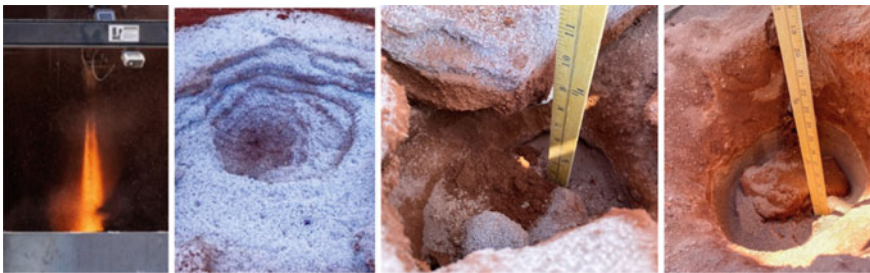


Fig. 14.7 Rocket mining tests showing a columnated 100 lbf plume (left); a shallow crater (center left); a pulse 1 deep crater (center right); and a pulse 2 large deep crater (right)

excavating ice from within and around rocks as it disaggregates regolith (Figs. 14.6 and 14.7).

Excavation rates for small rocket thrusters with gaseous propellants have been demonstrated at up to 70 kg per second during tests of frozen lunar regolith simulant at 77 K, which extracted 3.5 kg of water ice with two $\frac{1}{2}$ s engine pulses. This input of energy disrupted ice regolith, exposing frozen layers, pockets and areas of heterogeneous porosity, to direct chaotic mixing. Recent testing (June 2021) by Masten showed the effectiveness of deep cratering in disaggregating frozen regolith simulant to release ice particles.

14.3.1 Particle Breaking and Disaggregation

Particle breakage and disaggregation to liberate ice particles from rock and agglomerated fragments occur during the rocket pulse fluidization of regolith. The rocket plume disrupts the regolith, causing turbulent mixing, and incites displacement and rolling of larger fragments, such as boulder sized components (Figs. 14.8 and 14.9).



Fig. 14.8 Pulse 1 creates a deep crater and excavates some regolith (left), but much of the material is still contained in icy blocks (right) that require a second pulse to break up



Fig. 14.9 Pulse 2 is lower thrust, simulated here by increased engine height and far field plume effects (left) to break up the icy blocks and excavate more regolith. This leaves an enlarged crater (center and right) that grew 393% over the volume of the crater in Pulse 1. This demonstrates the effectiveness of chained pulses to break up and transport frozen regolith

Some percentage of particle grinding, fracturing and diminution occurs. This process will increase the separation of ice particles from their location as interstitial material, individual grains, and in fractures or fillings in regolith pore spaces.

Although very short (<1 s), the rocket plume blast introduces 215 J of thermal energy into the confinement dome and will result in some, although minor, sublimation of ice particles as the regolith is disaggregated. Preliminary estimates are that 5–10% of the particles will experience some degree of sublimation. Since the last component of the modified Aqua Factorem beneficiation system includes cold-plate condensation of exhausting gases, 50% of the sublimated ice will be captured, based on W.I.N.E. testing at Honeybee Robotics. There is also an unknown percentage of sublimated water vapor that will condense on other grains as it enters the beneficiation flow stream, and will likely be either captured (50%), if attached to ice particles, or lost if attached to lithic or mineral particles. The *rocketM* methodology combined with Aqua Factorem will eliminate 85% of the waste tailings and return them to the surface, reducing the transportation burden of the water ice and remaining regolith. This maximizes the efficiency of the system by collecting and transporting only water.

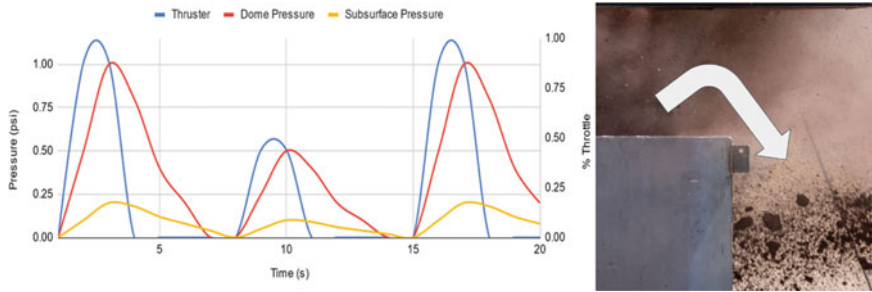


Fig. 14.10 Pressures in the rocket mining dome during operation pulses. (left) Large quantities of regolith were excavated and energetically transported away from the sample area. A 4 m/s wind was enough to drive all liberated regolith, even large chunks, to the right side of the test stand (white arrow). This demonstrates that small flow gradients can effectively direct the flow of particles. (right)

Based on the cratering response of the frozen regolith, a multi-pulse approach is used to generate small-size particulates to send through the beneficiation system (Fig. 14.10). Throttle points are alternated between 100 and 50% to modulate the shape of the plume and optimize breakup of the icy regolith. The grain relationships and ice occurrence sites are critical to achieving good recoveries of ice from icy regolith during the mining process. Because of the micro-sizes (<20 microns being common) of the ice, simple mechanical crushing may be insufficient to achieve high rates of grain separation, and therefore acceptable recoveries. Rocket mining induces sufficient energy to fluidize the regolith, thereby creating a mechanism to separate grains of all sizes.

14.3.2 Large Particle Rejection

The rejection of rocks and agglomerated particles >0.5 cm is accomplished through the use of a 0.5 cm grizzly screen (Fig. 14.5). The grizzly is placed in the dome at the input point of the PlanetVac intake pipe. Larger fragments fall back into the crater and do not have to be dealt with in the processing stream.

14.3.3 Excavation of Large Quantities of Icy Regolith

Deep cratering results when a firing rocket creates a columnated plume which can very efficiently excavate regolith. The technology for this rocket mining advancement is based on the physics of rocket mining deep cratering systematics, which can excavate tens of kilograms of material per second based on ground tests conducted at Masten. The system has demonstrated excavation to 1 m in depth in as little as 0.3 s

using the 440 N rocket engine, and to deeper depths with multiple half second pulses (Kuhns et al. 2020). A benefit of this design is the ability to excavate many types of regolith and handle large rocks, while simultaneously freeing water ice particles and coatings from boulders, pore spaces, fracture linings, breccias, and other hosting sites that would be more difficult with mechanical excavators. Excavation of icy regolith simulant frozen at 77 K using *rocketM* has been successfully demonstrated in testing at Masten.

14.3.4 *Delivery of Large Quantities of Water*

The *rocketM* system, as currently designed and through early testing, can recover over 426,000 kg of water or ice to commercial customers in one year of operation (355 days per year). This can be scaled up by adding *rocketM* rover systems (multiple simultaneous excavations) and through optimization of system performance. Recovered water ice is then delivered to NASA and commercial customers.

14.3.5 *Containment Dome*

The *rocketM* system is under an enclosure dome that contains the fluidized icy regolith. Dome dynamics and deployment are still in design. One consideration is to use the rover suspension to lower the chassis, which minimizes mechanisms exposed to the ejecta from mining. Another consideration is to use an extendable dome with sealed flexible interfaces. Power requirements and dome integrity testing is underway to refine the design elements of the dome system.

Excavation depth and width are controlled via throttling of the rocket engine. Pressure is regulated within the dome as that force influences the collimation of the plume and its excavating power. The lower throttle settings produce a shorter and wider plume, which is effective at removing overburden, breaking up chunks of fractured icy regolith, and collapsing the walls of deep craters to expand the excavation area. The higher throttle settings produce a more columnated and longer plume, with the columnation coming from the transient pressure in the dome. This plume is effective at performing the initial fracturing of icy regolith into chunks and creating a deep starting hole to reach the presumably richer deep ice deposits. The mechanism for crater formation is shown above in firing tests, and schematically in the Fig. 14.11.

This shows:

- (1) The columnating plume creating a jet that impinges on the soil. This kicks up the regolith and fluidizes it out of the crater at an angle of 30–60° for the initial formation.

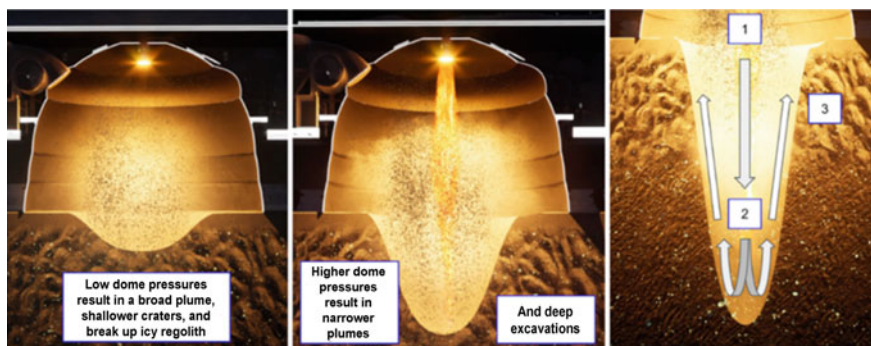


Fig. 14.11 Schematic representation of rocket plume excavation within the containment dome

- (2) As the crater is excavated deeper the ejecta angle increases to 80° , and the depth of the crater is then determined by plume pressures reaching equilibrium with the cohesive strength of the regolith.
- (3) The plume then turns 180° at the stagnation point and removes ejecta at a steep angle from the crater, with the regolith being transported through shear stresses along the crater wall.

There is very little lateral subsurface pressure exerted outward on the regolith which lowers the risk of dome leaks under the seal. This process will be enhanced by the nature of the icy regolith, which could further minimize pressure transfer in the subsurface. This erosion excavation method is very effective at breaking up and dislodging icy regolith.

As the containment dome is deployed a blade-like edge at the base of the dome seats it into the regolith to create a labyrinth seal (Fig. 14.12). This seal contains plume gases and icy regolith particles. Masten testing indicates the blade-like edge is pushed into the regolith to a depth of 6.5 cm will create a successful seal with margin for thruster pulses up to 1 s in duration. This is based on subsurface transient pressure measurements during PSI testing of the 440 N thruster for NASA, and shows the design is realistic and achievable to prevent pressure loss during firing (Kuhns et al. 2020). The method is assisted by the plume compacting the area around the crater as well as the icy regolith impeding gas flow through the subsurface. Icy regolith will enhance this effect.

To seat the dome vibrocore technology is used; this creates a short burst of vibrations in the dome material allowing the edge blade to penetrate surficial regolith and create the seal. This stabilizes the edge of the dome-regolith contact and the 0.8 m diameter crater under an interior dome pressure of 6.8 kPa. The vibrocore dome will be able to successfully seal itself in icy regolith, move small rocks out of its path, and either fracture larger rocks or reposition itself to find an optimal mining location.

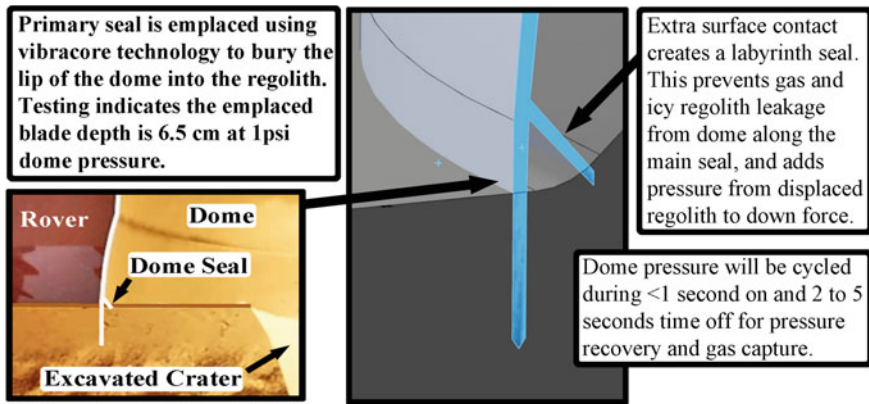


Fig. 14.12 Illustration of dome seal using a blade-like vibracore emplacement device

14.4 Honeybee Robotics PlanetVac Pneumatic Transport System

The PlanetVac system is a family of pneumatic samplers that has been successfully tested in multiple environments, including several test flights of Masten's Xodiac rocket, where it demonstrated the ability to capture rocket disrupted regolith (Fig. 14.11). The system uses compressed gas to transport large volumes of particles. It is currently manifested on multiple missions to the Moon (NASA CLPS 19D) and Mars (JAXA MMX).

As part of the *rocketM* system, PlanetVac captures regolith from the rocket mining dome and transports it to the modified Aqua Factorem beneficiation system. Honeybee Robotics testing has demonstrated successful pneumatic transport and sorting of fluidized regolith in a simulated lunar environment. The PlanetVac separates particles per the ratio of inertial force and aerodynamic drag. Aqua Factorem innovator Phil Metzger and Honeybee Robotics note that this method scales as density time particle diameter. Lithic and mineral particle densities are at least three times the density of ice, and therefore this process will separate ice from the other regolith material and help sort particles going into Aqua Factorem.

The system is driven through pneumatic flow and pressure differentials. The system has demonstrated capture of both dry and sticky particles, including ices. Flow from the thruster is 60 g/s, with the additional benefit of being a hot expanding gas, which provides additional energy over a cold working fluid. Testing by Sullivan showed 1 g of gas can lift 6000 g of regolith in lunar gravity environments (Thorpe and Sullivan 1972). Sixty (60) grams of thruster gas will dislodge 70 kg of regolith, and has the energy to transport 364 kg of material in a lunar environment. This results in a system margin of 520% giving the *rocketM* team high confidence in PlanetVac for this task.

14.4.1 PlanetVac Icy Regolith Interface

PlanetVac hardware that is exposed to regolith is primarily the flow surface in the piping and the grizzly filter to keep rocks from entering the beneficiation system (Fig. 14.13). The flow paths will be manufactured out of hard materials, such as metal matrix ceramic composites, and will be coated with Type III anodization layers or hard-wear resistant ceramics. The grizzly and other choke points will have reverse flow facing nozzles, which can be pulsed in the event of a clog. The system is designed to operate for 5 years with no on-site maintenance.

14.5 Ice Beneficiation System

The *rocketM* system is utilizing a modified Aqua Factorem beneficiation system designed at the University Central Florida, developed through the NASA NIAC Program (Metzger et al. 2021), with modifications by the Masten team. The system embraces tried-and-true technologies for sorting fine particles on the basis of their



Fig. 14.13 The PlanetVac pneumatic system schematic (upper left) and test conditions (upper right). The regolith simulant sample gathering results from tests with Masten’s rocket (lower left), and the laboratory testing equipment at Honeybee Robotics (lower right)

magnetism, electrostatic susceptibility, and density contrasts using pneumatic separation. Because of the use of pneumatic systematics, the system is not fluid- or gravity-dependent.

The system utilizes the rocket exhaust pressure created by the rocket plume as the crater is excavated in harmony with the PlanetVac pneumatic device described above (Figs. 14.5 and 14.13). This pressure, and its load of fluidized icy regolith, is directed through a grizzly to remove oversized particles and into the batch-processing Aqua Factorem beneficiator via the PlanetVac. The ice grains are intermixed with all the other minerals in the fluidized regolith, so a simple, ultra-low-energy grain-sorting process can extract the ice without phase change. The process offers potential economic viability because of the extreme mass reduction by producing an ice concentrate before transport to the NASA Delivery Site and significant reduction in energy requirements over thermal or mechanical approaches.

The Fig. 14.14 illustrates the modified Aqua Factorem beneficiation system. Fluidized regolith is passed over a magnetic separator in a pneumatic stream. Non-magnetic (ice) fraction proceeds through Cyclone 1 to remove rock and dense mineral particles. Material not removed is then passed over an electrostatic separator and non-conductive particles (ice) continue on to the final two cyclones for cleaning and upgrading. Ice is collected from Cyclone 3 and gases from sublimated ice and plume gases.

In the pneumatic stream, magnetic minerals are removed by passing the particles over an electromagnet and rotating drum. The magnetic mineral separates can be collected for further mineralogic/metallurgical study or disposed of via the tailings exhaust pipe. From studies drawing upon the bulk phenomenon of magnetic separation for lunar soil, it has been demonstrated that both magnetic minerals and glass particles containing nanophase iron (npFe) particles can be removed with this method (Oder 1992; Taylor and Oder 1990). This occurs because the soft saturation magnetization is temperature dependent for many minerals (Thorpe and Sullivan 1972). This condition is a benefit, in that it can extract the percentage of metallic/magnetic minerals, free metals and glass with npFe that are known to be in lunar soil and doing so with very little energy. These susceptible minerals and metals include magnetite, chromite, ilmenite, pyrrhotite, and meteoritic iron, but this collected fraction will likely be impure due the presence of npFe in glass. Due to the profound difference between ice and other regolith particles, magnetic separation can begin the process of concentrating the ice. As noted by Metzger in his Aqua Factorem program, the paramagnetic susceptibility of minerals is inversely related to temperature minus the Curie Constant (Biedermann et al. 2014). This means for certain minerals, such as olivine and pyroxenes, the magnetic susceptibility is 6 times stronger at 293 K in the PSR. This means magnetic separation is greatly improved under lunar conditions.

The modified Aqua Factorem material stream is then directed through the first of three pneumatic density sorting cyclones to continue removing non-ice grains, such as aggregated grains, rock fragments, and dust. The remaining concentrated material is then rapidly passed over an electrostatic separator to separate ice, a non-conductor, from non-ice. This is a tribocharging (surface phenomenon) process that draws upon the buildup of static electricity to separate particles (Agosto 1985, 1992). NASA has

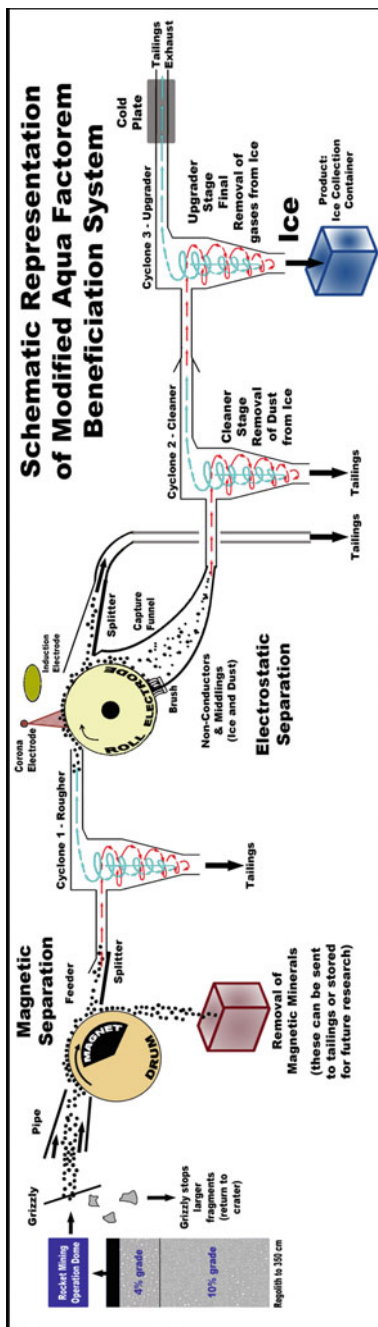


Fig. 14.14 The main components of the modified Aqua Factorem ice beneficiation system

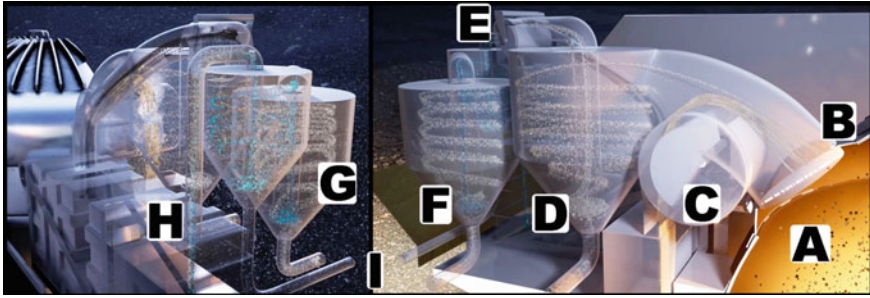


Fig. 14.15 Beneficiation system: **A** fluidized regolith, **B** grizzly filter to eliminate larger particles, **C** magnetic separator, **D** pneumatic rougher cyclone 1, **E** electrostatic separator, **F** pneumatic cleaner cyclone 2, **G** pneumatic upgrader cyclone 3, **H** cold plate unit, and **I** exhaust port

been successful in utilizing electrostatic beneficiation for ilmenite from lunar soil (Trigwell 2012). The separation of ice from silicate grains. As noted by Metzger (2021), Thorpe and Sullivan (1972), “Ice is known to tribocharge, and the properties of ice and silicate grains are so different that electrostatic separation will be highly efficient. Comparing tribocharging against aluminum for silicate versus ice of the same particle diameter, the ice will experience 100–10,000 times the acceleration in an electric field as the silicates” (Agosto 1985, 1992).

The remaining non-conductive fraction is passed through two additional pneumatic density sorting cyclones, the first is a cleaner to remove additional < 20 micron-sized mineral and glass dust from the water ice fraction (Fig. 14.15).

The second cyclone, called the upgrader, delivers the ice to a storage container, while allowing gas from sublimated ice and rocket plume fumes to pass over a cold-plate system to recover this volatile fraction from the *rocketM* process. Any additional ultra-fine tailings or uncollected gases are then exhausted through the tailings pipe via one-way valves to lunar ambient pressure.

Tailings ejection is achieved using the rocket-generated dome pressure which drives the icy regolith material through the beneficiation system. Should stoppage or accumulation enroute inhibit flow, the system can be flushed by briefly pulsing the rocket forcing exhaust gasses through the system. This type of gas pressure purging is effective in the mining industry on Earth. As the system architecture is tested in the next phase of development, Masten will identify beneficiation system flows and whether or not the system can benefit by in-line release valves and pressure boosters.

The ice can then be hauled to the chemical processing unit in solid phase and converted into rocket propellant.

The modified Aqua Factorem hardware exposed to regolith includes the flow surfaces and rotating drums. The flow paths will be manufactured out of hard materials, such as metal matrix ceramic composites, and will be coated with Type III anodization layers or hardware resistant ceramics. System choke points uses blower nozzles, which can be pulsed in the event of a clog to remove debris that may have gotten stuck.

14.6 Lunar Outpost Rover

The Lunar Outpost Heavy In-situ Propellant Production Off-world Rover (HIPPO) is Lunar Outpost's largest rover class designed for continuous operation in PSRs on the lunar surface. The platform draws extensively from technologies developed for the Lunar Outpost Mobile Autonomous Prospecting Platform (MAPP), currently undergoing protoflight qualification for lunar day missions as well as COLD-MAPP, a rover developed to survive the lunar night under a NASA-funded SBIR. The rover has been reconfigured to accommodate the *rocketM* system (Fig. 14.16).

The rover platform is 3 m long, 1 m wide, and 1 m tall. The total unloaded system mass (including ISRU payloads) is 818 kg unloaded and includes storage capacity for up to 300 kg of ice (Fig. 14.17). The fully loaded mass is 1118 kg and can scale slopes above 20° on the moon without exceeding the stall torque of the drivetrain. The Lunar Outpost HIPPO includes 8 independently actuated 0.3 m-diameter wheels. Each wheel is topology optimized for increased strength and powered by an in-hub motor for increased traction. A space-qualified motor in 36 V configuration paired with a space-qualified gearbox drives each wheel. The drivetrain is optimized to minimize drive time while being power efficient. The fully loaded rover over an average slope of 15° has a top speed of 3.31 km/h.

Regenerable hydrogen fuel cells are used to provide primary rover power during excavation sessions in total darkness. The fuel cells use the same oxygen and hydrogen tanks that power the *rocketM* system. This shared fuel reduces the typical mass cost of adding fuel cells to a rover. All water produced by the fuel cells is recovered and recycled into hydrogen and oxygen by the electrolysis system mounted on the rover. Current commercial-off-the-shelf proton-exchange membrane fuel cells for Earth-based applications can provide power densities of up to 5.5 kW/kg, ensuring a feasible density of 2.0 kW/kg capabilities from HIPPO's fuel cells (Hyzon Motors 2020). Fuel cells with a power density of 2.0 kW/kg also work well in tandem with the capacity provided by the 2000-Whr battery.

Near-vertical solar panels consisting of 198 cells per side provide up to 424.8 W to the rover when exposed to direct sunlight. This supplemental power serves two



Fig. 14.16 Rover in operation exhibiting tailings exhaust

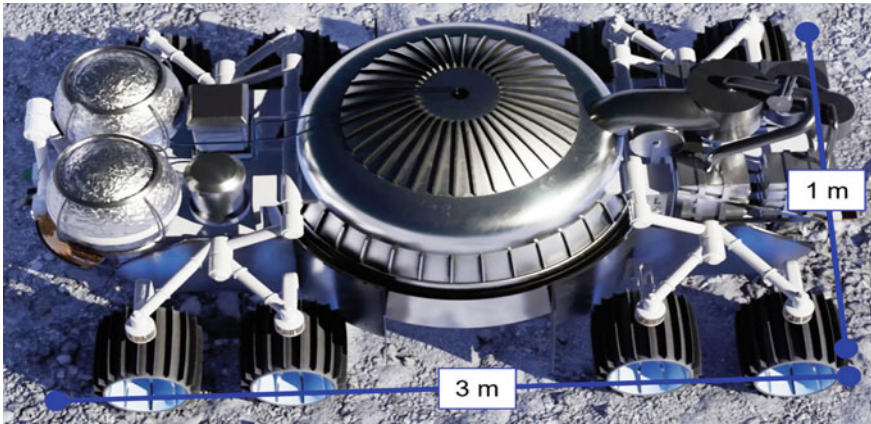


Fig. 14.17 Rover shown without outer protective coverings to illustrate geometry of the dome, the location of the modified Aqua Factorem beneficiation system (right), and fuel system (left)

purposes. First, the additional power is used to enhance mining operation speed, range, and yield during parts of the year where the delivery point is sunlit. Second, the solar panels increase operational flexibility for prospecting missions exploring for new resources. The solar cells are supported by a 2000-Whr battery system to store supplemental and backup power.

The rover has an advanced thermal control system that leverages Lunar Outpost's past experience designing robotic platforms for extreme environments. The rover also makes use of proprietary Lunar Outpost technologies developed to overcome the challenges that come with the extreme temperature differentials experienced on the lunar surface when transitioning from sunlit areas to PSRs. The thermal control system consists of the following components: the rover shell that acts as a radiator to reject heat and optical solar reflector to protect from solar heating, insulation between the shell and internals to keep the rover warm when in shadow, heaters to keep sensitive components within operating temperatures, and advanced active control mechanisms to increase the effective turndown ratio to allow for hot-cold transitions. The radiator is capable of rejecting in excess of 5 kW of heat. This will be used to manage the heat produced when the excavation system, PlanetVac, and modified Aqua Factorem are all operating and when the rover is sunlit. The active thermal control technologies are capable of reducing the heat loss to space from the radiator when the rover is operating in cold environments to less than 400 W.

The rover uses a 360° perception system for autonomous obstacle avoidance and navigation. Each side of the rover includes a pair of cameras for stereo vision and LiDAR for active depth perception inside the PSR. The front of the rover includes lights and additional perception sensors for scouting and prospecting. This perception system not only allows for autonomous navigation, but also characterization of the environments to provide data for future missions that will utilize the architecture enabled by *rocketM*.

Table 14.1 *rocketM* annual water production

Ice delivered per landed mass	520.8 kg _{H2O} /kg _{landed}
Ice delivered per energy used	75.8 kg _{H2O} /kWh
Mass of ice per crater (hole)	100 kg (85% recovery of 120 kg ice)
Craters per site	3 craters (holes) excavated per site
Sites per day	4 sites (3 craters × 4 sites = 12 craters)
Ice (kg) recovered per day	100 kg × 12 craters = 1,200 kg/day ice
Mass of ice recovered per year	355 × 1,200 kg/day = 426,000 kg

14.6.1 Lunar Mission Analysis

The *rocketM* system delivers 426,000 kg of water in 12 months (355 mining days and 10 maintenance and tech days) requiring 818 kg of landed mass using an 87% duty cycle. This makes the system extremely competitive both in terms of mass and cost, fulfilling the goal of designing a system capable of near-term deployment and also capable of generating commercially viable quantities of water to make it attractive to governments as well as private investors. At commercial prices it would cost \$21.3 billion to send this much water to cislunar space; the *rocketM* mission cost is \$500 million, a 42.6 × improvement (Table 14.1).

14.6.2 *rocketM* System Mass Performance Analysis

The *rocketM* system mass was built from the bottom up, using aerospace best practices and employing the mass growth allowance (MGA) method based on component maturity. Power and volume are treated in a similar manner, with values taken whenever possible from existing prototypes or hardware of the system components. Whenever practicable avionics and power system hardware designs with flight heritage will be reused from current flight designs from Masten, Lunar Outpost, and Honeybee.

Each of the three primary systems were built into Size, Weight, and Power (SWAP) tables (Table 14.2). These tables were then used to populate conops phases. The primary phases were flat driving, uphill drive, downhill driving, excavation, beneficiation, and unloading. These were further subdivided into sunlight or shade portions and built into a mission timeline. This was used to size the system and ensure that there was appropriate power and propellant margin, while achieving the target 818 kg mass. The following components have flight or extensive test heritage, and the *rocketM* team has high confidence in their SWAP values: the thruster assembly and valves (MM1), propellant COPV tanks (MM1), fuel cell (MOWS Tipping Point), avionics

(MM1 and MAPP), PlanetVac (CLPS 19D and MMX), Aqua Factorem (NIAC), cold plates (W.I.N.E.), radiators (MM1), beneficiation motors (Masten E-Pump), and wheel motors (MAPP).

The total landed system mass for the *rocketM* system is 818 kg. This is represented by 64 kg for the rocket mining system, 245.5 kg for the Aqua Factorem system, and 508.5 kg for the Lunar Outpost rover.

A bottom-up approach was used for the energy systems design and analysis for *rocketM*. Similar to the MGA, a Power Growth Allowance is used and is the same magnitude as the MGA based on item maturity levels. Each mining cycle, which returns 300 kg of water to the customer, requires 3.3 kWh of power for mining and transportation of ice. With 4 trips a day, and 355 days of mining operations per year (leaving 10 days for maintenance and troubleshooting), that is 1,420 mining cycles per year. Energy consumption for the year-long mission is 4.6 MWh on mining and ice transportation.

14.6.3 rocketM System Water Excavation Performance Analysis

An economic geology approach anchored by testing data was used for determination of water excavation performance of the *rocketM* system, enabling 1200 kg of water extraction per day. This results in 426,000 kg of water per year (Fig. 14.18).

14.6.4 Grade, Mass, Continuity, and Recovery of Ice “Ore” in Regolith

The in-situ resources outlined here includes an area 132,500 m² to an accessible depth of 3.5 m. Part of our evaluation of this resource is the knowledge that, in the science and business of exploration and mining, no resource is completely uniform in grade, extent, or geotechnical parameters, and rarely does it occur in a homogenous state or consistent geometry. Therefore, the model used herein draws upon the versatility of the *rocketM* system to operate in all known lunar regolith conditions. No mining method recovers 100% of the resource, so we have assumed an 86 percent recovery rate, based on our geologic, engineering, and geotechnical knowledge.

The recovery baseline for icy regolith (IR) and ice from each *rocketM* system crater is shown in Table 14.3. *Volume* is the estimated percent of ice in each of the different occurrences or situations within the icy regolith based on our knowledge of lunar regolith. *Released Ice* is the percent estimate the ice is released from the regolith as solid ice based on Masten rocket cratering tests. *Sublimation* is the estimated percentage changed due to thermal input from plume, and *Condensation* is

Table 14.2 Concept of operations data

#	Activity	Lunar day ops description	Elapsed time (h)	Duration (s)	Distance (m)	Power gain (kW)	Power use (kJ)	Heat gen <kW)	Fuel cell (Wh)	Battery (Wh)	GOX (*g)	GH2	Ice (kg)
1	Sunlit flat drive	Drive from lander to slope	0.11	412.4	430	175.2	-542.3	98.1		1898.03	6.60	1.10	9.0
2	Sunlit downhill drive	Drive downhill from lander into mining area	0.57	1654.3	1725	702.7	-895.0	572.9		1844.63	6.60	1.10	7.4
3	Shadow downhill drive	Drive downhill from lander into mining area	1.02	1606.3	1675	0.0	-869.0	460.8		1603.24	6.60	1.10	6.0
4	Shadow flat drive	Drive to Mining area	1.07	191.8	200	0.0	-252.2	34.2		1533.18	6.60	1.10	5.9
5	Mining	2 m deep excavation	1.08	29.4	0	0.0	-1.8	7.3		1532.68	5.07	0.84	108.9
6	Beneficiation	Sort ~2000 kg of regolith	1.25	600.0	0	0.0	-1244.4	256.2	333.33	1520.35	4.97	0.83	108.2
7	Shadow flat drive	Transit to next mining site	1.25	9.6	10	0.0	-12.6	1.7		1516.84	4.97	0.83	108.2
8	Mining	2 m deep excavation	1.26	29.4	0	0.0	-1.8	7.3		1516.34	3.44	0.58	211.2
9	Beneficiation	Sort ~2000 kg of regolith	1.43	600.0	0	0.0	-1244.4	256.2	333.33	1504.01	3.33	0.56	210.5

(continued)

Table 14.2 (continued)

#	Activity	Lunar day ops description	Elapsed time (h)	Duration (s)	Distance (m)	Power gain (kW)	Power use (kJ)	Heat gen <kW)	Fuel cell (Wh)	Battery (Wh)	GOX (*g)	GH2	Ice (kg)
10	Shadow flat drive	Transit to next mining site	1.43	9.6	10	0.0	-12.6	1.7		1500.51	3.33	0.56	210.5
11	Mining	2 m deep excavation	1.44	29.4	0	0.0	-1.8	7.3		1500.01	1.80	0.31	313.5
12	Beneficiation	Sort ~2000 kg of regolith	1.60	600.0	0	0.0	-1244.4	256.2	333.33	1487.68	1.70	0.29	312.7
13	Shadow flat drive	Return drive	1.66	191.8	200	0.0	-252.2	34.2		1417.62	1.70	0.29	312.6
14	Shadow uphill drive	Drive uphill from the mining area	2.08	1537.7	1675	0.0	-3212.2	441.1	854.25	1379.60	1.43	0.26	311.4
15	Sunlit uphill drive	Drive uphill from the mining area	2.52	1583.6	1725	672.7	-3308.0	454.3		647.56	1.43	0.26	310.1
16	Sunlit flat drive	Drive towards the processing plant	2.64	412.4	430	175.2	-542.3	98.1		545.59	1.43	0.26	309.8
17	Idle and offload	Unload Ice, Comms, housekeeping, software updates	2.78	500	0	100.0	-29.0	210		1997.42	6.86	2.98	9.0

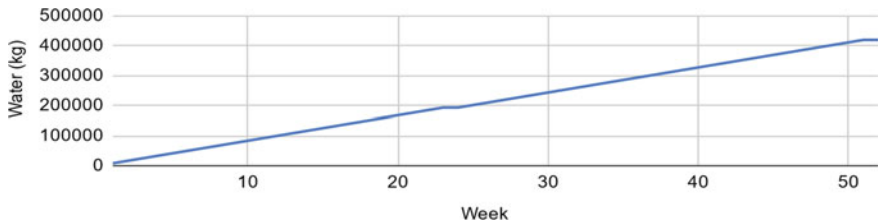


Fig. 14.18 Water delivered by week, with two planned off weeks for margin

the estimated percentage re-precipitated and recoverable. This is used to estimate ice recovery from icy regolith fluidized during the mining process (Table 14.4).

Table 14.3 Estimation of water ice recovery from IR using *rocketM*

Regolith ice occurrence	Volume est. (%)	Released ice (%)	Sublimation of ice by plume (%)	Condensation lost/recovered (%)	%Ice recovered (%)
Continuous matrix filling	15	90	10	50/50	85
Discontinuous bridging	5	90	5	50/50	88
Discrete grains	60	90	5	50/50	88
Fracture in-fill	2	50	5	50/50	48
Nodule in-fill	6	50	5	50/50	48
Coatings	10	90	5	50/50	88
Massive lenses	2	90	5	50/50	88
Weighted average of ice recovery					86

Table 14.4 Total ice deposit volumes and masses for the three zones

Zone	Depth (m)	Thickness (m)	Volume (m ³)	Density (g/cc) ^a	Mass (kg)
A	0–0.2	0.2	26,500 (4.6%)	1.47	38,955,000
B	0.2–1.0	0.8	106,000 (23.6%)	1.79	189,740,000
C	1.0–3.5	2.5	331,250 (76.4%)	1.89	612,812,500
Volume & mass of excav. site (A+B+C)			437,250 (100%)		841,507,500
Volume & mass of icy regolith (B+C)			331,250 (95.4%)		802,552,500
Volume & mass of contained ice (B+C)			37,365 ^b		68,870,850

^aNote g/cc have been converted to kg/m³ for mass in kg calculations. And, for all contained ice density calculations the maximum density of 1.89 g/cc is assumed

^bNote Since the exact density of cryogenic ice is not known, a one-to-one relationship between mass and volume is assumed here

Table 14.5 IR and ice recovered from each *rocketM* system crater

Zone	Ice grade (%)	Crater radius (m)	Crater vol. (m ³)	Regolith density (kg/m ³)	Mass ^a (kg)	Ice Mass ^d (kg)
A	0	0.4	0.1 ^b	1470	147 ^b	0
B (ore)	4	0.4	0.4	1790	716 ^c	29
C1–2 (ore)	10	0.4	0.5 ^c	1850	925 ^c	92
V & M B+C1–2		IR & ice per crater	0.9 ^c		1,641 ^c	184
C3–5 (ore)	10	0.4	0.75	1850	1,388	138

^aNote Mass (kg) total for each regolith zone

^bNote The Zone A Lunar soil is swept away prior to crater excavation

^cNote Represents icy regolith total mass in crater excavation

^dNote Assumes 100% of contained ice in icy regolith Zones B+C

The reference mission deposit dimensions encompass an area of 132,500 m² [500 m × 250 m], and a thickness of 3.5 m in three zones (A, B, C), two of which are ice-bearing (B & C), as shown in Table 14.4.

The weighted average ice grade of the B + C zones is 8.6%. This is important due to possible partial recovery of regolith, incomplete mixing during mining, and of course grade variability and recoverability in the zones; all common in real mining conditions (Table 14.5). The *rocketM* system rapidly excavates a crater with the dimension 0.8 m diameter and 2.0 m depth, modeled as cylinders for ice mining and recovery calculations. Although the ore zone extends down to beyond 3.5 m, the current *rocketM* system is designed to excavate most efficiently to 2 m depth. Because of the efficiency of the *rocketM* system, a considerable amount of water ice can be recovered by excavating to the 2 m depth in a short amount of time. The following table considers rocket mining to a depth of 2 m (Zone C1–2). Ice at 10% grade occurs to a depth of 3.5 m, with depth-grade increments of 46 kg ice per 0.5 m interval in the 10% grade zone. The table (and Fig. 14.19) also illustrates the ice resource for the lower 1.5 m (from 2.0 to 3.5 m depth, Zone C3–5) that may be accessible with modifications to *rocketM*.

14.6.5 Preliminary Excavation Plan of *rocketM*

The preliminary excavation plan for mining icy regolith with resources in the top 3.5 m is shown in Fig. 14.20, and the geometry of the excavations is shown in Fig. 14.9. A grid system allowing the excavation of meters deep craters delineates 51 mining blocks to accommodate mining for one year, encompassing an area approximately 90 × 60 m, or 5,400 m².

In this excavation model, each mining block is 10 m × 10 m and includes 25 crater excavations. Excavations are spaced 2 m apart, with each crater being 0.8 m in

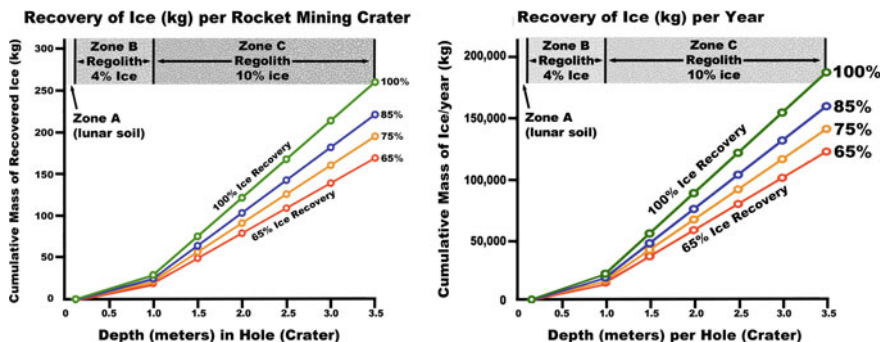


Fig. 14.19 With an understanding of grade, recovery percentages, and the *rocketM* system cratering/excavation, the recovered ice is shown at progressive depths down to 3.5 m at mining rates per day (left) and per year (right). The operating year is 355 days, allowing ten days for maintenance and other functions

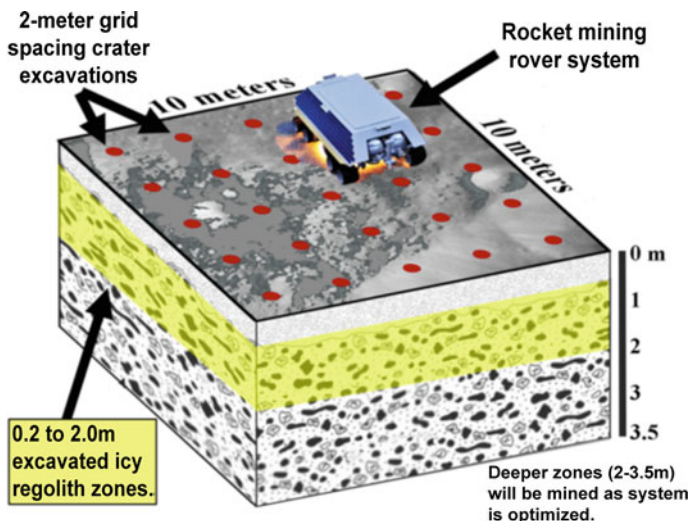


Fig. 14.20 Illustration of an idealized 10 × 10 m mining block. One year of mining would mine 51 of these blocks, recovering about 12% of the contained resource

diameter and at least 2.0 m deep (Fig. 14.11). This spacing allows movement of the *rocketM* rover system and ensures clearance of the rover over the excavations. Each mining block requires two days to mine, as each rover trip will mine three craters per site. *rocketM* will mine four sites per day totaling 12 excavations per day, recovering between 100 and 125 kg water ice per crater at an 85% recovery of resource. This excavates 4,260 craters per year, recovering 426,000 tonnes of water ice per year. The stripping ratio is 0.1.

This approach does not recover all the water ice resource in a given mining block but focusses on recovering enough water ice to achieve an economically viable mining venture. The area can be revisited to recover remaining reserves of water ice.

14.7 Preliminary Economics of *rocketM*

Developing an economically feasible and financially profitable cislunar economy depends on keeping front end capital costs low and operating costs manageable. In developing the *rocketM* system Masten has continually assessed economic components with the aim of reaching profitability as quickly as possible in the early stages of the program.

The general economic model for *rocketM* estimates \$500 million in development and deployment to the excavation site at the south lunar pole. The model also estimates an operating budget of \$10 million annually, and a contingency of \$5 million in the first several years of operation. With these highly generalized parameters we can examine the potential cash flow for operations (Table 14.6).

This economic model, as presented here, is highly simplified. Masten has modeled cashflow and profitability at different capital and operating costs, as well as at different water ice grades and tonnages in defining mineable (proven) ice reserves. The preliminary model shows the high profitability of the *rocketM* system.

Table 14.6 Generalized model of *rocketM* program economics

	Per excavation	Per day	Per year (350 operating days)
Excavations	1	12	4260
Recovered ice (kg)	120	1,440	511,200
Value at \$10k/kg	\$1,200,000	\$14,400,000	\$5,112,000,000
Value at \$8k/kg	\$960,000	\$11,520,000	\$4,089,600,000
Value at \$5k/kg	\$600,000	\$7,200,000	\$2,556,000,000
Project CAPEX (all up)^a			\$500,000,000
Annual operating cost			\$10,000,000
Annual contingency			\$5,000,000
Estimated EBIT Year 1 @ \$10k/kg water			\$4,597,000,000
Estimated EBIT Year 2 @ \$10k/kg water			\$5,097,000,000

^aNote in EBIT Year 1, the entire all up capital costs are included to achieve rapid payback. Note Project CAPEX (all up) includes all research and development costs, launch and deployment costs, and set-up to first operational mining day. Note that in EBIT Year 2, only the operating and contingency costs are subtracted from the cash flow. All cashflow calculations use \$10,000 per kg water ice; \$8 and \$5 k values are shown for illustration purposes. EBIT = earnings before interest and taxes; all dollars in US\$

At this point in time the greatest uncertainty exist in the water ice grades, tonnages, resource continuity, geometry, depths and stripping ratios, and geotechnical information regarding the regolith. Because of the very low certainty of resource information, the mining system must also be versatile enough to conduct exploration. This requires mobility and processing of sample results on the platform – functions capable of the *rocketM* system. Additionally, this translates into the development and carrying out of a focused exploration program before mining actually begins.

The *rocketM* system would allow such exploration, and then mining using the same platform could commence as soon as proven ice reserves are delineated. This is the fastest and most innovative exploration and mining methodology identified to date by Masten.

Masten is planning the construction of a pilot plant to test all components of the rocket mining and beneficiation system using regolith simulant. This rover-mounted system will allow testing of exploration and sampling methodologies, testing of grid mining and recovery, terrane management and navigation systems, especially around cratered mine areas, and product transport and transfer systems. The pilot plant and rover will allow testing of efficiencies, regolith flow within the beneficiation system, maintenance, and trouble shooting other components and processes of the rocket mining system.

References

- Agosto. 1985. Electrostatic concentration of lunar soil minerals. *Lunar bases and space activities of the 21st century*, 453.
- Agosto. 1992. Lunar beneficiation. *Space Resources* 509: 153–161.
- Biedermann, et al. 2014. Anisotropy of magnetic susceptibility in natural olivine single crystals. *Geochemistry, Geophysics, Geosystems* 15 (7): 3051–3065.
- HoneybeeRobotics Ltd. *W.I.N.E. The world is not enough*, 14 January 2019. <https://www.youtube.com/watch?v=-BhnL1GqhxM&t=100s>.
- Hyzon Motors. 2020. Third party tests confirm Hyzon Motors’ new liquid-cooled fuel cell stack leads the world in power density, 27 October 2020.
- Kornuta, et al. 2018. Commercial lunar propellant architecture. United Launch Alliance, 2018; Kornuta et al. (2019).
- Kuhns, Pelfrey, et al. 2019. Electric pump for launch and planetary landers. In *JANNAF liquid propulsion conference*, December 2019.
- Kuhns, et al. 2020. Testing deep cratering physics to inform plume effects modeling. Phase I Final Report NASA SBIR 80NSSC19C0603. February 2020. Masten Space Systems.
- Kuhns, et al. 2021a. Deep regolith cratering and plume effects modeling for lunar landing Sites. *ASCE Earth and Space*, April 2021.
- Kuhns, et al. 2021b. Deep regolith cratering and plume effects modeling for lunar landing sites. In *ASCE/earth & space conference*, 19–23 April 2021.
- Kuhns, et al. 2021c. Instant landing pads for lunar missions. *ASCE Earth and Space*, April 2021.
- Metzger, et al. 2021. Aqua factorem: Ultra low energy lunar water extraction: NASA NIAC Phase I Grant No. 80NSSC 20 K 1022, Final Report, March 15, 2021, 71 p.
- NASA. 2017. A new space policy directive calls for human expansion across solar system. NASA Press Release 17-097

- Oder. 1992. Beneficiation of lunar soils: Case studies in magnetics. *Mining, Metallurgy & Exploration* 9 (3): 119–130.
- Taylor, and Oder. 1990. Magnetic beneficiation of highland and hi-Ti mare soils-Rock, mineral, and glassy components. *Engineering, Construction and Operations* 1.
- Thorpe, and Sullivan. 1972. Temperature-dependent magnetic properties of individual glass spherules, Apollo 11, 12, and 14 lunar samples. In *Proceedings of third lunar science conference*, Supplement 3, *Geochimica et Cosmochimica Acta*, v. 3, 2465–2478.
- Trigwell, et al. 2012. Electrostatic beneficiation of lunar regolith: Applications in in situ resource utilization. *Journal of Aerospace Engineering* 26 (1): 30–36.
- “Underground Rocket,” Wikipedia, 16 January 2021. https://en.wikipedia.org/wiki/Underground_rocket.

Chapter 15

Penetration Investigations in Lunar Regolith and Simulants



Jared Atkinson

Abstract The American Apollo and Soviet Luna missions to the Moon during the ‘space race’ led to a vast collection of knowledge regarding the properties of the lunar surface. A critical but often under-appreciated investigative tool used in the missions is the penetrometer, a simple device which was successfully operated both manually and semi-autonomously to penetrate and characterize the unknown lunar regolith. Since that time, penetrometers have seen little use in investigations of returned lunar soil (also called regolith) or—more often—regolith simulants, though a few intrepid researchers have continued using the penetrometer in various forms. Recent work provides evidence that both the penetration and relaxation behavior of the regolith can help to determine useful physical properties, including important indications of ice content, cohesion, and particle angularity. Current plans to return to the Moon’s polar regions to explore icy regolith are being developed along with in-situ resource utilization (ISRU) demonstration missions, and some will likely include instruments for determining in-situ regolith properties using penetrometer technology.

15.1 Introduction

The American Apollo and Soviet Luna missions to the Moon during the space race led to a vast collection of knowledge regarding the properties of the lunar surface. A critical but often under-appreciated investigative tool used in the missions is the penetrometer, a simple device which was successfully operated both manually and semi-autonomously to penetrate and characterize the unknown lunar regolith. Since that time, penetrometers have seen little use in investigations of returned lunar soil (also called regolith) or—more often—regolith simulants, though a few intrepid researchers have continued using the penetrometer in various forms. Recent work provides evidence that both the penetration and relaxation behavior of the regolith can help to determine useful physical properties, including important indications of ice content, cohesion, and particle angularity.

J. Atkinson (✉)
Honeybee Robotics, 2408 Lincoln Ave., Altadena, CA 91001, USA
e-mail: jwatk@alum.mit.edu

Plans to return to the Moon's polar regions to explore icy regolith are currently in development some of which will include instruments for determining in-situ regolith properties using the technology and techniques discussed in this chapter. The discussion will start with a brief description of the penetrometer and the physical mechanisms involved in soil penetration and relaxation, followed by a brief history of lunar penetrometer investigations and the subsequently developed lunar simulants. The chapter will end with a review of the more pertinent and interesting research using penetrometers in regolith and their simulants since those first (and last) steps onto the lunar surface over half a century ago.

15.2 Penetrometer History, Measurements, and Applications

15.2.1 Introduction

A penetrometer can be thought of as any sort of rigid object—generally rod-like—that is pushed into a material to derive some sort of qualitative or quantitative measurement of its firmness, hardness, compaction, or strength. The earliest penetrometers were fists and thumbs, fingernails, sticks, and metal rods (Kirkham 2014), for millennia used to determine the consistency of a mixture, the strength of mud for building shelter, or the safety of the ground beneath an explorer's feet in a soggy wetland. The basic idea is this: the firmer or more solid the material, the more it *resists* penetration. As it turns out, the actual mechanics of penetration are more complicated than one would expect, and researchers have devoted entire careers to understanding the physics of how soils and other materials deform under penetration, and what that deformation can tell us about the material's fundamental properties.

Modern field penetrometers are generally metal rods (Sanglerat 1972) connected to a force-measuring device (electronic sensor, proving ring, etc.), pushed into a medium at a specified rate, that determine the resistance to vertical penetration with depth. The quantitative measurement of resistance is then correlated to soil characteristics (Kirkham 2014) such as bearing capacity, safe soil pressure, rolling resistance, wheel trafficability, relative density, crop yields, and a whole host of other—typically non-fundamental—properties. New research, however, aims at correlating the penetration resistance and subsequent relaxation to more fundamental soil properties (Oravec et al. 2010; Cil 2011; Atkinson et al. 2019, 2020).

Penetrometers used in the field exist in two main forms: portable hand-operated (Fig. 15.1), or machine-operated and stationary (Blok et al. 2019). Those used in laboratory testing have often been manually operated (and consequently prone to user error), while newer studies tend to focus on controlled mechanisms that limit lateral motion, maintain consistent penetration rates, and record penetration resistance continuously during operation (e.g., Atkinson et al. 2019). Two types of penetration tests also exist: static and dynamic (Kirkham 2014). Static tests consist of a

Fig. 15.1 A standard field penetrometer. *Source* Kirkham (2014)



penetrometer pushed steadily into the soil, such as the traditional cone penetration test (CPT) (Lunne et al. 1997), while dynamic tests involve a penetrometer driven into the soil by a hammer or falling weight resulting in a direct measurement of depth per blow rather than resistance as a function of depth.

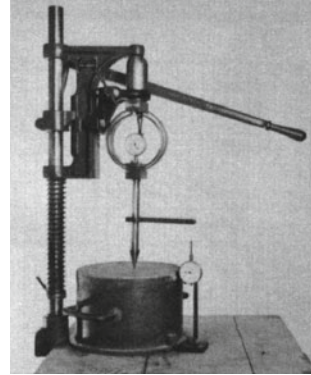
15.2.2 History

While humanity has used rod-like tools to probe the ground far before the first recorded instance, the method of measuring the strength of sub-surface soil using a rod has been attributed to French researchers (1846), who used a 1-mm diameter needle and 1-kg weight to probe clays of various consistencies and estimate the resulting cohesion (Sanglerat 1972). A comprehensive review of the penetrometer history is given by Sanglerat (1972) and an excellent overview by Lunne et al. (1997).

The invention of the modern cone penetrometer, arguably the most widely used device for field determinations of soil properties, is alternatively attributed to the US Army Corps of Engineers in the early 1940s (Oravec 2009; Kirkham 2014) and to the Dutch in the 1930s (Lunne et al. 1997). The Dutch cone penetrometer was developed in the Laboratory for Soil Mechanics at Delft University of Technology. It had a base area of 10 cm² and an apex angle of 60° (Durgunoglu and Mitchell 1973), and the first tests were conducted in 1932 (Lunne et al. 1997). The US version was developed at the Waterways Experiment Station during WWII and was composed of a 1.59-cm diameter rod, a proving ring with dial gage (for measuring force), a cone tip of 30°, and a 323-mm² base area (Fig. 15.2) (Oravec 2009, and citations within). Originally intended to predict the carrying capacity of fine-grained soils for off-road military vehicles, it provided a single value (Bekker 1969) that combined mechanical soil properties (such as soil drag and thrust) into one convenient parameter that could be interrelated with soil trafficability—a particularly important measurement in “go/no-go” analyses for military vehicles.

Concurrently, the electronic penetrometer—providing nearly continuous and sensitive penetration data—was developed in Berlin during WWII and has become a

Fig. 15.2 Army Corps of Engineers original laboratory cone penetrometer. Right: Typical field penetrometer.
Source Department of the Army Corps of Engineers Mississippi River Commission (1948)



common cone penetrometer for use in soil exploration (Lunne et al. 1997). It is now considered the standard modern field penetrometer, manually operated, providing automatic data acquisition and digital readouts of penetration resistance during use, and producing graphs of resistance as a function of depth.

In the laboratory, controlled-mechanism penetrometers have been developed for more sensitive testing of soils of various consistencies and volumes. Generally deployed on a laterally constrained z-stage to enable only vertical motion and equipped with force sensors capable of digital recording, they have been used to explore the mechanisms of deformation during penetration (Kochan et al. 1989; Cil 2011) and, when upwards vertical motion is prohibited via a lead screw, to examine the relaxation of the soil post penetration (Atkinson et al. 2019).

15.2.3 Measurements and Applications

Depending on the application, the standard measurements for cone penetration testing generally involve the vertical force imparted on the cone (often called the *resistance*)—measured in N or other units of force—and the depth of penetration in m or ft. Readouts show the force encountered at a certain depth, or alternatively the depth a penetrometer reaches under a specific weight (force). Additional complexity can be introduced through measurement of the friction along the penetrometer shaft (which contributes to the overall resistance) or the measurement of pore pressure using a tapered piezocone at the penetrometer tip (Lunne et al. 1997; Varney et al. 2001; Jiang et al. 2006).

The vertical force applied to press a cone to a certain depth in the soil is dependent on the cross-sectional area of the cone itself, so the force is often reported as a dimensionless cone index (CI) (Oravec 2009). CI represents the force per unit base area and generally takes the form (Rohani and Baladi 1981)

$$CI = \frac{4F_z}{\pi B^2}, \quad (15.1)$$

where F_z is the vertical force (in N) and B the cone diameter (in m).

The CI is an index of the resistance or *impedance* of the cone and is a compound parameter that involves components of shear, compressive, and tensile strength of the soil in addition to friction along the metal penetrometer shaft (Mulqueen et al. 1977). However, because it is a compound parameter it cannot be used to discern any individual property, and relatively little is known about how CI is affected by soil mechanical properties. CI does not provide an actual physical measurement of the soil strength, only an index to the penetration resistance (Oravec 2009). Even in homogeneous soils, variability in the soil condition will alter the proportion of shear, compressive, and tensile components determined in the CI (Mulqueen et al. 1977).

Over the years, researchers have discovered a number of correlations between CI and various soil characteristics. Rohani and Baladi (1981) developed relationships between CI and civil engineering properties such as shear strength, friction angle, cohesion, density, and shear modulus. While analytical predictions for the standard Waterways Experiment Station cone penetrometer showed good agreement between CI and these basic engineering properties, the relationships were only valid for homogeneous, frictional soils. Alshibli and Hasan (2009) claim that soil properties such as shear strength, permeability, in-situ stress, and compressibility can all be calculated using CPT data, and Carrier et al. (1991) point to the fact that the shear strength of soil, a key component of the resistance to penetration, governs engineering properties like ultimate bearing capacity, slope stability, and trafficability. In contrast, Wong (1989) showed that it was simple to obtain CI from a soil with known properties but difficult to determine the properties independently from the CI values. Mulqueen et al. (1977) investigated the relationship of CPT resistance to engineering properties such as soil strength and moisture content and found that changes in shear and compressive strengths were not reflected in the resulting CI values of soils with high moisture content: that is, the effect of the moisture content was predominant.

Another common index used in cone penetrometer investigations is the cone index gradient with depth (G), which is the slope of the linear portion of a resistance vs. depth curve. It has been shown to indicate relative soil density and strength over a range of depths, whereas CI indicates soil strength at a specific depth (Oravec 2009). As with CI, generally a higher G value indicates stronger soil.

Interpretation of CPT data still relies largely on empirical correlations developed in laboratories and calibration chambers, where soil properties are carefully controlled (Johnson 2003; Butlanska et al. 2012). When these correlative relationships are applied to soil conditions that differ from those of the testing environment, significant errors have been noted (Johnson 2003). Even with such complications, CPT results are routinely and successfully used in multiple industries to obtain valuable soil information.

An economical procedure, the cone penetrometer test is a common investigative tool in geotechnical engineering. The CPT has been widely used in soil studies related to off-road traffic and cultivation, and its use in offshore geotechnical work

is commonplace due to ease of deployment. The cone penetrometer is the reference tool for obtaining geotechnical data in burial engineering (often in conjunction with other continuous geophysical profiling techniques), and for assessing burial conditions along pipeline or telecom cable routes (Puech and Foray 2002). Pore pressure-predicting piezocones have also found uses in estimating the consolidation coefficient of soils (Jiang et al. 2006 and references therein) and in offshore geotechnical site investigations (Lunne et al. 1997; Varney et al. 2001). The CPT is even used as a rapid empirical method in the food industry to determine the consistency of a wide variety of solid, semisolid, and nonfood products (Muthukumarappan and Swamy 2017).

A short comparison of relevant parameters of terrestrial versus lunar (and lunar simulant) cone penetrometer testing is presented in Table 15.1 to orient geotechnical engineers to the similarities and differences between the two.

Table 15.1 Comparison of terrestrial and lunar geotechnical investigation methods

Character	Terrestrial	Lunar (in situ and analog)
Main purpose	Civil engineering projects	Scientific exploration
Soil type	Onshore/offshore clay, silt, sand, or gravel	Lunar regolith, lunar regolith simulants
Soil state	Partly or fully saturated	Dry
Penetrometer type	Standardized (10 cm ² , 60° cone)	Not standardized, mini-CPT often used in lab
Penetration mechanism	Generally truck/rig with hydraulic force	Manual or robotic
Penetration force available	Generally 5–20 kN	Generally 0.1–1 kN
Penetration depth	Generally 5–50 m ^a	Lab: mm In situ: m
Measurement bias	Electrical sensors inside penetrometer	Force gage above penetrometer
Measured parameters	Cone resistance, sleeve friction, pore pressure	Penetration resistance, relaxation
Measurement timing	During penetration, during dissipation (if conducted)	During penetration, during relaxation (if conducted)
Use of measured parameters	Site characterization for foundation design, slope stability, etc. Layering and soil parameters obtained through empirical correlations	Scientific understanding of regolith properties, estimates of bearing capacity

^aCan be carried out to larger depths in boreholes

15.3 Physical Mechanisms

15.3.1 Introduction

Penetration of a cone into a granular material—while a simple procedure—is a complicated process. The failure of grains around the cone leading to an increase in fine material, the contributions of stress at the cone tip and friction along the sleeve/shaft, the development of soil bodies ahead of the advancing cone tip: all make for a mechanically complex process which has been subjected to considerable theoretical treatment. To this day, however, there is no widely accepted theory of failure mechanics during penetration. Rather, empirical correlations dominate terrestrial use after decades of intense laboratory study in a wide range of natural and synthetic materials. This section provides a brief overview of research into the physical mechanics of penetration and, less closely studied, relaxation behavior.

15.3.2 Penetration

Researchers and engineers analyze cone penetration problems using three main methods: experimental investigations using calibration chambers and controlled environments, theoretical analyses concerning bearing capacity and/or cavity expansion, and numerical methods including finite- and discrete-element modeling (Jiang et al. 2006).

Theoretical treatments of the physical mechanisms of deformation at play during the penetration of a cone penetrometer into a granular material are generally based on continuum mechanics models of behavior and ignore the influence of microstructures (individual grains) (Johnson 2003). Most theories assume that shear strength is typically defined by Mohr–Coulomb

$$\tau_f = c - \sigma \tan \phi, \quad (15.2)$$

where c is apparent cohesion (in Pa), ϕ is the angle of internal friction or shearing resistance (in degrees), and σ is the normal pressure (in Pa), and incorporate some form of the ultimate bearing capacity equation introduced by Meyerhof (1957)

$$q_u = cN_{cq} + \frac{1}{2}\gamma BN_{\gamma q}. \quad (15.3)$$

Here q_u is known as the ultimate bearing capacity or penetration resistance (in N/m^2 or Pa), c is the soil cohesion (in Pa), γ the unit weight of the soil (in N/m^3), B the diameter of the penetrometer base and shaft (in m), and finally N_{cq} and $N_{\gamma q}$ are the dimensionless bearing capacity factors for cohesion surcharge and friction surcharge respectively. Another common theory assumes that penetration occurs

through a similar Mohr–Coulomb (elastic–plastic) granular medium that produces a monotonically increasing pressure loading that results in the expansion of a series of spherical cavities around the penetrometer (known as the cavity expansion theory), simulating the geometry of the cone (Vesic 1972; Rohani and Baladi 1981; Johnson 2003). Yu and Mitchell (1998) showed that the cavity expansion approach provides more accurate predictions than bearing capacity theory (Jiang et al. 2006).

While most of the most important recent theoretical treatments of penetration theory have involved the use of finite-element and discrete-element modeling (among others), numerical methods will not be discussed in detail in this chapter due to their complexity. Numerical models are instrumental in increasing our understanding of the physical mechanisms of deformation at the granular level, and the reader is directed to Jiang et al. (2017) and the references therein.

What physically occurs during penetration of a granular material is still an area of active research. Two approaches describe slightly different physical mechanisms, one based on continuum mechanics and the other on the interaction of microstructures at the granular level.

The continuum mechanics approach treats the penetrometer and the granular medium as single, separate bodies. Traditional theory, which predicts a linear increase in stress with depth for homogeneous, unstratified soils, states that during penetration the stresses near the penetrometer tip increase with depth to large peak stresses then decrease upon material failure to constants slightly larger than their initial values. The penetration causes the soil near the penetrometer tip to undergo combinations of compression, shear, and tensile stress in various directions and leads to a complex displacement path, often resulting in high displacement gradients and velocity fields (Jiang et al. 2006). Soil body formation at the leading edge of penetrometer tips (particularly blunt ones) have also been noted as having significant impact on the resistance (Mulqueen et al. 1977).

An approach that predicts nonlinear increases in resistance with depth was introduced by Puech and Foray (2002), refining a model for interpreting shallow penetration cone penetrometer testing in sands. Two phases of penetration were identified: the first phase was characterized by a parabolic increase in resistance associated with the dilational movement of the overburden around the rod, followed by a quasi-stationary linear regime dominated by compression. The first, parabolic phase tends to disappear in loose sands and the change in concavity occurs at the first occurrence of compressional mechanisms at the penetrometer tip.

Similar observations of nonlinearity were reported by Meyerhof (1976) and ElShafie (2012). ElShafie et al. (2012) presented a nonlinear model to describe penetration resistance force results in Martian regolith simulants, taking the form.

$$F_T = F_c + F_s = q_c A_c + q_s A_s, \quad (15.4)$$

where q_c is the cone resistance and q_s the sleeve/shaft resistance (in Pa), A_c and A_s the cone and sleeve area (in m^2).

When combined with estimates of q_c (Puech and Foray 2002) and q_s (Harr 1977), Atkinson et al. (2019) showed that ElShafie's estimate of F_T can be expressed as a parabolic equation in the form

$$F_T = F(z) = \alpha z + \beta z^2, \quad (15.5)$$

with α (in N/m) a function of the unit weight of the soil, bearing capacity factors, and cross-sectional area of the cone, and β (in N/m²) a complicated function of α and many soil properties including lateral slip lines, coefficients of lateral pressure and angle of internal friction. This model accurately predicted responses of various regolith simulants in a carefully conducted set of laboratory experiments (Atkinson et al. 2019, 2020). While nonlinearity has been identified as being applicable mainly to shallow penetrations of noncohesive, granular, sand-like materials (including regolith simulants), much remains to be discovered concerning the physical and mathematical descriptions of penetration within a continuum mechanics perspective.

The discrete-element approach suggests that granular materials support penetration forces through the development of microstructure elements that consist of individual grains/particles connected to each other by either cohesive bonds or friction contacts (Johnson 2003). During penetration, a microstructural element in contact with the penetrometer deforms elastically until a critical deflection is reached and the element fails in a brittle manner. Once failure occurs, the element fragments are compressed around the penetrometer surface forming a compaction zone extending from the cone tip to its base (Fig. 15.3). The microstructural approach attempts to address contradictions in the application of continuum mechanics theory, which predicts that resistance should not vary with cone angle and base area (Johnson 2003).

15.3.3 Relaxation

The relaxation of stresses around a penetrometer tip has been given insufficient treatment in the literature. Few experiments have been performed and very little has been investigated in terms of physical mechanisms specific to penetrometer testing. Stress relaxation phenomena in general have been successfully modeled using rheological models to aid in identifying the elastic and viscous components of deformation (Roylance 2001; Liingaard et al. 2004; Mitchell and Kenichi 2005; Atkinson et al. 2019).

Rheological models are conceptually useful and, while they reflect the real behavior of soils (Liingaard et al. 2004), they assume simple linear relationships in both the elastic and viscous components of deformation in describing the complex relationships in granular materials (Atkinson et al. 2019). The most common application of rheological models to relaxation behavior has been in the description of soil relaxation (Lacerda and Houston 1973; Rao et al. 1975; Kuhn 1987), but it has also found use in food science (Peleg and Normand 1983).

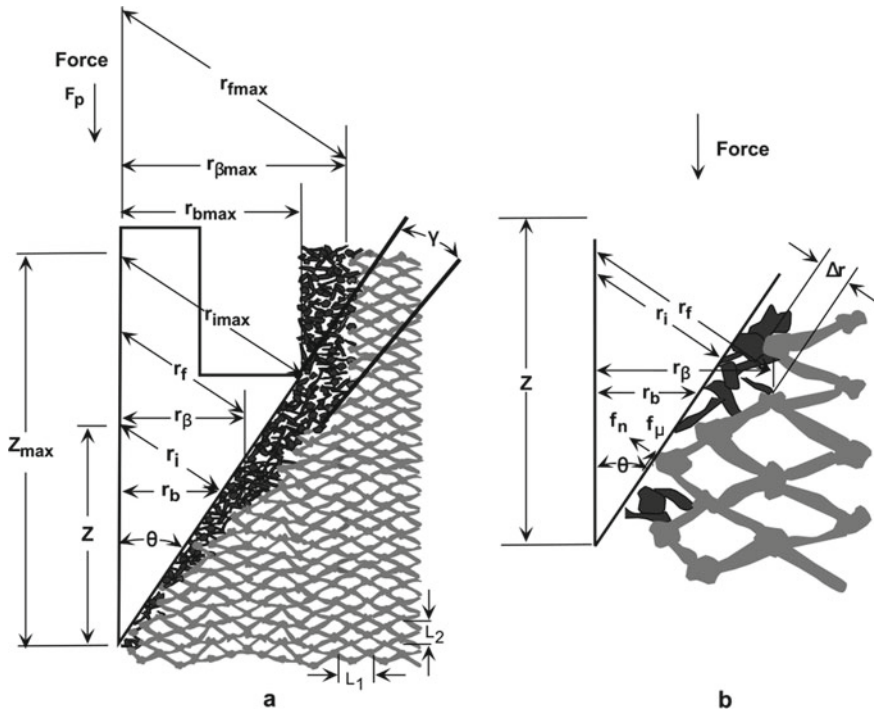


Fig. 15.3 A representation of the process of cone penetrometer moving through a granular material, including the geometric parameters and compaction zone. Note that the various parameters indicated are described in detail in the original publication and not described here. Modified from Johnson (2003)

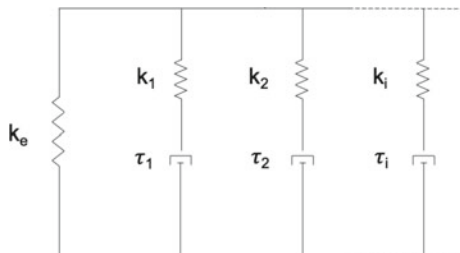
The most widely accepted form of the rheological model is the “Maxwell” model (Liingaard et al. 2004), consisting of an external Hookean spring connected in parallel to any number of Maxwell arms (themselves consisting of a spring and viscous Newtonian dashpot in series) (Fig. 15.4). The springs represent instantaneous elastic deformation of the body while the dashpots provide a viscous, time-dependent response to deformation (Atkinson et al. 2019). Upon deformation, all the input energy goes into compression of the springs and the dashpots then energy is gradually dissipated resulting in exponential decay (Rao et al. 1975).

A normalized mathematical formula for this relaxation behavior is presented by Peleg and Normand (1983) and modified by Atkinson et al. (2019) as

$$\frac{\sigma(t)}{\sigma_{max}} = \epsilon \left(k_e + \sum_{i=1}^n k_i e^{-\frac{t}{\tau_i}} \right), \tag{15.6}$$

which closely resembles the more general formula for universal relaxation provided by Snieder et al. (2017)

Fig. 15.4 Rheological relaxation model including springs and dashpots representing elastic and viscous behavior (respectively). *Source* Atkinson et al. (2019)



$$\frac{\sigma(t)}{\sigma_{max}} = \epsilon \left(k_e + \int_{\tau_{min}}^{\tau_{max}} \frac{1}{\tau} e^{-\frac{t}{\tau}} d\tau \right). \tag{15.7}$$

In these equations, $\sigma(t)$ is the vertical stress (in Pa) exerted on a probe tip as a function of time (t), σ_{max} the maximum penetration resistance experienced by the probe (in Pa), ϵ the resulting strain in the surrounding material, k_e the residual load supported by the material after relaxation has occurred, k_i and τ_i the elastic and viscous components of the Maxwell arms, and τ_{min} and τ_{max} the limiting relaxation times.

While the physical mechanisms of both penetration and relaxation in cone penetrometer testing are poorly understood, the information gathered during decades of use has yielded extremely useful results for investigating and predicting the behavior of soils. From buried cables, foundations for buildings, landing strips for airplanes through to the regolith on the surface of extraterrestrial bodies, the penetrometer is an excellent tool for characterizing and predicting soil behavior.

15.4 Lunar In-Situ Penetrometer Investigations

Despite repeated missions to Earth’s nearest celestial neighbor in the 1960s and 1970s (and, notably, none thereafter), the lunar environment, its soils, and the interplay between the two is not well understood. Direct in-situ measurements of the lunar regolith were made possible by the landings and subsequent exploration of the Luna 9 and 13 rovers in 1966, the Surveyor 7 surface sampler in 1968, two Lunokhod rovers in 1970 and 1973, and the manned Apollo missions from 1969 to 1972. Laboratory measurements of returned surface samples represent a very small fraction of the overall surface material (Oravec 2009).

The Luna 9 spacecraft was the first to survive a lunar landing, giving immediate information about the surface strength, while Luna 13 carried a conical indenter that used the impulse from a small solid-fuel jet engine to press into the lunar regolith (Cherkasov and Shvarev 1973). Initial regolith properties such as bearing capacity were investigated by the Surveyor 7 surface sampler through impact and trenching tests, though as further exploration would show, many of the inferred values of

regolith strength determined from these measurements were near the lower bounds (Carrier et al. 1991). The lunar regolith data from Surveyor missions were augmented by the later Apollo missions, whose measured soil property values are expected to be much closer to reality (Oravec 2009).

Autonomous Lunokhod (1 and 2) operations resulted in over 1000 measurements of the physical properties of the lunar regolith and covered over 50 km of terrain, representing the broadest coverage of lunar regolith strength available (Gromov 1998). An original analysis of these data was undertaken by Durgunoglu and Mitchell (1973), and more recent re-analysis by ElShafie and Chevrier (2014) confirmed many of the previous results. The astronauts of the Apollo missions performed extensive soil mechanics experiments that generally increased in complexity with each subsequent landing. From the interaction of the Apollo lunar module with the lunar surface, to famous footprints and specially designed penetrometer tests, the Apollo program provided the most detailed investigation of lunar regolith mechanics to date.

Penetrometer testing constituted a significant part of the overall soil property investigation of the lunar surface by both Soviet and American scientists. Penetrometers deployed autonomously on the Lunokhod rovers or operated manually by Apollo astronauts, along with additional geotechnical testing, provide the best estimates of lunar regolith bearing capacity, density, cohesion, friction angle, and void ratio. A broad overview of the geotechnical results of all lunar missions is presented in Table 15.2.

The most important measurements of the in-situ strength of lunar regolith come from the cone penetrometer tests made on the Lunokhod 1 and 2 robotic roving vehicles and manually operated tests on Apollo missions 14, 15, and 16 (Carrier et al. 1991).

The Lunokhod 1 robotic rover, deployed in 1970 on the Soviet Luna 17 mission, was equipped with a cone-vane penetrometer, a specialized device consisting of a combination conical penetrometer (5-cm² base area and 4.4-cm height, with a 60° apex angle) and shear-vane (7 cm in diameter, with four cone vanes at 90°) for measuring both penetration and torque resistance (Fig. 15.5). The device operated when the rover was stationary and deployed vertically into the soil to a maximum depth of 10 cm (and 196 N) and rotated while a set of sensors recorded the penetration depth, resistance force, rotation angle, and rotation force (torque) (Oravec 2009). In total, 327 tests were performed along a 5-km traverse near the Sea of Rains.

Typical Lunokhod 1 results, shown in Fig. 15.6, were analyzed (Mitchell et al. 1972) using the bearing capacity theory specifically developed for evaluating lunar penetrometer data (Durgunoglu and Mitchell 1973). While the surface locations of each penetration were not specifically described, a generalized horizontal section of the lunar surface (including crater slopes, rims, and rocky areas) was inferred, along with evidence that the strength of crater rims was generally higher than that of intercrater locations and that a decrease in the crater diameter resulted in a decrease in the strength of the regolith at the rim (Oravec 2009). The Lunokhod 2 rover traversed through a region of the Lemonnier crater for a distance exceeding 40 km in a transitional zone from the lunar mare (generally basaltic) to the highlands (generally composed of anorthosite), taking many additional penetrometer readings (Leonovich

Table 15.2 An overview of lunar soil/surface properties and the associated missions. Combined and modified from Gertsch et al. (2008) and Heiken et al. (1991). For complete references of the data, please see the original publications

Mission	Basis	Density [g/cm ³]	Cohesion [kPa]	Friction angle [°]
Orbiter	Boulder track analysis		0.35	33
Surveyor 1	TV and landing data		0.15–15	55
Surveyor 1	TV and landing data	1.5	0.13–0.4	30–40
Luna 13		0.8		
Surveyor 3	Soil mechanics surface sampler, TV and landing data			>35
Surveyor 6	Vernier engine		>0.07	35
Surveyor 6	Attitude jets		0.5–1.7	
Surveyor 3,7	Soil mechanics surface sampler		0.35–0.7	35–37
Lunar Orbiter	Boulder track analysis		0.1	10–30
Lunar Orbiter	Boulder track analysis	0.5	21–55	
Apollo 11	Footprints, lunar module landing data, crater slope stability		Consistent with Surveyor model	
Apollo 11	Penetration tests in LRL bulk soil sample		0.3–1.4	35–45
Apollo 11	Penetration of core tubes, flagpole, SWC shaft		0.8–2.1	37–45
Apollo 11		1.54–1.75		
Apollo 11		0.74–1.75		
Apollo 11		1.81–1.92		
Apollo 11		1.6–2.0		
Apollo 12	Footprints, lunar module landing data, crater slope stability		Consistent with Surveyor Model	

(continued)

Table 15.2 (continued)

Mission	Basis	Density [g/cm ³]	Cohesion [kPa]	Friction angle [°]
Apollo 12	Penetration of core tubes, flagpole, SWC shaft		0.6–0.8	38–44
Apollo 12	Cone penetrometer		0.17–2.7	25–45
Apollo 12		1.80–1.84		
Apollo 12		1.55–1.90		
Apollo 12		1.7–1.9		
Luna 16		1.2		
Lunokhod 1		1.5–1.7		
Apollo 14	Soil mechanics trench		<0.03–0.3	35–45
Apollo 14	Apollo simple penetrometer		≥ Surveyor model	
Apollo 14	MET tracks			37–47
Apollo 14		1.45–1.6		
Apollo 15		1.35–2.15		
Apollo 15	Measured at Station 8	1.92–2.01		47.5–51.5
Apollo 11, 12, 14, 15		1.76	0.55	43
Apollo 15	SRP data and soil mechanics trench		1.0	50
Apollo 16	SRP Station 4		0.6	46.5
Apollo 16	SRP Station 10		0.37	49.5
Apollo 16	SRP Station 10		0.25–0.6	47–50
Apollo 16	Drill core open hole		1.3	46.5
Apollo 17	Drill core open hole		1.1–1.8	30–50
Apollo 17	LRV		0.17	35
Apollo 17	North, East, and South Massifs		1	26–50

et al. 1971, 1976). The results of the Lunokhod measurements indicated a bearing capacity ranging from 0.2 to 1.0 kN/m², with a most probable of 0.34 kN/m², and a range of shear strengths from 0.03 to 0.09 kN/m², with a most probable value of ~0.048 kN/m² (Cherkasov and Shvarev 1973; Zacny et al. 2010).

Fig. 15.5 The Soviet Lunokhod 1 robotic roving vehicle. The rover landed in Mare Imbrium in 1970. *Source* Carrier et al. (1991)

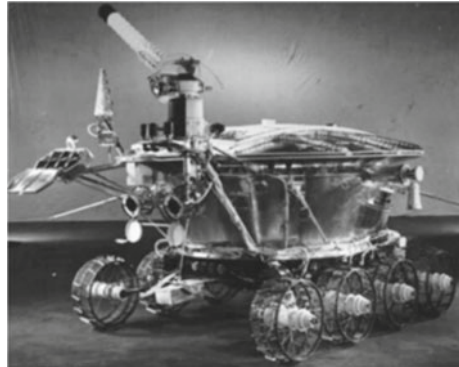
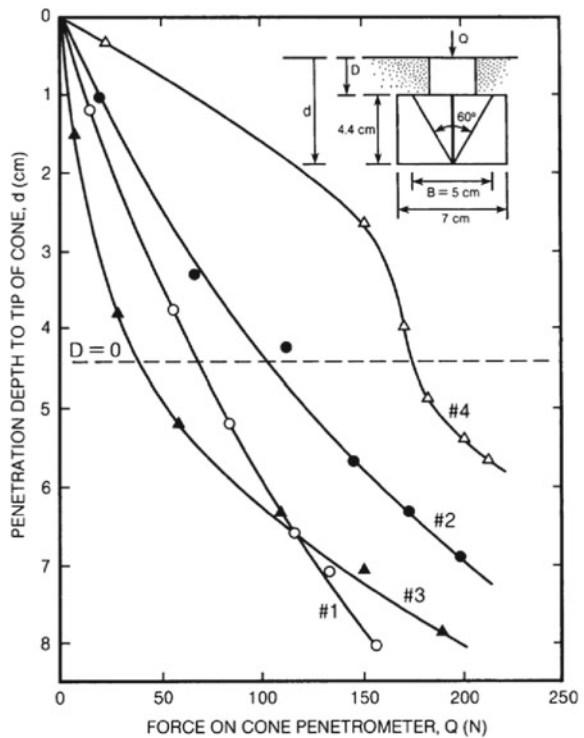


Fig. 15.6 Typical cone penetrometer resistance data, obtained by the Lunokhod 1 automated rover, for the lunar surface material in different areas of its landing site. *Source* Carrier et al. (1991)



The manned Apollo missions, beginning with Apollo 11 in 1969, ushered in a three-year period of intense study of the lunar surface including the properties of the regolith. Apollo 11 and 12, the first two American lunar missions, carried no specific lunar soil testing devices. Estimates of shear strength were limited to interactions with the lunar surface including the landing of the Lunar module, astronauts walking on the surface (the famous footprint), penetration into the soil by coring tubes, an equally

famous flag pole, and the solar wind composition shaft (Carrier et al. 1991). These various interactions suggested that the surface was at least as strong as predicted by the Surveyor estimates (Costes et al. 1969).

Apollo 14, landing in early 1971, deployed what became known as the Apollo Simple Penetrometer (ASP): a 0.95-cm diameter, 68-cm long, 30° cone penetrometer used to determine the difference in penetration resistance at various locations along the lunar surface (Carrier et al. 1991). Measuring resistance using the ASP was performed in a rather circumspect manner: Astronaut Mitchell (having had his one- and two-handed pressing force measured prior to the mission) operated the ASP by pressing it as far as he could into the surface using one hand, marking the depth of penetration, then using both hands to penetrate to a maximum depth thereby generating a rough resistance force vs. depth curve. These estimates of force were related to values of cohesion and internal friction angle, which were later compared to Surveyor data and found to be somewhat higher.

The Apollo 15 and 16 missions in 1971 and 1972 made use of an advanced cone penetrometer for measuring lunar regolith properties, known as the Self-recording Penetrometer (SRP) (Fig. 15.7), developed in the Geotechnical Research Lab at the Marshall Space Flight Center Space Science Lab. Designed with a detachable penetrometer portion, a rotating drum recording unit, and various probe components, the instrument provided a constant force-versus-depth profile. As the probe was pushed into the surface with a downward force, a gold-plated cylindrical drum rotated corresponding to the applied force and was simultaneously scratched by a stylus according to the depth of penetration (Carrier et al. 1991). The drum was scribed in situ and returned to Earth for analysis. The SRP included three interchangeable 30° cones of base areas 129, 133, and 645 mm², capable of a maximum load of 111 N and depth of 75 cm (Johnson et al. 1995).

Six cone penetrometer tests were performed during Apollo 15, all near the lunar module and all performed by Astronaut James Irwin. Costes et al. (1969) reported that two SRP measurements were made within and adjacent to a lunar roving vehicle track, and two others made adjacent to and at the bottom of a 30-cm deep trench with

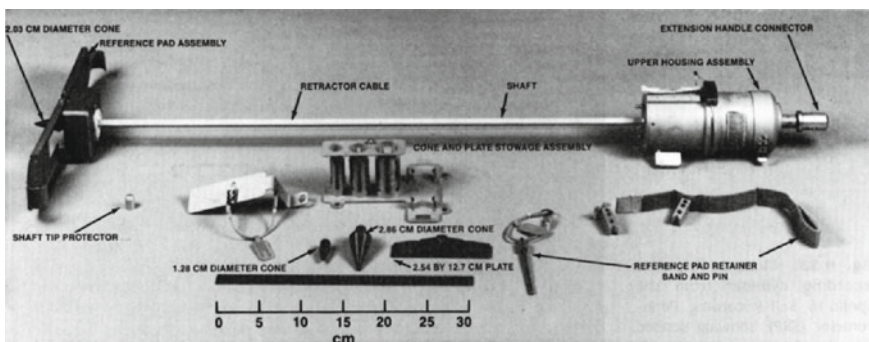


Fig. 15.7 Photo and explanatory diagram of the Self-Recording Penetrometer (SRP) used on the Apollo 15 and 16 missions. *Source* Carrier et al. (1991)

a vertical sidewall. During Apollo 16, ten measurements using the SRP were made by Astronaut Charlie Duke at Bench Crater and the ALSEP site.

The resulting soil mechanics data, in the form of handwritten plots of the penetrometer resistance stress as a function of the depth of penetration, mainly provide a lower bound to the soil strength as slippage of the surface reference pad made it difficult to accurately determine the depth of penetration (Oravec 2009). Estimates of cohesion and the internal friction angle of lunar regolith from the SRP are 0.25–1.0 Pa and 46.5–51.5°, respectively (Carrier et al. 1991).

Oravec (2009) used CI measurements from Apollo 15 and 16 SRP data to determine the cone index gradient G as a function of depth. G ranges from <3 to >9 kPa/mm were calculated, with little apparent correlation with depth (though it is noted that high G values may correspond to encountering rocks in the subsurface), as shown in Table 15.3.

Integrating all the data available, Carrier et al. (1991) derived and recommended “typical” intercrater values of cohesion and friction angle to use when modeling the behavior of the lunar surface (Table 15.4).

Table 15.3 Lunar cone index gradient terrain estimates near Apollo 15 and 16 landing sites. Modified from Oravec et al. (2010), reproduced by permission of Heather Oravec

Mission	Location	Estimated depth [cm]	Cone index gradient, G [kPa/mm]
Apollo 15	Adjacent to trench	8.25	4.06
Apollo 15	In rover track	5.25	4.36–7.59
Apollo 15	Adjacent to rover track	<11.25	>2.98
Apollo 16	Uphill, top of crater	20	3.37–3.86
Apollo 16	Near rover track	8 ^a	6.30–9.85

^aPenetrometer may have hit rock

Table 15.4 Recommended typical values of lunar soil cohesion and friction angle (intercrater areas). Modified from Carrier et al. (1991), reproduced by permission of the Lunar and Planetary Institute, Houston

Depth range [cm]	Cohesion, c [kPa]		Friction Angle, φ [°]	
	Average	Range	Average	Range
0–15	0.52	0.44–0.62	42	41–43
0–30	0.90	0.74–1.1	46	44–47
30–60	3.0	2.4–3.8	54	52–55
0–60	1.6	1.3–1.9	49	48–51

15.5 Lunar Simulants

15.5.1 History

In addition to providing us with important data about the properties of the lunar surface and near-subsurface through the investigations described in Sect. 15.4, the 12 Apollo astronauts also returned 382 kg of lunar material for study here on Earth between 1969 and 1972. To date, an estimated 350 kg of this original material remains for study (Sibille et al. 2006).

The success of future lunar operations (as well as those on other bodies) depends critically on the ability to predict and simulate lunar regolith behavior accurately. Tasks such as construction of lunar habitats, operating surface vehicles, lunar mining, and mitigating the hazard of excessive lunar dust all rely on a fundamental knowledge of regolith behavior. Due to the limited supply of real lunar material and the need to preserve it, the scientific community has turned to the manufacture of suitable regolith simulants intended to represent specific properties of the lunar surface.

A simulant is a material manufactured from natural or synthetic terrestrial components (including meteors) for simulating one or more physical and/or chemical properties of the lunar soil (Sibille et al. 2006). Due to the rather limited variation in regolith composition on the lunar surface, most terrestrial stimulants contain some basaltic or sandy-silicate materials, often ground to a grain-size distribution resembling that of the returned lunar material. The manufacturing of terrestrial simulants generally requires knowledge of the special properties needed for the intended exploration disciplines. For example, terrestrial simulants needed for resource-focused extraction disciplines require chemical and mineralogical similarity to the lunar regolith, while geotechnical researchers require large volumes of simulants with similar mechanical/physical behavior.

Simulants can only approximate the behavior of lunar soil. The unique lunar environment creates regolith properties that are not found in terrestrial soils. Lunar regolith is expected to be dramatically frictional and dilatant compared to terrestrial analogs, particularly at low confining pressures (caused by the absence of a significant lunar atmosphere), which can lead to nonlinear behavior and will strongly affect the behavior of engineered lunar structures such as foundations (Klosky et al. 2000). Lunar regolith also contains agglutinates, glass spheres, nanophase iron, and micrometeoroid impact craters on grain surfaces not found in terrestrial soils (Carrier et al. 1991). The extreme angularity, abrasiveness, and invasiveness of lunar regolith and its associated dust has been remarked upon by many, including the Apollo astronauts subjected to its extraordinary behavior on the lunar surface.

Simulants were also critical in predicting the behavior of lunar regolith *before* humans ever landed on the surface. Despite limited knowledge of the lunar surface prior to the first landings, a highly successful set of standard simulants were developed during the Apollo program to test surface systems in preparation for the lunar landings (Sibille et al. 2006). These simulants, known as Lunar Surface Simulants (LSS) 1–5, were used in the development of drills, tools, and lunar roving vehicle

Table 15.5 Soil properties and parameters for single-wheel tests in lunar simulants LSS1–3. Modified from Green and Melzer (1971)

Soil Parameter	Method	Average
Penetration resistance gradient [g/cm ³]	–	87.7
Dry density [g/cm ³]	Gradient <i>G</i>	1.59
	Gravimetric	1.67
Moisture content [%]	–	0.9
Relative density [%]	Gradient <i>G</i>	42
	Gravimetric	54
Average friction angle [°]	Triaxial (secant)	39
	In-situ plate	35
	Sheargraph	29
	Bevometer	22
Average cohesion [kPa]	Trenching tests	0.45
	Bevometer	1.01
	Sheargraph	2.07

maneuvers/systems using crushed basalts from Napa, CA, USA (Sibille et al. 2006; Oravec 2009). Classified as “granular with angular to sub-angular grains” (Green and Melzer 1971), these materials no longer exist and the library of documents describing their compositions and properties is incomplete (Sibille et al. 2006). The most comprehensive overview of LSS properties is provided in Oravec (2009), and Table 15.5 presents some general parameters derived from trafficability tests.

Since the creation of the LSS materials, additional simulants have been developed to serve a variety of purposes and investigations throughout the past several decades. Three of these—JSC-1A and its predecessor JSC-1, the NU-LHT series, and the GRC series (GRC 1 and 3)—will be introduced and briefly described. While other simulants have been created for specific purposes, these three simulant families are of interest for two reasons: (1) they represent lunar mare, lunar highland, and specifically geotechnical simulants, and (2) they have all been used in both manual and controlled-mechanism penetrometer investigations for the purposes of predicting lunar surface behavior. The Planetary Simulant Database at the Colorado School of Mines (<https://simulantd.com/>) contains a complete listing of current lunar simulants.

15.5.2 JSC-1 and JSC-1A

15.5.2.1 JSC-1

The JSC series of lunar simulants is one of the best known and widely used simulant families ever produced, beginning in the 1990s with the JSC-1 lunar all-purpose

simulant generated at the Johnson Space Center (JSC) for the purposes of developing lunar EVA suits (Sibille et al. 2006).

JSC-1 is a general-use mare simulant with low titanium content made from volcanic ash in the San Francisco lava field near Flagstaff, AZ, on the flank of the Mirriam cinder cone (Sibille et al. 2006). It is a glass-rich crushed basaltic ash containing rich oxidized forms of silicon, aluminum, iron, calcium, and magnesium that approximates the bulk chemical composition and mineralogy of the Apollo 14 sample 14163 (McKay et al. 1993; Klosky et al. 2000). Its mineralogy includes olivine, pyroxene, ilmenite, plagioclase, and basaltic glass (Sibille et al. 2006), and it is considered a well-graded silty sand (Klosky et al. 2000).

The most thorough geotechnical analysis of JSC-1 was performed by Klosky et al. (2000), though they note that previous authors (McKay et al. 1993; Willman et al. 1995; Perkins and Madson 1996) had already investigated the simulant's specific gravity, grain-size distribution, and mineral content. Using vibratory compaction to simulate the assumed depositional characteristics of real lunar soil, they performed triaxial compression and isotropic vacuum unloading experiments to determine JSC-1's shear and elastic properties: deviatoric stress and axial strain to axial stress, friction angle, cohesion, Young's modulus, and bulk modulus. They describe high values of cohesion (from ~4 kPa to over 14 kPa) and friction angle (44.4–53.6°) that increase with relative density with maximum and minimum densities of 1.83 and 1.43 g/cm³ respectively. Perkins (1991) reported friction angles between 41 and 60° and cohesion values between 0.1 and 2.5 kPa.

While ~12,000 kg of JSC-1 was produced, it was widely distributed to researchers and not tracked, stored, or utilized properly. As a result, little is known about how much is left, and what condition it is in (Sibille et al. 2006).

15.5.2.2 JSC-1A

After the original volume of the JSC-1 simulant was exhausted, NASA commissioned the production of another 16 tons (~14,500 kg) of a similar simulant through a coordinated grant in 2005. This included 14 tons of a JSC-1 clone called JSC-1A and one ton each of a coarse (JSC-1AC) and fine (JSC-1AF) version, all produced at the same quarry as the original (Zeng et al. 2010a). JSC-1A is no longer commercially obtainable, but costs ~\$20,000 per ton when available.

As with JSC-1, JSC-1A approximates a low-titanium mare regolith and contains major crystalline phases of plagioclase, pyroxene, olivine, and minor oxide phases of ilmenite and chromite (Alshibli and Hasan 2009), though the presence of plagioclase is disputed by Ray et al. (2010).

The chemical/mineralogical composition of JSC-1A, along with its physical and strength properties, engineering properties, and geotechnical properties have all been investigated and characterized by various authors. Ray et al. (2010) characterized JSC-1A by X-ray diffraction (XRD), scanning electron microscope (SEM), differential thermal and thermo-gravimetric analyses, chemical analysis, and Mössbauer

Table 15.6 Chemical compositions of the JSC series of lunar simulants compared to a lunar sample from Apollo 17. Modified from Ray et al. (2010)

Constituent oxides	JSC-1 [%]	JSC-1A [%]	Apollo 17 Sample 70051 [%]
SiO ₂	47.2	45.7	42.2
Al ₂ O ₃	15.0	16.2	15.7
CaO	10.4	10.0	11.5
MgO	9.0	8.7	10.3
FeO	7.4	–	12.4
Fe ₂ O ₃	3.4	12.4	–
Na ₂ O	2.7	3.2	0.2
K ₂ O	–	0.8	0.1
TiO ₂	1.6	1.9	5.1
P ₂ O ₅	–	0.7	–
MnO	–	0.2	0.2

spectroscopy. The results, showing the weight percentage (wt%) composition of JSC-1A as compared to JSC-1 and samples from Apollo 17, are presented in Table 15.6. The high glass content—similar to the lunar soil—also allowed for the creation of various glass preforms such as glass hairs and beads (Fig. 15.8).

The physical and strength properties of JSC-1A were investigated by Alshibli and Hasan (2009), who compared its particle-size distribution to that of the range of returned Apollo samples and found it to be within ± 1 standard deviation (SD). The specific gravity of the simulant was found to be 2.92 compared to 2.90 for JSC-1 (McKay et al. 1993) and 2.9–3.4 for Apollo samples (Carrier et al. 1991), with a maximum and minimum density of 2.106 and 1.556 g/cm³ respectively, compared to reported values of 1.93 and 0.87 g/cm³ (Carrier et al. 1991). Triaxial tests provided average ranges for the Young's modulus, shear modulus, and Poisson's ratio (Alshibli and Hasan 2009), while scanning electron microscope (SEM) images showed highly

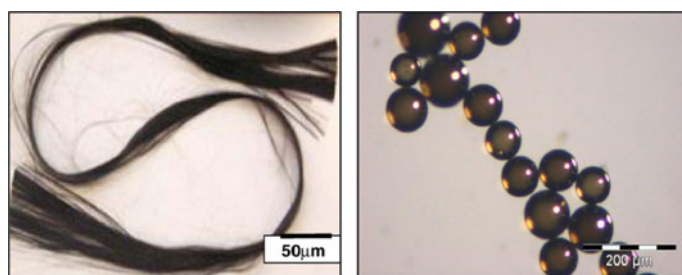


Fig. 15.8 Left: Glass fibers prepared from JSC-1. Right: Hollow glass microspheres produced from JSC-1. Modified from Ray et al. (2010)

angular shapes and surface crevices reminiscent of lunar regolith images. Finally, a peak friction angle range of $\sim 40\text{--}59^\circ$, increasing with density, was measured.

A geotechnical analysis of the simulant was performed by Zeng et al. (2010a). In addition to defining particle-size distributions, specific gravity, and maximum/minimum densities similar (though not identical) to Alshibli and Hasan (2009), triaxial testing determined the stress/strain characteristics as well as the shear behavior under increasing normal stress. Cohesion was found to be too low to measure, and the peak friction angle was found to be high and to increase with density. The low cohesion value described by Zeng et al. (2010a) was eventually determined to be exceptionally small, from 0 to 1.1 kPa by Li et al. (2013).

15.5.2.3 NU-LHT

The first general lunar highlands regolith simulant was developed in the early 2000s by the USGS and named the NU-LHT series. NU-LHT-1M was a pilot simulant, and -2M a prototype, matching the modal mineral and glass content, average chemical composition, and grain-size distribution of Apollo 16 regolith samples as closely as possible (Stoeser et al. 2010). It is not known if NU-LHT simulant is currently available, but it had a cost similar to JSC-1A at $\sim \$20,000$ per ton.

The composition of NU-LHT is a combination of mostly intrusive igneous rocks: Stillwater norite, anorthosite, hatzburgite, and Twin Sisters dunite. The simulant included pseudo-agglutinates formed of partially melted Stillwater mill waste (from the Stillwater Mining Company of Nye, MT), while fully melted waste constituted what was termed “good glass” (Stoeser et al. 2010). NU-LHT-1M consisted of 80% crystalline, 16% agglutinate, and 4% glass components, while -2M consisted of 65, 30, and 5% respective components. The bulk chemistry is reported in Table 15.7.

Geotechnical properties of NU-LHT-2M were investigated by Zeng et al. (2010b), including the particle-size distribution, specific gravity, maximum and minimum densities, triaxial testing, and peak friction angles. The particle-size distribution of NU-LHT-2 M falls within ± 1 SD of the lunar soil reported by Carrier et al. (1991), except at the finest particle sizes. It is classed as a well-graded silty sand, and a specific gravity of 2.749 was identified as being lower than that of typical lunar regolith. The maximum dry density was 2.057 g/cm^3 and the minimum 1.367 g/cm^3

Table 15.7 Composition of NU-LHT-1M prototype. From The Planetary Simulant Database (<https://simulantdatab.com/>), reproduced by permission of the USGS

Constituent oxides	Wt %
SiO ₂	47.6
Al ₂ O ₃	24.4
FeO	4.3
MgO	8.5
CaO	13.1
Na ₂ O	1.4

(compared to 1.93 and 0.87 g/cm³ respectively for lunar soils). A peak friction angle of 36–40.7° was determined to be lower than typical lunar soils, but that it increased with density.

15.5.2.4 GRC-1 and GRC-3

The GRC-1 and GRC-3 lunar simulants were developed at Glenn Research Center around 2009–2011 as purely geotechnical simulants designed for testing roving vehicle wheel traction in lunar soils (GRC-1, Oravec 2009) and excavation studies (GRC-3, He et al. 2011). They were developed as a readily available sand mixture and, at an affordable cost of \$250 per ton (compared to ~\$20,000 per ton for JSC-1A or NU-LHT) (He et al. 2011), facilitated the use of large quantities for vehicle mobility and excavation testing. Composed primarily of quartz sand, the GRC series is a combination of commercially available sands from the Best Sand Corporation of Chardon, Ohio and, in the case of GRC-3, a natural loess (Bonnie silt) from Burlington, Colorado comprises a finer component. GRC-1 is currently available from Black Lab (Covia) in Chardon, Ohio; it is not known if GRC-3 is commercially available at this time.

Geotechnical properties of GRC-3 were investigated by He et al. (2011), with the standard determination of particle-size distribution, specific gravity, maximum and minimum densities, peak friction angle, cohesion, shear and stress–strain behavior. The particle-size distribution slightly exceeds the ± 1 SD limit of typical lunar soils at both median and very fine particle sizes and it is classified as a silty sand. The specific gravity, 2.633, is lower than that of lunar soils, while the maximum and minimum densities (1.939 and 1.520 g/cm³) are within the typical lunar soil range. The peak friction angle of 37.8–47.8°, as with both JSC-1A and NU-LHT-2M, is lower than that of typical lunar soils but increases with density. Cohesion was determined to be essentially negligible (as expected for sands).

A summary of the pertinent geotechnical, physical, and strength properties of the simulants JSC-1, JSC-1A, NU-LHT-2M, and GRC-1 and -3 are shown in Table 15.8, and compared to the values of typical lunar soils as determined by Carrier et al. (1991).

Table 15.8 Summary of pertinent simulant properties compared to recommended values representing lunar regolith

Property	Lunar Soil ^a	JSC-1	JSC-1A	NU-LHT	GRC-1	GRC-3
Specific gravity	2.9–3.4	2.90	2.92	2.75	2.58	2.63
Max density [g/cm ³]	1.93	1.83	2.11	2.06	1.89	1.94
Min density [g/cm ³]	0.87	1.43	1.56	1.37	1.60	1.89
Peak friction angle [°]	41–55	41–60	40–59	36–41	~ 39	38–48
Cohesion [kPa]	0.4–3.8	0.1–2.5	0.0–1.1	–	–	–

^aAll depth ranges

15.6 Penetrometer Tests in Lunar Simulants

15.6.1 Introduction

The return of over 300 kg of lunar rocks and soil from the Apollo missions of the 1960 and 1970s enabled detailed investigations—including penetrometer testing—of the regolith's mechanical properties. The first lab measurements of the lunar regolith's shear strength were performed in 1969 in the Lunar Receiving Lab at the NASA Manned Spacecraft Center (now known as the JSC) and were the first of many performed on the samples returned by Apollo 11.

Shear investigations consisted of a standard penetrometer test where a flat hand penetrometer was pressed into several hundred grams of compacted lunar regolith (Carrier et al. 1991) then sieved to remove larger clasts (>1 mm) and kept in a dry nitrogen atmosphere to prevent adsorption of ambient moisture. The results show generally increasing penetration force as a function of depth, with higher-density samples more resistant to penetration at all depths (Table 15.9). Similar penetration tests were performed by Jaffe (1971) on 6.5 g of returned Surveyor 3 regolith.

While no additional laboratory penetrometer tests were apparently performed on the returned lunar soil, it is worth mentioning relevant shear investigations that complement the CPTs. Carrier et al. (1972, 1973) performed three direct shear tests

Table 15.9 Laboratory hand penetrometer measurements on lunar soil samples from Apollo 11 sample no. 10084. Modified from Carrier et al. (1991), reproduced by permission of The Lunar and Planetary Institute, Houston

Test	Density [g/cm ³]	Force [N]	Area [cm ³]	Pressure [kPa]	Penetration [cm]
1	1.36	<1.8 ^a	0.316	<57	0.64
2	1.36	<1.8 ^a	0.316	<57	1.96
3	1.36	<1.8 ^a	0.316	<57	1.96
4	1.36	<1.8 ^a	0.316	<57	1.96
5	1.36	3.1	2.68	11.4	2.01
6	1.77	1.8	0.316	57	0.81
7	1.77	5.1	0.316	171	1.70
8	1.77	<1.8 ^a	0.316	<57	0.64
9	1.77	9.8	0.316	308	2.54
10	1.77	5.8	0.316	183	2.11
11	1.77	38.7	2.68	143	1.70
12a	1.80	28.9	2.68	108	0.66 ^b
12b	1.80	79.8	2.68	297	1.96 ^b

^aPenetrometer did not meet with sufficient resistance, tabulated force is weight of penetrometer

^bPenetrometer was removed after achieving 108 kPa at 0.66-cm depth, then reapplied at the same place until achieving 297 kPa at 1.96-cm depth from original sample surface

(in a vacuum) on 200 g of Apollo 12 soil. They noted that the resulting measured cohesion and friction angle were significantly lower than those of a basaltic simulant, which they attributed to the crushing of weak particles such as agglutinates and breccias unique to the lunar regolith. More precise shear tests (triaxial, direct shear, etc.) were performed on Surveyor 3 scoop samples (Scott 1987) and Luna 16 and 20 samples (Leonovich et al. 1974, 1975).

15.6.2 Apollo Era

The early 1970s saw the first recorded penetrometer experiments on the newly created lunar soil simulant (LSS) (see Sect. 15.5). Costes et al. (1971) performed CPTs in LSS and Yuma sand of various grain-size distributions and consistencies under terrestrial conditions and on-board parabolic flights achieving 1/6, 1, and 2 g in order to investigate the effect of gravity. The results indicated that the average penetration resistance (q_c) and the average rate of change in q_c with depth (z) of the simulants decrease monotonically with decreasing gravity (g) and are sensitive indicators of soil bulk dry density, void ratio, and relative density (Fig. 15.9). They further claimed that q_c and G could be used with bearing capacity theory to determine in-situ shear strength and developed these analytical methods for application to crude soil penetration data from Apollo 11 and 12, determining preliminary measures of cohesion and other soil properties.

An extensive experimental program was undertaken in 1971 to determine the penetration resistance of an unnamed lunar soil simulant and translate the detailed relationships to the lunar surface (Houston and Namiq 1971). The simulant was

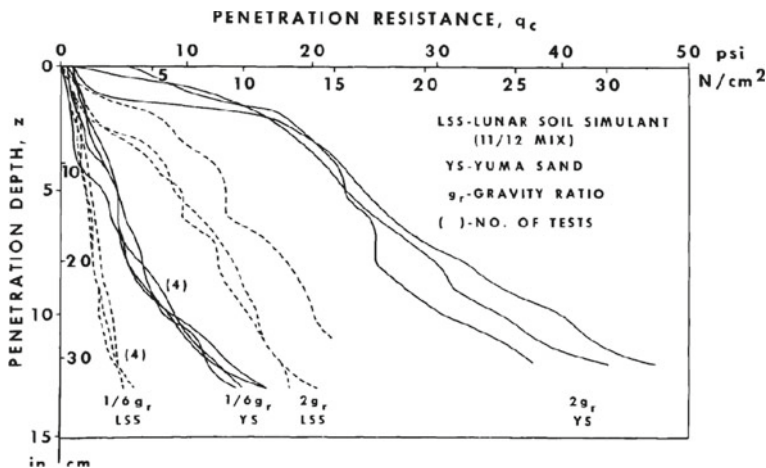


Fig. 15.9 Cone penetration resistance (q_c) versus penetration from tests on Yuma sand and a mix of LSS 11/12 performed under varying gravity conditions. Modified from Costes et al. (1971)

prepared by mixing crushed basalt powder with basalt sand, selected and modified based on Surveyor and Apollo 11 compositions, gradation curves, and cohesion values. Of particular note was that the authors found the addition of 2% water to the simulant generated enough cohesion to mimic that estimated for lunar regolith cohesion values. This mass percent of moisture is still in use today.

Ultimately, the experimental campaign generated an estimate of ultimate bearing capacity that could be applied to in-situ lunar soils (Houston and Namiq 1971)

$$q_{ult} = \frac{B\gamma}{2} N_{\gamma} s_{\gamma} + c N_c s_c + q' N_q s_q, \tag{15.8}$$

where q_{ult} is the unit ultimate bearing capacity (in N/m^2), q' the surcharge stress (in Pa), N_{γ} , N_c , N_q the dimensionless bearing capacity factors for friction, cohesion, and the surcharge respectively, and finally s_{γ} , s_c , s_q dimensionless shape factors. The analysis generated an estimate of the variation in G with average void ratio (Fig. 15.10).

Houston and Namiq (1971) concluded that the bearing capacity equation provided a reasonable estimate of G if local shear strength parameters were used (assuming these could be obtained) and predicted that the factor by which the penetration resistance is reduced in lunar gravity ($\sim 1/4$) is less than the reduction in gravity ($1/6$). Finally, they predicted that stress penetration gradients could be used to indicate heterogeneity in lunar soil.

A final set of penetration experiments in LSS2 was performed by Durgunoglu and Mitchell (1973), where the application of the authors' analytical models to static CPT measurements showed good agreement (Fig. 15.11). Using a 15° conical penetrometer 2 cm in diameter, manual insertion into simulant prepared over a wide variety of densities resulted in penetration profiles that were compared with

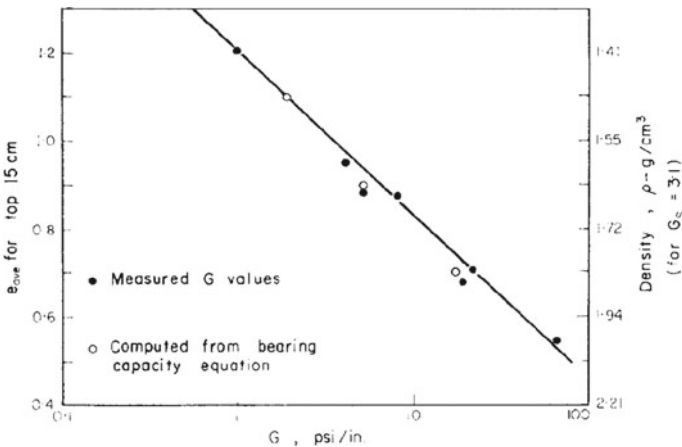


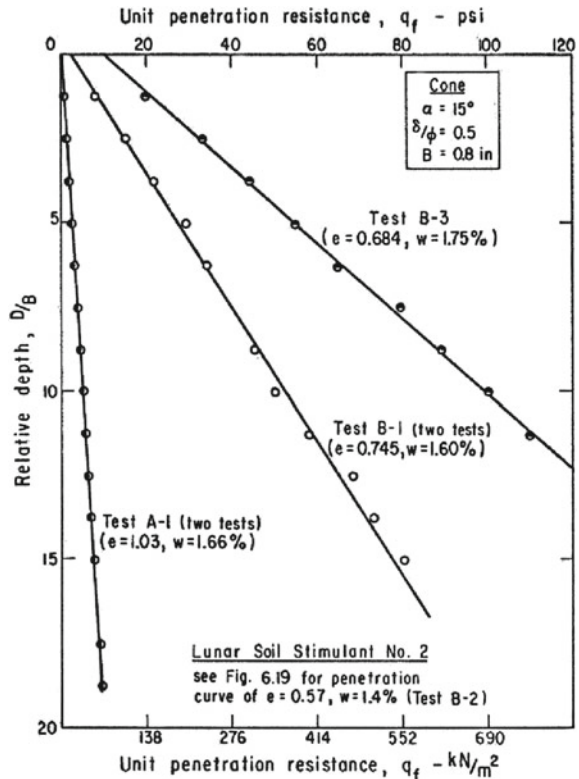
Fig. 15.10 Comparison of measured and computed G values for an unnamed lunar soil simulant under full gravity. Source Houston and Namiq (1971)

analytically-predicted penetration resistance using the formula

$$q_f = cN_c s_c + \gamma BN_{\gamma q} s_{\gamma q} \tag{15.9}$$

While not apparent from the figure, low-density (high void ratio, i.e., Test A-1 in Fig. 15.11) predictions were less accurate. The authors explained this decreased model accuracy at lower densities as due to the significant influence of soil compressibility on resistance.

Fig. 15.11 Measured penetration curves for LSS2. Source Durgunoglu and Mitchell (1973)



15.6.3 Manual

There was a general hiatus in penetration testing in lunar regolith simulants for roughly three decades until the early 2000s, when a new geotechnical simulant (GRC-1) was being developed at Case Western Reserve University—under a cooperative agreement with the NASA Glenn Research Center—by H. Oravec during her doctoral

thesis, in which she used CPT measurements of various permutations to compare the new simulant to lunar trafficability estimates (Oravec 2009).

The penetration experiments were performed by manual insertion (Fig. 15.12) of a 30° field cone penetrometer of two interchangeable cone areas—130 and 323 mm²—into several different-sized bins (from 55 to 75 cm in diameter) filled in uniform layers using a hopper and compacted to specified densities. The cone-to-container radius ratio ($CCR = \frac{r_{container}}{r_{cone}}$) for these experiments ranges from 36 to 43, indicating that dominant edge effects would not be expected. CCR relates the radius of the penetrometer to the radius of the sample container and should generally be greater than 10 to avoid edge effects in confined sample testing.

Four tests of 16 insertions each (at an attempted rate of ~2 cm/s) were analyzed to determine the effect of cone size and repeatability of testing in ambient conditions. For each cone size and depth, the cone index gradient (G) was determined and a generally expected increase in gradient with depth observed (Table 15.10). Oravec (2009) states that the increase is likely due to compaction and compression ahead of the probe.

The determination of G , however, involves an assumption of linearity in the penetration profiles, and nonlinearity can cause a significant difference in the resulting gradient values. Repeated mentions of nonlinear behavior are noted in the study and are most clearly seen in the penetration profiles of higher-density samples for both cone sizes (Fig. 15.13), while low-density samples show more linear behavior. Irrespective of linearity, the mean G showed the expected increase with density (Fig. 15.14).

Fig. 15.12 Cone penetration tests in GRC-1 at the NASA Glenn soils lab. *Source* Oravec (2009), credited to NASA GRC



Table 15.10 Cone index gradient (G) in GRC-1 as a function of depth intervals. Modified from Oravec (2009), reproduced by permission of Heather Oravec

Soil depth [mm]	Cone index gradient (G) [kPa/mm]
5–15	4.03
15–25	4.01
25–35	4.56
35–45	4.09
45–55	5.06
55–65	4.93
65–75	5.08
75–85	5.79
85–95	5.78
95–105	5.74
105–115	6.91

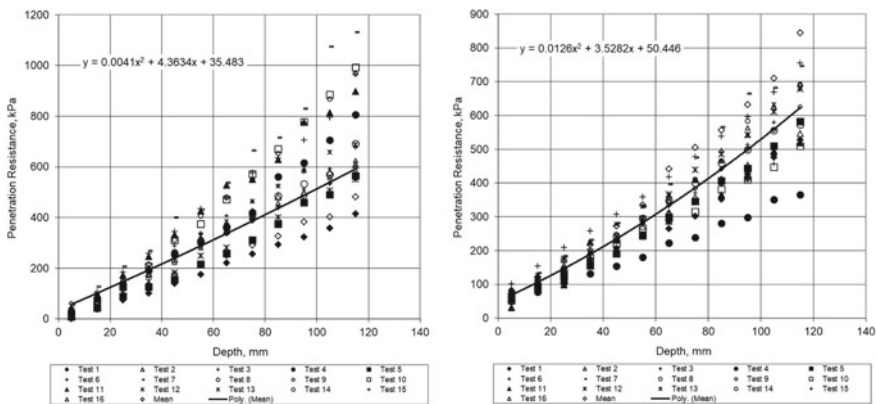


Fig. 15.13 Penetration profiles of the small (left) and large (right) cone penetrometer tests in GRC-1, showing nonlinearity. Modified from Oravec (2009)

After determining the effect of density on G at ambient conditions for the GRC-1 simulant, SRP data from Apollo 15 and 16 were analyzed in the same manner. A large range in G was predicted, potentially demonstrating regional variation in lunar soil conditions. The lower end of the predicted values (2.22–5.08) for the Apollo 15 tests (Tests 1–4) correspond generally well to those estimated by Mitchell et al. (1972) of 2.98–5.97 (as cited by Costes et al. 1972). Differences between the G values for GRC-1 and the lunar estimates are claimed to be due to the ambient testing conditions (pressure and temperature), the presence of moisture in terrestrial samples, boundary conditions of the plastic testing bins, and the artificial sample preparation method using vibration.

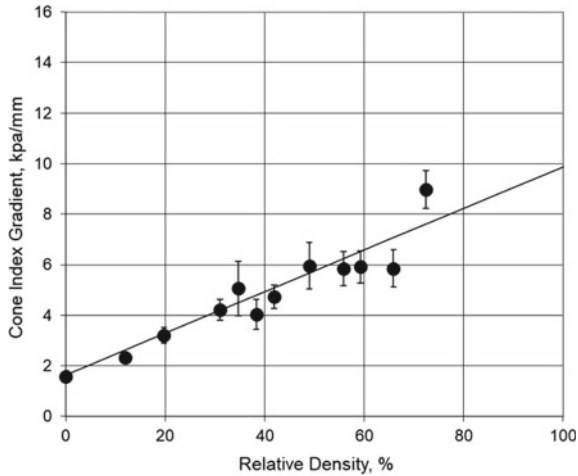


Fig. 15.14 The expected linear increase in cone index gradient G as a function of relative density in GRC-1 (with standard deviation). *Source* Oravec (2009)

The response of manual penetration resistance to ice content in icy regolith simulants has been investigated by Mantovani et al. (2016) and Pitcher et al. (2016), using considerably different approaches and control of ambient conditions.

Mantovani et al. (2016) built upon previous work by Honeybee Robotics showing that a percussive cone penetrometer was capable of penetrating lunar regolith with a fraction of the force required using an ordinary field penetrometer. Using a percussive cone penetrometer developed by Honeybee Robotics (Fig. 15.15) capable of delivering 2.6 J of energy per blow at a frequency of 1500–1700 blows per minute, the penetration rate (a proxy for resistance) into samples of JSC-1A containing ice contents of 0–8% by mass was measured.



Fig. 15.15 The Honeybee Robotics percussive cone penetrometer. *Source* Mantovani et al. (2016)

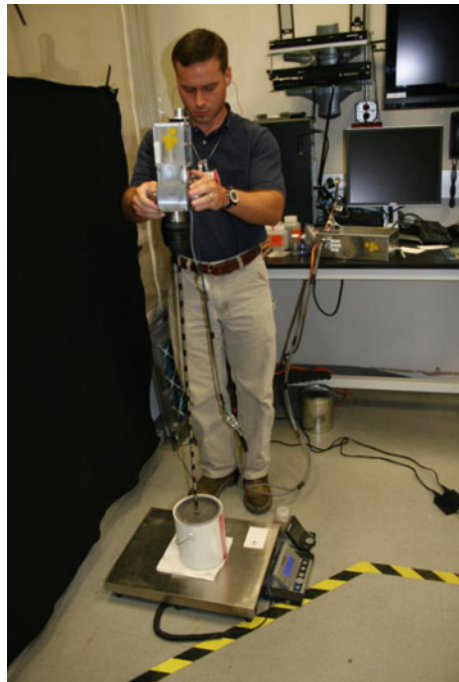
Samples were created by mixing JSC-1A with water (in 1% increments) and compacting varying layers into paint cans, followed by an undescribed quick-freezing process to an unknown temperature. Additionally, a 1-m column containing 10 layers

of alternating icy/dry simulant was prepared in a similar fashion and frozen to -60°C (213 K) overnight. Manual penetration of the samples took place outside the freezer in ambient conditions (Fig. 15.16), where some warming of the samples is expected to have occurred. An electronic scale below the samples measured the applied vertical force, and the rate and depth of penetration were determined by analyzing video footage of the tests.

It was noted that the speed of penetration decreased with increased ice content, and that the operator was unable to penetrate the simulant with ice contents of 6% or greater. The observed deeper penetrations into some samples (e.g., one sample of pure 100% ice) is expected to be due to the fact that fractured ice/regolith “chunks” have room to move into the surrounding volume, thus allowing additional penetration. Such a theory might be correlated to the concepts of compression and dilation described in Puech and Foray (2002). Additionally, the penetration rate was not a strong function of downward force and suggests that it could be used as an indication of ice content.

The mechanics of penetration into dry or icy materials is assumed to be quite different, as the grains in frozen soils are unable to rearrange themselves when subjected to stress (and subsequent strain) from a penetrating probe. The apparent transition in these two mechanical states occurs between 3 and 5% of ice by mass in JSC-1A in this particular study. Penetration into a multi-layered column (Fig. 15.17) shows that the interleaving dry layers allow room for particle motion, which facilitates

Fig. 15.16 Manual (percussive) cone penetration into a sample can. *Source* Mantovani et al. (2016)



penetration into icy layers that were unable to be penetrated in the previously layered samples (paint cans). It was also noted that, in this case, increased downward force aided penetration as the additional force helped to push fractured material into the surrounding dry material space and create space for the penetrometer to pass.

The experiments, while not well-controlled in terms of experimental conditions, show that a percussive cone penetrometer is capable of penetrating icy regolith at ice contents that a static penetrometer is not, and in a manner insensitive to downward force. Of particular note is the fact that the ultimate penetration depth was not only a function of ice content, but also of the availability of space around the penetrometer tip for relocation of fractured and dislodged material.

Another approach to manual penetration testing was taken by Pitcher et al. (2016), who used both a pencil and a field penetrometer to investigate the properties of icy NU-LHT-2M samples. Preliminary attempts to identify the saturation point of the icy simulant and explore its properties involved mixing six samples of simulant with increasing amounts of water in small ($1.34 \times 10^{-4} \text{ m}^3$) containers, which were frozen overnight at $-20 \text{ }^\circ\text{C}$ (253 K). Qualitative measurements of penetration were

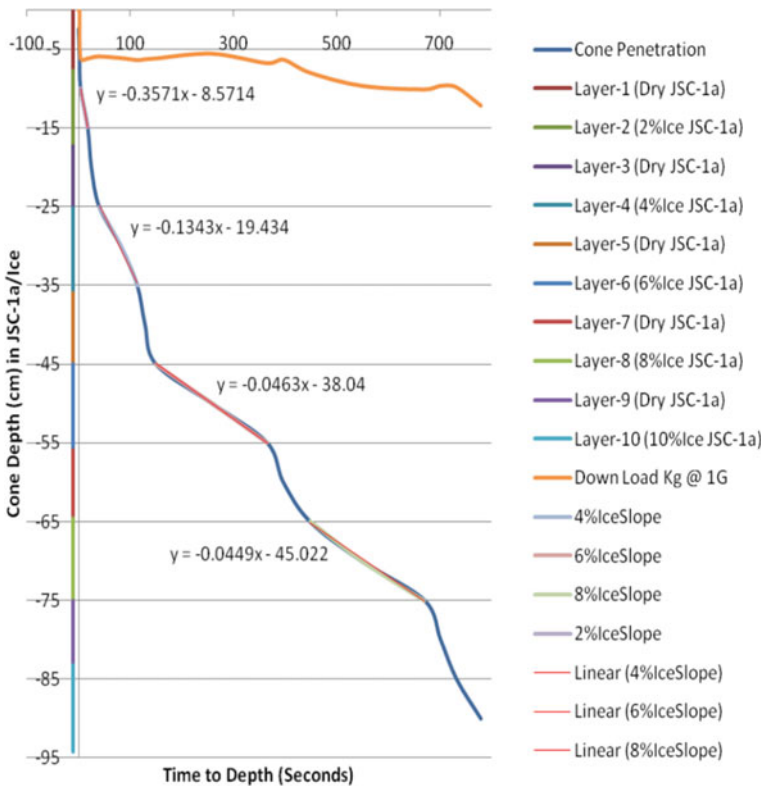


Fig. 15.17 Depth of penetration and down-force as a function of time for a 1-m column, showing interleaving dry layers that allow room for particle motion. *Source* Mantovani et al. (2016)



Fig. 15.18 Six frozen samples of NU-LHT-2M with increasing volumes of water added. Each container has a diameter of 7.2 cm and a height of 3.3 cm. *Source* Pitcher et al. (2016)

performed by pushing a pencil with a 5-mm conical tip into the samples (potentially in ambient conditions) and taking subsequent pictures (Fig. 15.18) to accompany a table of observations.

The qualitative assessment indicated that the frozen regolith experienced a rapid change from ‘soft’ to ‘hard’ when the ice content (water mass) was in the range of 5–9%. To further investigate the rapid change over such a narrow range of ice content, the penetration resistance of icy samples containing 3–4, 4–5, and 7–8% ice was measured (Figs. 15.19 and 15.20). However, useful interpretation of the results is complicated by various uncontrolled experimental factors including the testing temperature, repeated penetrations into single samples, and the creation of subsequently lower ice content samples by allowing a higher moisture sample to dry overnight and be refrozen.



Fig. 15.19 The manual field penetrometer and measurement technique in a frozen NU-LHT-2M sample. *Source* Pitcher et al. (2016)

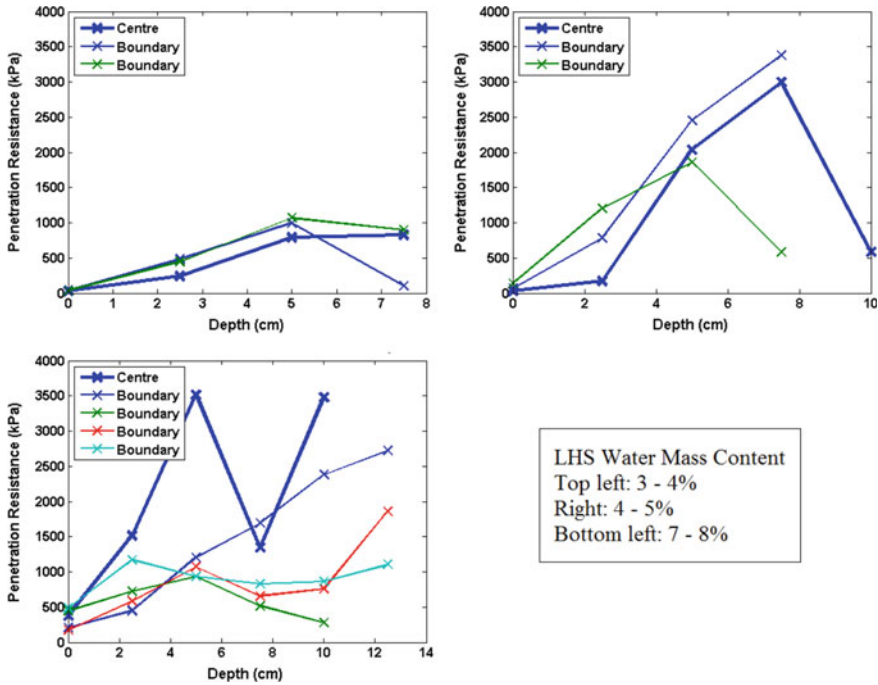


Fig. 15.20 Results of the penetration tests of the frozen NU-LHT-2M sample with different water contents, measuring the resistance experienced by the penetrometer at incremental depths in the sample. Source Pitcher et al. (2016)

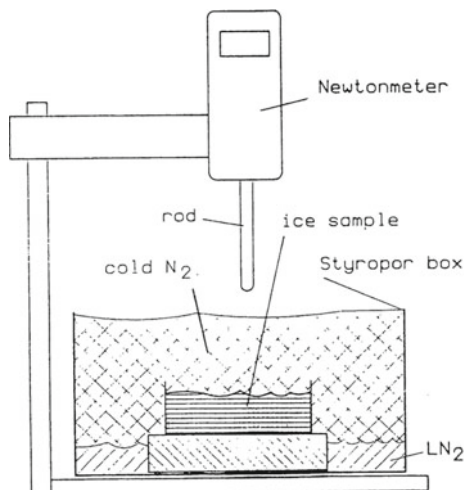
15.7 Controlled Mechanism

15.7.1 Introduction

Penetration into regolith simulants using a controlled mechanism—that is, an autonomous or semi-autonomous device capable of maintaining a vertical penetration angle with limited deviation, penetrating at a constant rate or maintaining a constant force, and in general controlling as many aspects of the penetration as possible—apparently began with the KOSI experiments (*Kometen-Simulation*) at DLR-Köln in the late 1980s (Kochan et al. 1989 and others). Intended to test comet analogs at cryogenic temperatures, the experiments aimed to help support the eventual Rosetta mission and used a specially designed testing apparatus to penetrate fluffy ice-mineral samples prepared by injection of an aqueous mineral suspension into LN₂.

The testing machinery (Fig. 15.21) consisted of a 5-mm diameter teflon rod with a hemispherical tip fixed to a force gage (piezocone forcemeter) that penetrated at a rate of 0.2 mm/s into the sample. The samples contained H₂O and CO₂ ice (15% by weight) and grains of olivine and montmorillonite of approximately 1 mm (10%

Fig. 15.21 KOSI hardness testing device with cold box and sample. *Source* Kochan et al. (1989)



by weight). Of interest is that some icy samples were exposed to solar radiation in a vacuum environment and developed a distinctive crust that was highly resistant to penetration.

Tests were carried out in specially designed boxes and N_2 -purged compartments to eliminate atmospheric moisture and maintain sample temperatures that ranged from ~ 115 K at the near surface to 110 K in the center and 90 K near the underlying cryogenic plate (as shown in Fig. 15.21). In all cases, LN_2 was used as the cooling agent of a baseplate upon which the sample sat to cool, while cold N_2 gas produced by the boiling LN_2 created a cold environment that also served to chill the penetrometer. This is the first example of a system in which the probe temperature was lowered towards that of the sample, though no attempt was made to monitor the temperature of the probe during penetration.

The well-controlled cryogenic tests showed the formation of a hard crust underneath a dusty mantle. While the dusty mantle showed almost no resistance to penetration, the icy crust was highly resistant, and resistance depended on the length of irradiation of the sample. Non-irradiated samples showed an initial parabolic increase to a constant ~ 200 kPa, while radiated samples demonstrated a ~ 7 mm thick crust with a strength of up to 1400 kPa, followed by a quick drop to 200–300 kPa below (Fig. 15.22). Samples irradiated for ~ 41 h maintained a crust of close to 5 MPa strength, almost that of dense crystalline ice at temperature (T) < 200 K. The authors attributed the crust formation to three physical processes occurring simultaneously: sublimation, diffusion, and condensation of volatiles in a porous medium.

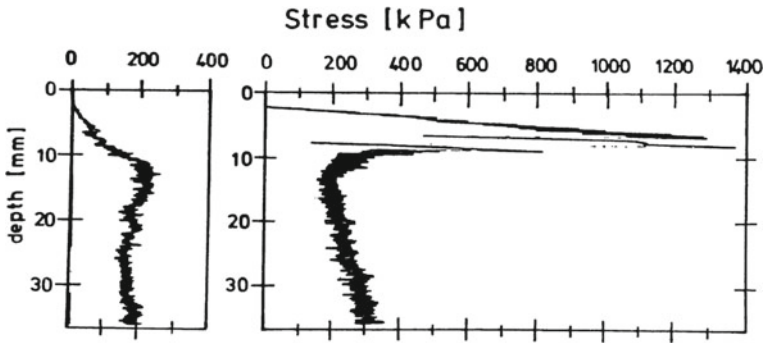


Fig. 15.22 Stress-depth profiles of an unirradiated (left) and an irradiated (right) model comet, showing the dramatic increase in crustal strength occurring after irradiation. *Source* Kochan et al. (1989)

15.7.2 Indentation

In the early 2000s, Gertsch and colleagues undertook an extensive indentation testing campaign on samples of JSC-1 with water content from <1% to full saturation (>12%) at cryogenic temperatures using an electro-hydraulic closed-loop servo-controlled indenter (Gertsch et al. 2006, 2008). While not precisely penetration tests, they nonetheless provided very useful information on the expected behavior of icy lunar simulant by providing estimates of “specific penetration” and “specific energy” (Teale 1965).

Samples were prepared by mechanically mixing water with fully dried JSC-1 to the desired percentage water content, then compressing the mixtures into 10.9-cm diameter stainless steel test rings at 467 N, intending to simulate the effect of long-term regolith compaction due to meteorite impacts. Samples were then sealed and submersed into LN₂ to cool them to 77 K, as measured by a Type-K thermocouple embedded inside.

Once cooled, the samples were placed into the test machine in ambient conditions (indicating that the samples would be warming continuously). The upper platen of the machine was brought to bear on the sample at 1.24 mm/s, pushing a 19-mm diameter hemispherical indenter vertically into the center of the sample. Once the sample failed—as indicated by measured force drop (Fig. 15.23) or visual confirmation (Fig. 15.24)—the indenter was withdrawn.

Results of the indentation tests indicate that both the specific penetration and specific energy of icy JSC-1 increase with moisture content. Additionally, an apparent change in failure mechanism—identified by subtle changes in failure morphologies and shapes of the load penetration curves—occurs between 1 and 1.3% moisture content and could indicate a transition from brittle to ductile behavior.

Parabolic behavior is again noted with respect to the specific energy of samples near to saturation, while drier samples show a linear relationship between specific energy and water content. Interestingly, the excavated volume (Fig. 15.25) shows a

Fig. 15.23 A load penetration (indentation) curve from a 1.48% water content sample of JSC-1, showing multiple failures (load drops) during indentation. *Source* Gertsch et al. (2006)

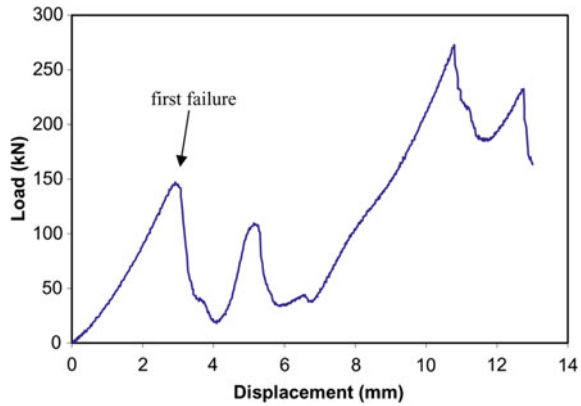


Fig. 15.24 Close view of a sample immediately after indentation, showing some of the chips and the fines produced. *Source* Gertsch et al. (2008)



power law decrease with increasing water content, with a particularly sharp transition between 1 and 3% water content. The authors note that there appears to be a bilinear function between the two regions.

Indenter penetrations into samples of various moisture contents clearly show increased brittle-like behavior with increased saturation (Gertsch et al. 2008). Penetrations into dry JSC-1 proceeded up to 16-mm depth with maximum loads under 5 kN; moist samples (0.6–1.5% water content) experienced depths of 12–14 mm and maximum forces of 15–27 kN; ~8 to 9% samples reached 200 kN at depths of roughly 6 mm; and samples with 10–12% water content behaved like strong sandstone with maximum loads of 200+ kN at 12-mm depth. Additionally, the specific penetration results were used to correlate unconfined compressive strength (UCS) estimates and compared with a single direct UCS measurement at 77 K, yielding an estimated UCS curve that predicts values from <20 MPa for low moisture content to >100 MPa at full saturation (>12% water content).

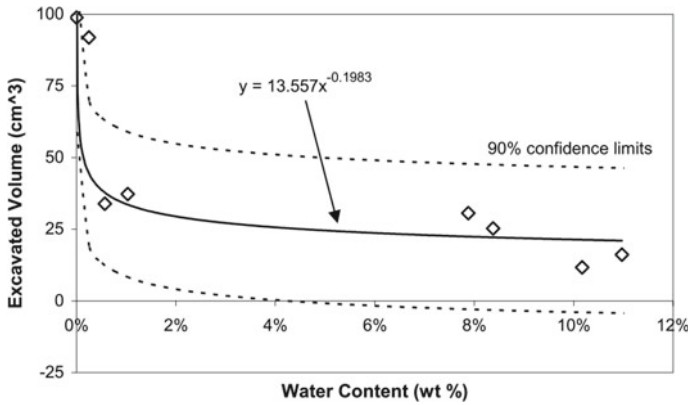


Fig. 15.25 The effect of water content (in JSC-1) on excavated volume, with 90% confidence limits. *Source* Gertsch et al. (2006)

While the authors note that the small number of experiments within the study should be augmented in order to provide details in the transition from ductile to brittle behavior—and in particular the linear or parabolic relationship between specific penetration and moisture content—they were confident in their assessment that at 77 K the icy mixtures behave more like a strong brittle material than a collection of noncohesive dry regolith particles. The transition is explained, according to the authors, by the sharp decrease in the mobility of weak hydrogen bonds in ice as temperature decreases. The reduced mobility increases the strength of the ice content, while the cementing behavior of ice in the unconsolidated granular regolith increases penetration resistance.

15.7.3 Penetration and Relaxation

True penetration tests on lunar simulants using controlled-mechanism penetrometers at both cryogenic and ambient temperatures, vacuum pressures, and dry and saturated states began in earnest only in the last decade. Additional attention to laboratory techniques and cryogenic methods have allowed for more robust explorations of the behavior of simulants at ambient and low-temperature conditions.

In the early 2010s, Kleinhenz and colleagues began a two-phase series of penetration experiments into GRC-3 and NU-LHT-3M simulants at NASA's Glenn Research Center. Using an electric cone penetrometer, they measured the strength, cohesion, friction angle, bulk density, and shear modulus of the simulants at both ambient and vacuum pressure conditions and ambient temperatures in a large sample bed with a depth of 64 cm and a surface area of $\sim 1 \text{ m}^2$, containing 1 ton of simulant. In Phase I, the CPT system was driven by a standard hand drill via a flexible shaft feedthrough into a jackscrew drive, pushing the cone at $\sim 1 \text{ cm/s}$ into the simulant of an unknown

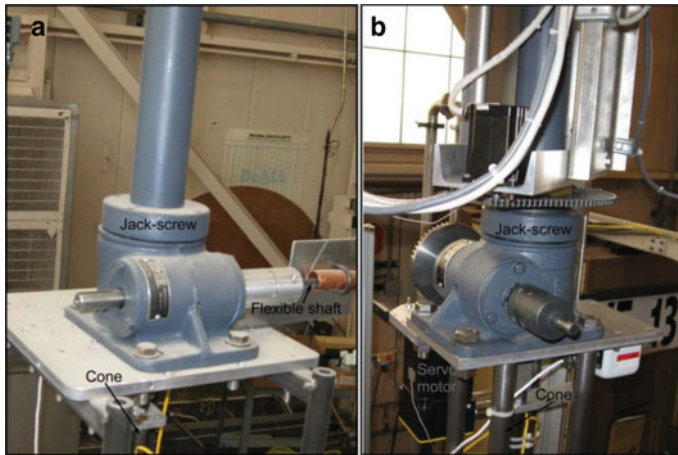


Fig. 15.26 The cone penetrometer drive system during Phase I (a, left) and Phase II (b, right). Source Kleinhenz and Wilkinson (2014)

density (Kleinhenz and Wilkinson, 2012). The pressure at the tip was recorded at 2–5-mm depth intervals, assuming the use of an internal strain gage common to electric penetrometers. Phase II saw the 2.54-cm diameter, 60° mini-electric cone driven by a servomotor (Kleinhenz and Wilkinson 2014) (Fig. 15.26).

Phase I results show the varying relationships of penetration resistance with depth for three penetrations: two in GRC-3 and one in NU-LHT-3M. Parabolic increases were seen most prominently in the NU-LHT-3M test, while the tests on GRC-3 showed undulating variations in resistance with depth that alternate between what appears to be a logarithmic behavior followed by parabolic behavior and may indicate two separate layers. The authors attribute the variation in resistances to both sample preparation technique and consolidation time, suggesting that uncontrolled experimental conditions affected the results. GRC-3 was rapidly pluviated into the bin while NU-LHT-3M was filled by dumping many large 5-gallon buckets. The beds were left to settle for different lengths of time.

Phase II, begun in November 2011, explored variations in pressure (vacuum versus ambient) and NU-LHT-3M simulant bed preparation (“tilled” versus tamping) (Fig. 15.27). Additional variations in the time between tests where sediment consolidation likely occurred create some difficulties in comparing the results, but in general the penetrations show that increased consolidation (increased density) resulted in increased penetration resistance. Interestingly, penetration resistance also appeared to increase with decreasing pressure, though no explanation of the phenomenon is provided.

In 2011, Cil (2011) performed penetrations into the lunar simulant JSC-1A and a simple Ottawa Sand using both a vehicle-mounted penetrometer system and a controlled mini-CPT. The study investigated the effect of cone and specimen size

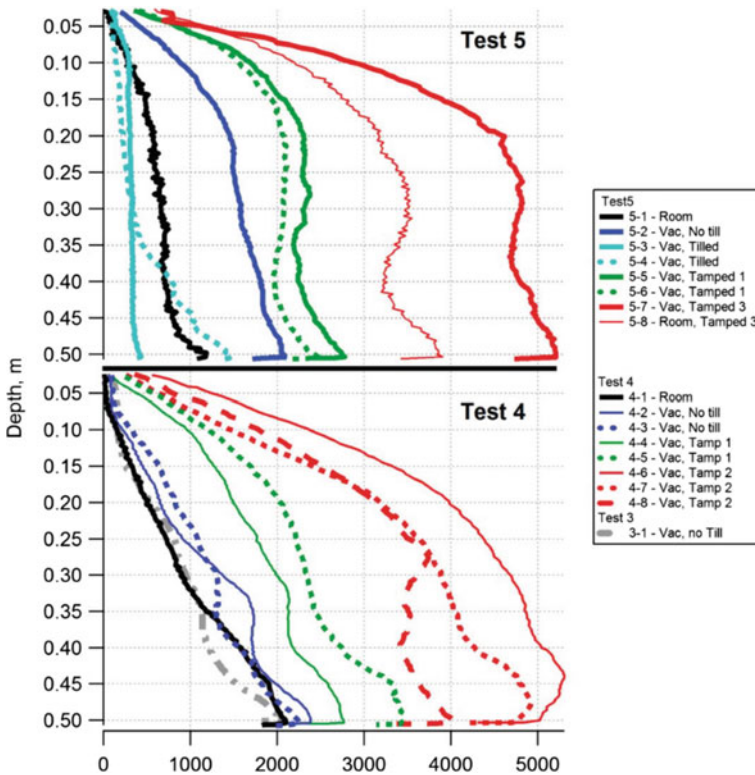


Fig. 15.27 CPT results in NU-LHT-3M, showing multiple tests at both ambient (‘Room’) and vacuum (‘Vac’) pressures and at various levels of tamping. An increase in penetration resistance with increased density (increased tamping) is seen, as well as an apparent increase in resistance with decreasing pressure. *Source* Kleinhenz and Wilkinson (2014)

(CCR) on penetration resistance, as well as the effect of boundary conditions on the behavior of the granular analog materials.

Vehicle-mounted penetrations were performed using a 20-ton truck into a cylindrical container 91 cm high and 13.84 cm in diameter, at various densities and pressures. While there is no mention of the probe size, Fig. 15.28 indicates that the CCR was quite low, perhaps on the order of 5. Similarly, no measurements of sample preparation density are provided, only “loose” and “dense”. The results of these penetrations show a general parabolic increase with depth, though one test shows the initial nonlinear increase followed by a “plateau” of resistance, similar to that predicted by Puech and Foray (2002) (Fig. 15.29 and Table 15.11). The reliability of these measurements, however, is questionable due to the low CCR, likely boundary effects, and limited number of data points.

Supporting measurements, in particular those providing supplementary data for the subsequent DEM model, were obtained using a controlled mini-CPT in JSC-1A and Ottawa Sand. The mini-CPT had a reported cone diameter of 3.125 mm and



Fig. 15.28 The testing set-up for CPT measurements in JSC-1A, showing the large-diameter truck-mounted penetrometer and the relatively narrow cylindrical testing container. *Source Cil (2011)*

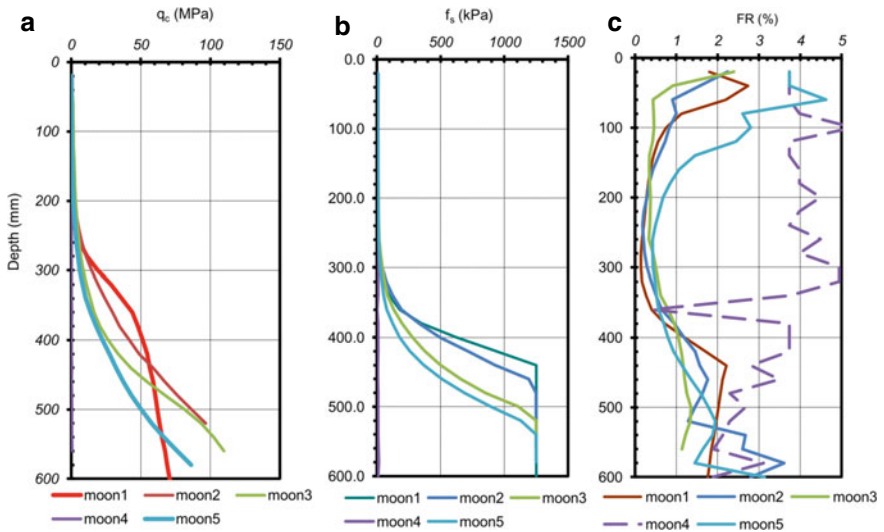


Fig. 15.29 Results of CPT measurements in JSC-1A showing tip resistance (**a**, left), sleeve friction (**b**, middle), and friction ratio (**c**, right). *Source Cil (2011)*

Table 15.11 Summary of field CPT experiments in JSC-1A. Modified from Cil (2011)

Experiment	Soil height before penetration [mm]	Soil height after penetration [mm]	Dry density [g/cm ³]	Pressure source
Moon1	726.4	Not recorded	1.78	Atmospheric
Moon2	706.1	Not recorded	1.78	Atmospheric
Moon3	709.9	571.5	1.78	+25 kPa
Moon4	769.6	670.5	1.75	-25 kPa
Moon5	607.1	Broken	1.75	-25 kPa

penetrations were performed in cylindrical containers of 25.4 and 40.5 mm radius, both 101.6 mm in height, giving CCRs of 8 and 13 respectively and suggesting that edge effects could influence results (Fig. 15.30). Sample preparation was noted as being the most challenging aspect of the experiment, and relatively little information on the resulting sample densities is provided. “Loose” samples were prepared by free-pouring simulant from a specified height through a funnel followed by vibratory compaction to a pre-determined surface level (back-calculated from the desired density), and “dense” samples were formed in three layers using a standard Proctor Method. Samples were loaded into a GeoJack machine, and penetration occurred at ~10 mm/min (0.17 mm/s), while a load cell-recorded resistance and displacement of the probe was measured with an LVDT sensor.

Results of multiple penetrations into both materials showed that penetration resistance increases drastically for low void-ratio samples and that high void-ratio samples see either a linear increase in resistance with depth or, in some cases, an initial

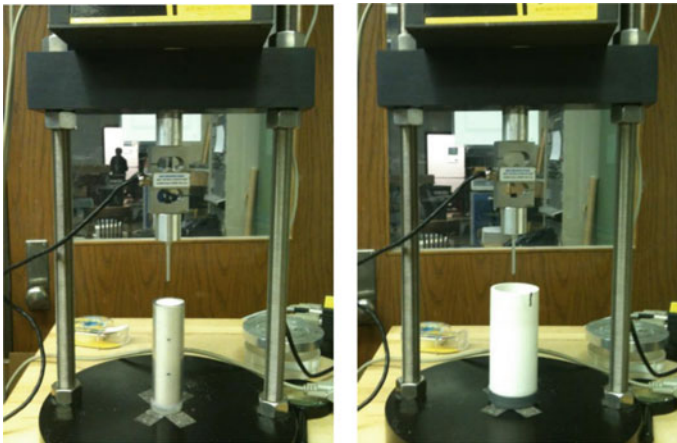


Fig. 15.30 Mini-CPT experimental set-up for supplementary input to a discrete-element model. Note the small penetrometer diameter but narrow container diameter (left) compared to the larger container (right). *Source* Cil (2011)

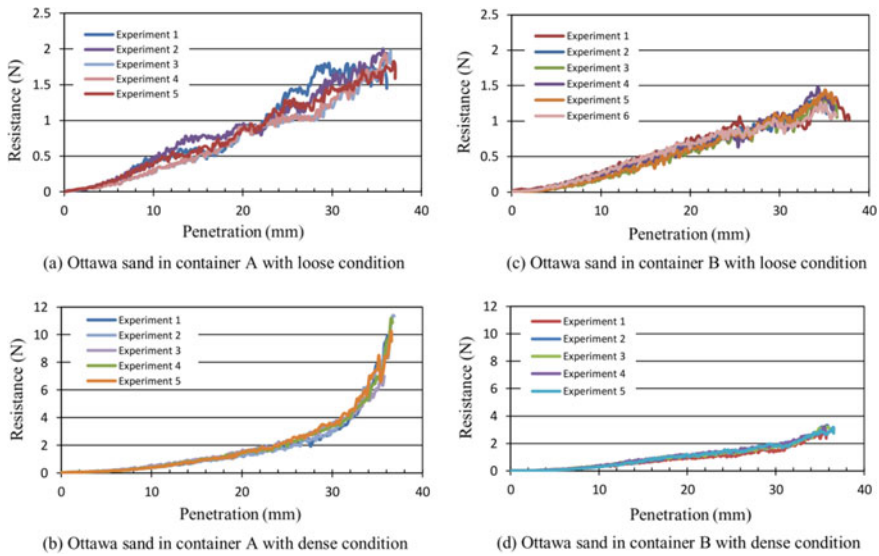


Fig. 15.31 Mini-CPT results in Ottawa Sand with dense and loose density conditions in the small (container A) and large (container B) containers. *Source* Cil (2011)

parabolic increase followed by a sustained constant force (similar to that predicted by Puech and Foray, 2002) (Figs. 15.31 and 15.32).

Cil (2011) claims that the variability seen in the penetration resistance of JSC-1A is due to inherent heterogeneity in identically prepared samples, a result of the simulants' particle-size distribution (as compared to that of Ottawa Sand). Furthermore, the sharp increase in resistance with depth (occurring at ~ 27 mm) in the "dense" condition for both materials is likely due to edge effects, observed to be most pronounced in the narrow container (A), and is a function of high particle confinement and particle interlocking. The plateau state achieved by JSC-1A in a dense condition and in the larger container (B) is taken as an indication that the soil boundary conditions have been removed. Ultimately, the experiments demonstrate the sensitivity of penetration resistance to sample density and container size (boundary effects), though various aspects of the penetration behavior (such as the flattening of the dense JSC-1A in the large container at depths > 20 mm while Ottawa Sand follows a predicted increase) remain unaddressed.

In 2014, Seweryn et al. (2014) proposed the use of a low-velocity penetrometer (LVP) for determining the geotechnical properties of regolith, in a method similar to a dynamic cone penetrometer but modified for use in space (low mass, low power). LVPs are penetrators that utilize low velocity, high stroke energy, and low power to autonomously generate forward motion in zero- or micro-gravity, and are designed to carry various sensors for in-situ investigations of planetary subsurfaces. Other examples of LVPs (Fig. 15.33) are the MUPUS system used on the Philae lander, the mole KRET, the CHOMIK, and the HP3 device used on the Mars InSight mission.

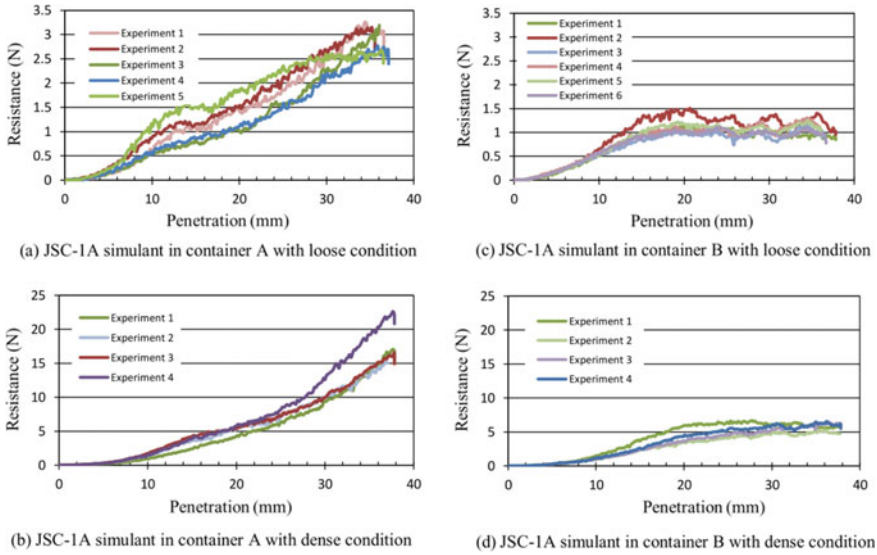


Fig. 15.32 Mini-CPT results in JSC-1A with dense and loose density conditions in the small (container A) and large (container B) containers. *Source* Cil (2011)

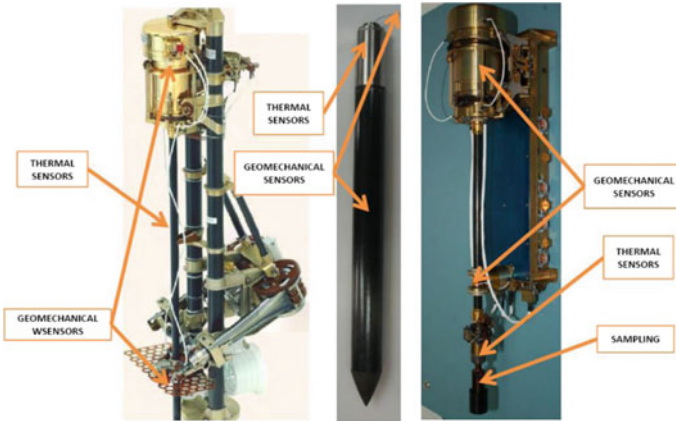


Fig. 15.33 Examples of the LVP devices. Left: the MUPUS instrument (Rosetta mission to comet 67P). Middle: the mole KRET penetrator. Right: the CHOMIK instrument (Phobos Grunt mission). *Source* Seweryn et al. (2014)

Results from penetration into lunar regolith simulants (AGK 2010 and a dry quartz simulant) using the KRET penetrometer show the expected increasing resistance with increased density (noted as “not compacted” to “highly compacted”), as indicated by the decreasing depth per stroke (DPI—dynamic cone penetration [DCP] penetration

index) in Fig. 15.34 and the shallower penetration depth of the penetrometer tip (Fig. 15.35) for highly compacted simulant.

A robust exploration of the penetration response of dry and icy lunar regolith simulants under controlled laboratory conditions—at variations in density, moisture content, pressure, and temperature—was performed beginning in 2019 by J. Atkinson during doctoral studies at the Center for Space Resources at the Colorado School

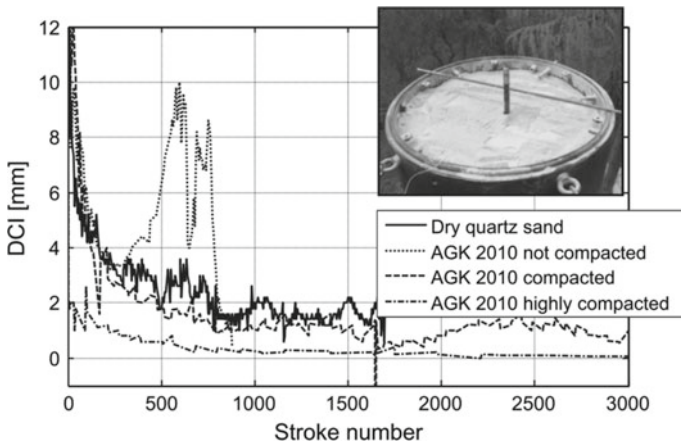


Fig. 15.34 A comparison of the DPI (DCP penetration index) parameters obtained using the KRET device in lunar simulant AGK 2010 under various compaction conditions, as well as in dry quartz sand. *Source* Seweryn et al. (2014)

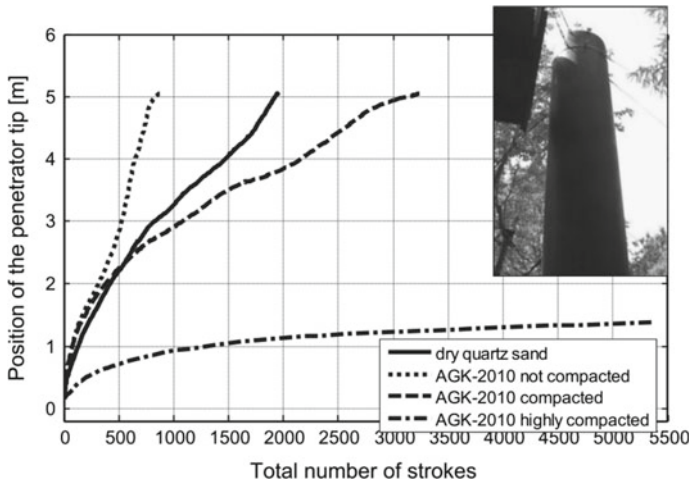


Fig. 15.35 A comparison of the depth of penetration to total number of strokes obtained using the KRET device in lunar simulant AGK 2010 under different compaction conditions, as well as in dry quartz sand. *Source* Seweryn et al. (2014)

of Mines. A specially designed penetrometer capable of making high-resolution measurements of force and depth (Dreyer et al. 2018) allowed for precise monitoring of not only the penetration behavior of the simulants, but the relaxation of the granular material after initial penetration as well. The relaxation behavior of penetrated granular materials (especially simulants) had not yet been studied and appears to be sensitive to geotechnical and environmental conditions that the penetration resistance is insensitive to.

Both dry and cryogenic testing of lunar regolith simulants utilized the ISRU Experimental Probe (IEP) (Fig. 15.36), a 6-mm diameter, 30° conical controlled-mechanism penetrometer capable of penetrating simulant samples to a depth of ~30 mm. Vertical motion was driven by a lead screw such that, when under load, the backlash (reactive movement) was negligible and continuous monitoring of the force at the probe tip using a mounted force gage allowed the relaxation of the simulant around the probe tip to be observed and measured. Cryogenic tests were performed under ambient and vacuum conditions at sample temperatures of ~110 K, while dry tests were performed at similar pressures and elevated temperatures typically approaching 323 K, both at a rate of 0.25 mm/s.

Atkinson et al. (2019) performed 24 penetration tests on dry GRC-3 and JSC-1A at low (~20%) and high (80%) relative densities at pressures of ~700 and ~0.05 Torr. Samples were prepared using both standard Proctor and vibratory compaction methods to relatively high degrees of density accuracy via specially designed compaction sleeves. Final sample containers measured 9.4 cm in diameter and 7.3 cm in depth, giving a CCR of ~16. The depth of penetration (30 mm) was

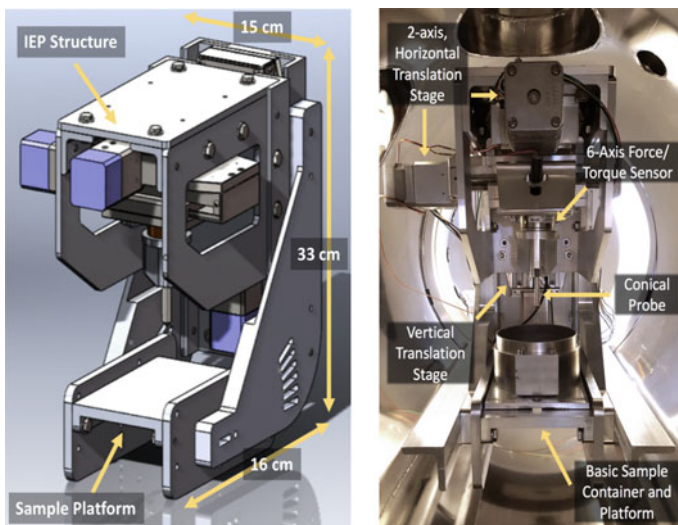


Fig. 15.36 The ISRU Experimental Probe (IEP). Left: CAD model. Right: IEP with a basic sample container inside a vacuum chamber. *Source* Dreyer et al. (2018)

less than half the container depth in order to minimize boundary effects from the sample base.

Dry penetration curves (Fig. 15.37) were fitted with a parabolic equation ($F(z) = \alpha z + \beta z^2$), as described in Sect. 15.3.2. Low-density samples (~20% relative density) showed a steady, nonlinear increase in resistance with depth—to ~5 N at 30-mm depth—while high-density samples (80% relative density) displayed a more dramatic nonlinear increase with depth to reach resistances of 20 N or more. The coefficient of the parabolic equation β shows high sensitivity to density and relative insensitivity to pressure (Fig. 15.37), is shown to be related to the model developed by Puech and Foray (2002), and gives moderate approximations of basic geotechnical properties such as bearing capacity and lateral slip line length. The linear penetration coefficient α , while displaying some correlation to geotechnical parameters, shows high volatility since higher-density samples demonstrate essentially parabolic penetration behavior with little to no linear component. Despite the high CCR, potential edge effects were identified at depths >20 mm by the onset of z^3 behavior.

Dry relaxation curves were fitted with a Maxwell-style rheological model (Sect. 5.3) whose parameters correspond to both the elastic (k_i) and viscous (τ_i) behavior in relaxing granular materials. Of particular importance is that the relaxation behavior showed significant sensitivity not only to sample density but to the simulant type and testing pressure (Fig. 15.38), predominantly the elastic coefficients k_i . Some sensitivities are non-unique as evidenced by the parameter k_1 , which decreases in value with both increasing density and pressure, in which case comparison to other parameters (such as β , sensitive only to density) can be used to determine the testing condition. Both elastic parameters k_e and k_1 are sensitive to simulant type, likely due to the increased angularity of JSC-1A inhibiting grain rotation and increasing the interlocking of grains leading to both a higher residual supported load (k_e) and a delayed onset of relaxation (k_1).

The increased cohesion in JSC-1A (Li et al. 2013) was also potentially observed in the relaxation behavior of both simulants to their coarse-grained (250–710 μm) and fine-grained (<75 μm) versions, sieved in house. Figure 15.39 shows that JSC-1A—with its full combination of all particle sizes in its original particle-size distribution—shows more cohesive behavior (as evidenced by the relaxation parameter k_e which relates to the external Hookean spring in Sect. 15.3.3) than either the fine or coarse versions. This phenomenon was predicted by Sanchez and Scheeres (2012) for application to the cohesive nature of rubble-pile asteroids, in which the cohesion is created through the interaction of many small particles with nearby larger ones, thus requiring a distribution of particle sizes for increased cohesion. GRC-3, a cohesionless sand, has a k_e at full particle-size distribution that appears to be a volumetric average of those of its fine and coarse versions.

The same models of nonlinear penetration and Maxwellian relaxation behavior were used to describe the results of similar penetration tests conducted on icy JSC-1A samples at cryogenic temperatures. A specially designed cryogenic cooling reservoir (Fig. 15.40) kept the sample at $\sim 110 \pm 20$ K while the probe was cooled to ~ 188 –211 K (as measured using an internally embedded Type-K thermocouple). The cryogenic cooler allowed for pre-cooling of the samples in LN_2 outside the

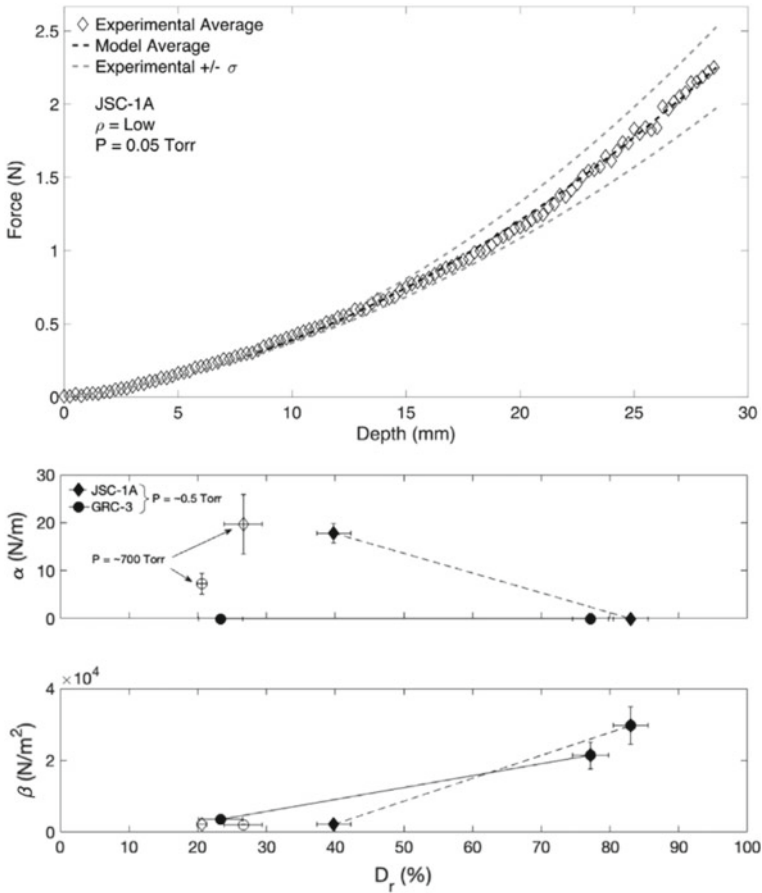


Fig. 15.37 Top: A sample penetration curve from a low-density JSC-1A tested at 0.05 Torr showing the experimental average (diamond markers) and one standard deviation (dashed gray line), with the associated model fit (dashed black line). Middle and Bottom: α (N/m) and β (N/m²) model values plotted against relative density (%) for all six full particle distribution tests. JSC-1A tests are represented as diamonds and GRC-3 as circles. Black markers are tested at ~0.5 Torr, white markers at ~700 Torr. Error bars are taken from the model fits to the standard deviation curves and thus represent experimental variability. *Source* Atkinson et al. (2019)

vacuum chamber before insertion and for free flow of LN₂ through the reservoir while in the vacuum to minimize heat gain and maximize cooling rate.

Twenty-four penetration tests into samples of JSC-1A at densities of 1.55–1.63 g/cm³ and levels of saturation from 0 to 12% (including one sample of pure water ice) using the same IEP device (Dreyer et al. 2018), modified to handle cryogenic temperatures, resulted in both penetration and relaxation curves that were strongly sensitive to ice content.

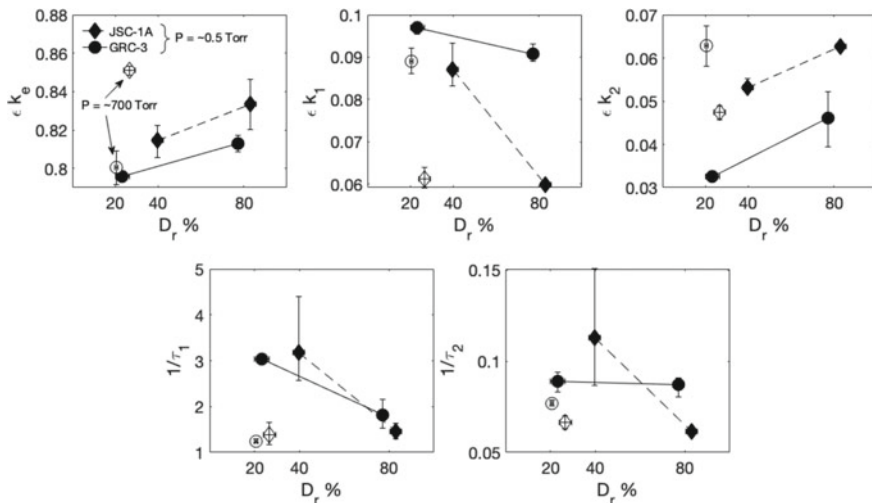


Fig. 15.38 Model parameter values versus relative density for all full particle distribution tests. JSC-1A tests are represented as diamonds and GRC-3 as circles. Black markers indicate testing at 0.05 Torr while open markers indicate testing at ~700 Torr. Error bars are taken from the model fits to the standard deviation curves and thus represent experimental variability. Elastic parameters are at the top while time-dependent parameters are at the bottom. k_1 is the only parameter capable of distinguishing simulant type, test pressure, and sample density. *Source* Atkinson et al. (2019)

Cryogenic penetration curves (Fig. 15.41) followed a parabolic increase for higher ice content samples ($\geq 1\%$) and linear responses for low ice content ($< 1\%$). Full penetration to ~30 mm was only possible for samples $< 1\%$ ice content, as the 120 N maximum threshold for the machine was achieved at shallow depths for higher ice contents. A step change in penetration resistance (similar to that seen by both Gertsch et al. 2006 and Pitcher et al. 2016) appears between 1 and 3% ice content. The phenomenon is explained by the authors to parallel percolation theory: it is the result of the progressive filling of pore space within the simulant to the point at which the pore ice (~35% content) can begin to act as an interconnected, load-bearing component, thus increasing the resistance to vertical force. Below this point, the majority of the load is still borne by the grains with cementing assistance from the distributed pore ice (Fig. 15.42). Since the phenomenon is dependent on the pore-size (and thus pore-volume) distribution, it would be expected to result in a step change at different percentages of ice content, which may explain the identification of such a step change at between 5 and 9% ice content in NU-LHT-2 M (Pitcher et al. 2016).

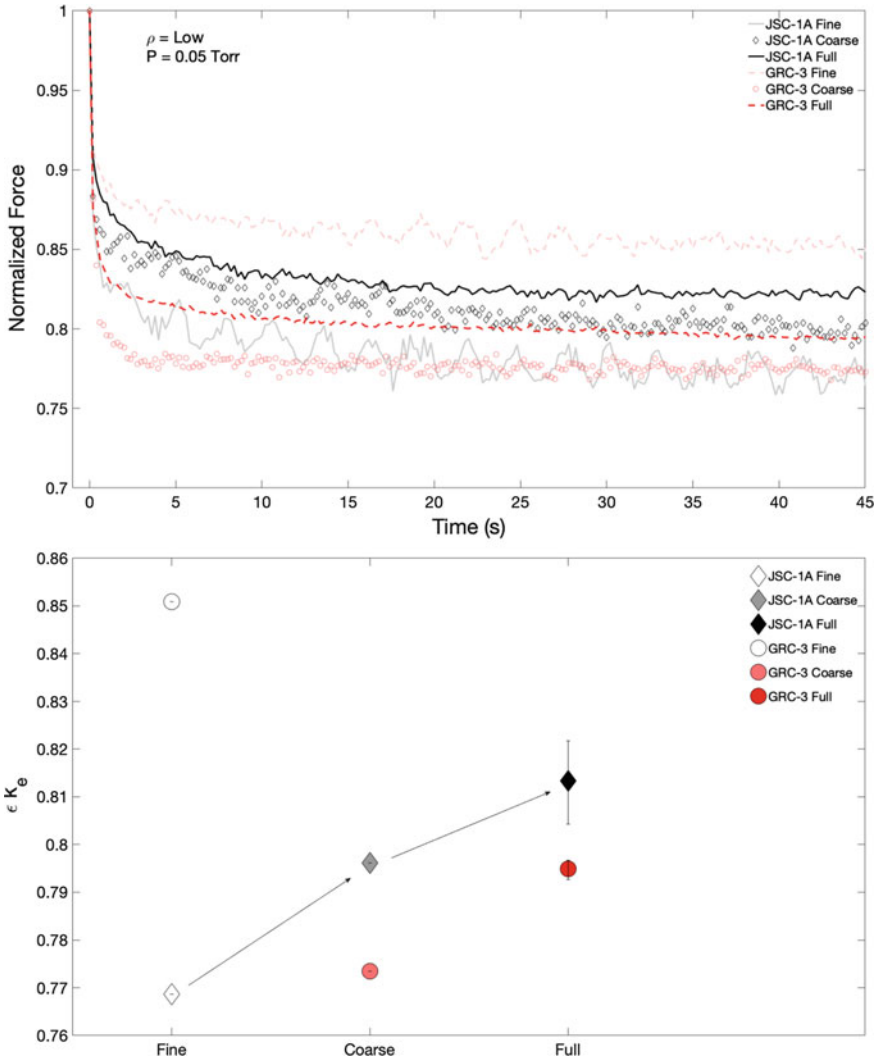


Fig. 15.39 Relaxation curves (note linear time) of fine, coarse, and full particle-size distribution JSC-1A and GRC-3. Bottom: Distribution of relaxation parameter k_e for the associated fine, coarse, and full JSC-1A (diamonds) and GRC-3 (circles) tests. *Source* Atkinson et al. (2019)

The model parameters for both penetration (α and β) and for relaxation (elastic parameters k_i) were sensitive to ice content (Figs. 15.43 and 15.44), with the relationships taking the forms

$$\alpha = 1162p_{ice} + 147.5p_{ice}^2, \quad (15.10)$$

$$\beta = 1E6p_{ice}^2, \quad (15.11)$$

$$k_e = 0.935(p_{ice}^{0.013}), \quad (15.12)$$

$$k_1 = 0.054(e^{-0.28p_{ice}}), \quad (15.13)$$

$$k_2 = 0.028(p_{ice}^{-0.096}), \quad (15.14)$$

where p_{ice} is the sample ice content expressed as a percentage.

Additionally, the relaxation behavior showed sensitivity to temperature, as evidenced by the increase in k_e value for dry (0%) JSC-1A tested at 110 K compared to that tested at 320 K in Atkinson et al. (2019). Consequently, this modified penetrometer has shown the potential for identifying simulant type, cohesion, density and

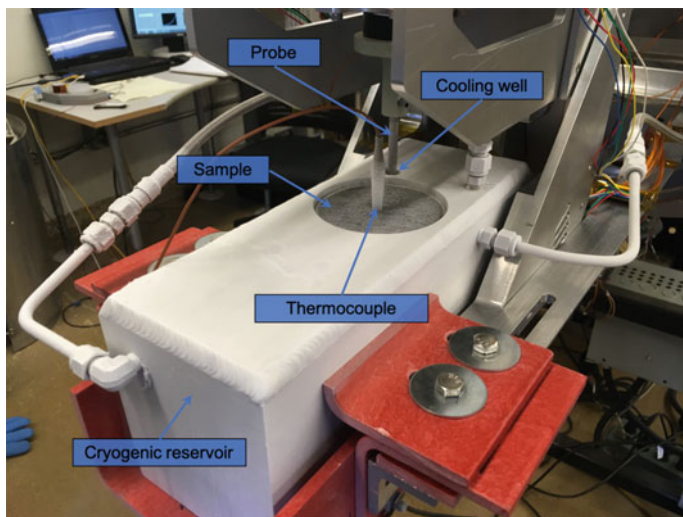


Fig. 15.40 The cryogenic cooler as seen connected into a specialized garolite cradle beneath the IEP penetrometer. Inlet and outlet hoses provide circulation of liquid nitrogen, as the cooler is hollow to create a cooling, stable reservoir surrounding the sample. The probe itself is seen pressed into the cooling well (at back), while a thermocouple is embedded in the sample center for temperature monitoring during thermal tests. *Source* Atkinson et al. (2019)

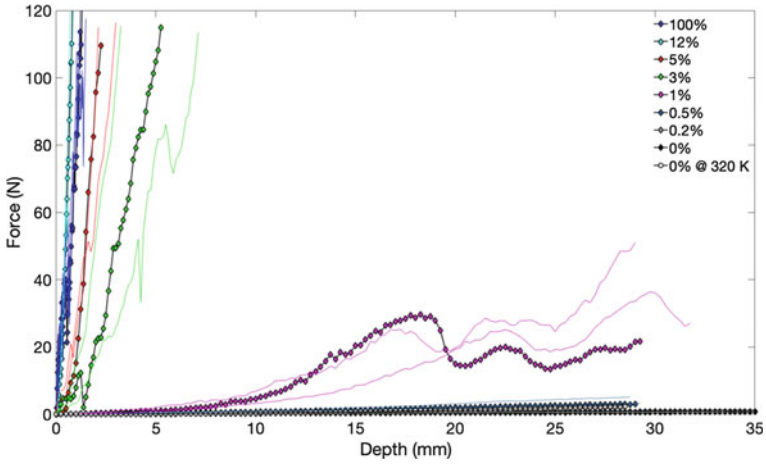


Fig. 15.41 Penetration curves for all ice contents showing resistance force as a function of probe depth, including a test of 0% at 320 K from Atkinson et al. (2019). All tests performed at ~110 K unless otherwise noted. Markers indicate one of three runs at the same ice content, while solid lines indicate the other two. *Source* Atkinson et al. (2019)

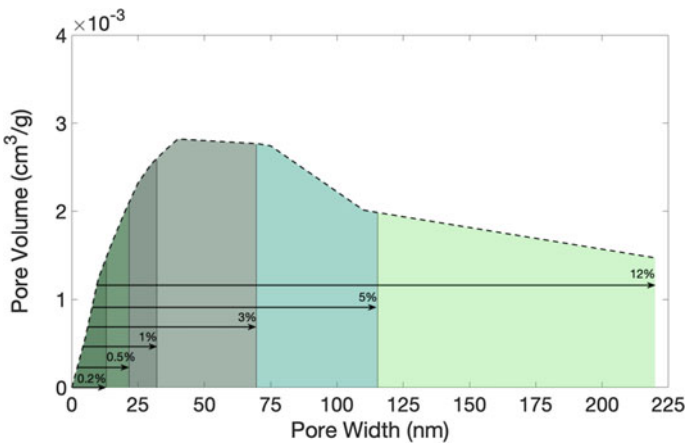


Fig. 15.42 Nitrogen adsorption analysis of JSC-1A, showing pore volume as a function of pore width. Arrows indicate the direction of pore filling (from smallest pores to largest) at each identified moisture content (%). Colors indicate the additional pore space and pore volume fluid filled at each associated moisture content. *From* Atkinson et al. (2020)

test pressure at ambient to elevated temperatures, and temperature and ice content at cryogenic conditions.

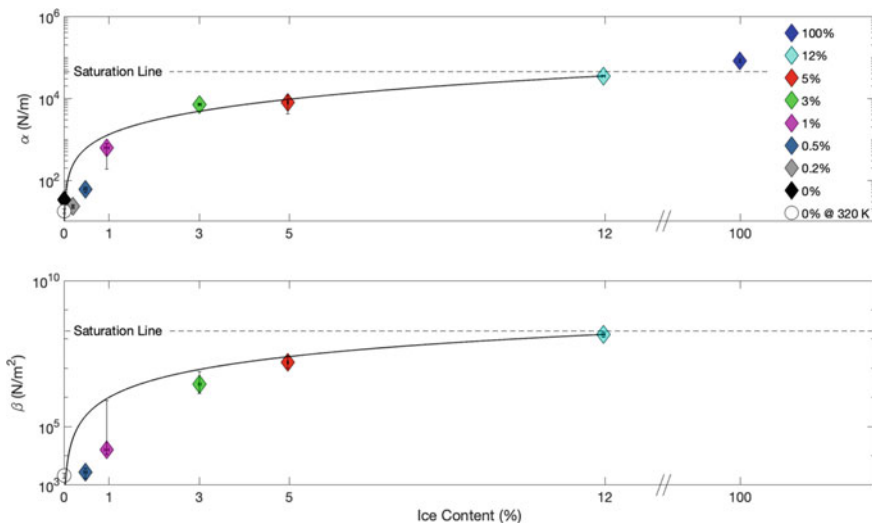


Fig. 15.43 Coefficients of the second-order model fits to penetration curves α (top) and β (bottom) as a function of ice content. Note the nonlinear increase in value for both coefficients to a saturated limit. β values at 0, 0.2, and 100% ice content are zero and therefore not displayed on the semi-log axis. All tests performed at ~ 110 K unless otherwise noted. Note that horizontal error bars indicating the range in ice content are present but extremely small. *Source* Atkinson et al. (2020)

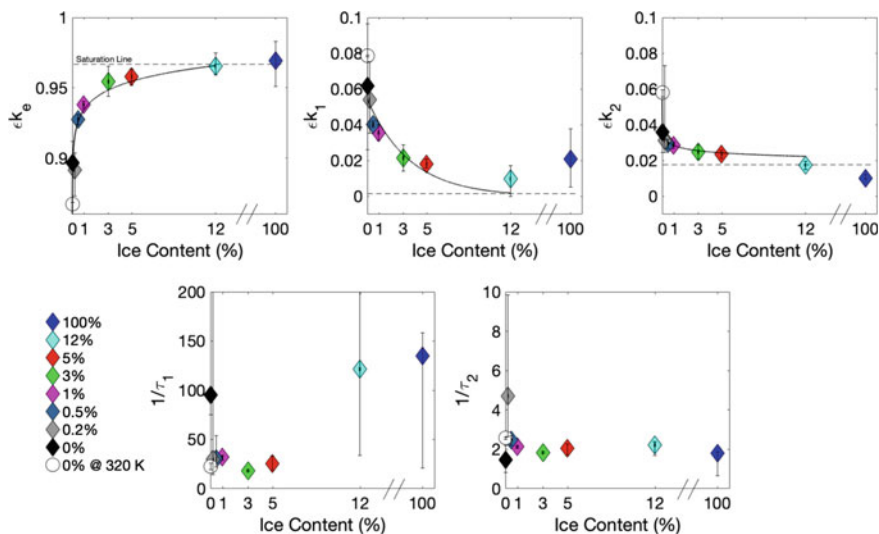


Fig. 15.44 Parameters of the two-arm Maxwell model fits to relaxation curves in the initial 1 s of relaxation as a function of ice content. Note the nonlinear increase in k_e and decrease in k_1 and k_2 to a saturated limit. All tests performed at ~ 110 K unless otherwise noted. *Source* Atkinson et al. (2020)

15.7.4 Synthesis

Penetrometer tests in lunar simulants, while relatively limited in number and performed in a wide range of environmental conditions using varied experimental methods, have nonetheless provided insight into the possible behavior and characteristics of both dry and icy lunar regolith.

Apollo-era researchers performed both penetrations into small amounts of returned lunar regolith and larger tests in more widely available lunar simulants and other granular materials. They determined that penetration resistance and its rate of change with depth are indicators of soil density and decrease with gravity. The research indicated generally linear increases in penetration resistance with depth, with minor evidence of nonlinearity.

More recent research has involved the use of both manual field and controlled-mechanism penetrometers into simulants containing water ice from zero to full saturation, and at temperatures and pressures approaching those expected in situ. Wide ranges in the experimental methods have led to difficulties in interpreting data in some cases, though in general there is ample evidence of nonlinear relationships between penetration resistance and depth (contrary to established analytical solutions). It should be noted, however, that the boundary effects induced by the bottom boundary of sample containers has not been quantified in sufficient detail.

Evidence also points to the sensitivity of the penetrometer to general soil properties like density, with a particular emphasis in this regard on the ice content of the simulants at low temperatures (reaching cryogenic in some cases). Increases in ice content correspond to nonlinear increases in penetration resistance, with large step changes in resistance potentially a result of the amount and distribution of ice in the granular matrix.

In addition to relative density and ice content, the relaxation of the granular material after penetration shows sensitivity to various bulk granular properties such as cohesion and grain angularity, as well as experimental conditions such as pressure and temperature.

The use of the penetrometer in exploring the characteristics and behavior of granular materials at a more detailed level than previously pursued should become a focus for researchers in the future. This simple device, robust enough to be deployed on the lunar surface and adaptable enough to perform highly technical tests in the laboratory, has unexplored potential both as a primary instrument in future missions and as a terrestrial investigative tool for more advanced soil research.

15.8 Permissions

Figure 15.1: Reprinted from M.B. Kirkham, *Principles of Soil and Plant Water Relations* (2014), 171–183 with permission from Elsevier.

Figure 15.2: Reproduced by permission of ERDC.

Figure 15.3: Approved for public release; distribution is unlimited.

Figure 15.4: Reprinted from Atkinson et al., Penetration and relaxation behavior of dry lunar regolith simulants (2019), *Icarus*, 328, 82–92 with permission from Elsevier.

Figures 15.5–15.7, 15.9: Reproduced by permission of the Lunar and Planetary Institute, Houston.

Figure 15.8: Reprinted from Ray et al., JSC-1A lunar soil simulant: Characterization, glass formation, and selected glass properties (2010), *Journal of Non-Crystalline Solids*, 356, 2369–2374 with permission from Elsevier.

Figure 15.10: Reprinted from Houston & Namiq, Penetration resistance of lunar soils (1971), *Journal of Terramechanics*, 8, 59–69 with permission from Elsevier.

Figures 15.11, 15.16–16.17, 15.26–15.27: Reproduced by permission of NASA.

Figures 15.12–15.14: Reproduced by permission of Heather Oravec.

Figure 15.15: Reproduced by permission of Honeybee Robotics.

Figures 15.18–15.20: Reproduced by permission of COSPAR.

Figures 15.21–15.22: Reproduced by permission of Harald Hellman.

Figures 15.23, 15.25: Reproduced from Gertsch et al., Effect of water ice content on excavatability of lunar regolith (2006) with permission from Missouri University of Science and Technology.

Figure 15.24: Reproduced from Gertsch et al., *Review of Lunar Regolith Properties for Design of Low Power Lunar Excavators* (2008) with permission from AIP Publishing.

Figures 15.28–15.32: Reproduced by permission of Mehmet Cil.

Figures 15.33–15.35: Reprinted from Seweryn et al., Determining the geotechnical properties of planetary regolith using Low Velocity Penetrometers (2014), *Planetary and Space Science*, 99, 70–83 with permission from Elsevier.

Figure 15.36: Reprinted from Dreyer et al., A new experimental capability for the study of regolith surface physical properties to support science, space exploration, and in situ resource utilization (ISRU) (2018), *Review of Scientific Instruments*, 89, with the permission of AIP Publishing.

Figures 15.37–15.39: Reprinted from, Atkinson et al., Penetration and relaxation behavior of dry lunar regolith simulants (2019), *Icarus*, 328, 82–92 with permission from Elsevier.

Figures 15.40–15.44: Reprinted from Atkinson et al., Penetration and relaxation behavior of JSC-1A lunar regolith simulant under cryogenic conditions (2020), *Icarus*, 346, 113,812 with permission from Elsevier.

References

- Alshibli, K.A., and A. Hasan. 2009. Strength properties of JSC-1A lunar regolith simulant. *J Geotech Geoenvironmental Engineering* 135 (5): 673–679. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000068](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000068).
- Atkinson, J., et al. 2019. Penetration and relaxation behavior of dry lunar regolith simulants. *Icarus* 328: 82–92. <https://doi.org/10.1016/j.icarus.2019.03.009>.
- Atkinson, J., et al. 2020. Penetration and relaxation behavior of JSC-1A lunar regolith simulant under cryogenic conditions. *Icarus* 346(113812). <https://doi.org/10.1016/j.icarus.2020.113812>.
- Bekker, M. G. 1969. *Introduction to Terrain—Vehicle Systems*, 1st ed. University of Michigan Press.
- Blok, C., et al. 2019. *Analytical methods used with soilless substrates*, 2nd ed. Soilless culture. Elsevier B.V. <https://doi.org/10.1016/b978-0-444-63696-6.00011-6>.
- Butlanska, J., et al. 2012. Multi-scale analysis of cone penetration test (CPT) in a virtual calibration chamber. *Canadian Geotechnical Journal* (December): 1–26. <https://doi.org/10.1139/Published>.
- Carrier, W.D.I., L. G. Bromwell, and R.T. Martin. 1972. Strength and compressibility of returned lunar soil. In *Proceedings of the 3rd lunar science conference*, 3223–3234. The MIT Press.
- Carrier, W.D.I., L.G. Bromwell, and R.T. Martin. 1973. Behavior of returned lunar soil in a vacuum. *Journal of the Soil Mechanics and Foundations Division* 99: 979–996.
- Carrier, W.D.I., G.R. Olhoeft, and W. Mendell. 1991. Physical properties of the lunar surface. In *Lunar sourcebook*, ed. G.H. Heiken, D.T. Vaniman, and B.M. French, 475–594. New York: Cambridge University Press.
- Cherkasov, I. I., and V. V. Shvarev. 1973. Soviet investigations of the mechanics of lunar soils. *Soil Mechanics and Foundation Engineering* 10 (4): 252–256. <https://doi.org/10.1007/BF01704945>.
- Cil, M. B. 2011. *Discrete element modeling of cone penetration in JSC-1A lunar regolith simulant*. Louisiana State University.
- Costes, N. C., et al. 1969. *Apollo 11 soil mechanics investigation*. Apollo 11 Preliminary Science Report.
- Costes, N. C., G. T. Cohron, and D. C. Moss. 1971. Cone penetration resistance test—an approach to evaluating in-place strength and packing characteristics of lunar soils. In *Lunar and planetary science conference proceedings*, 1973–1987.
- Costes, N. C., J. E. Farmer, E. B. George. 1972. *Mobility performance of the lunar roving vehicle: terrestrial studies—Apollo 15 results*. Washington, DC.
- Department of the Army Corps of Engineers Mississippi River Commission (1948) *Trafficability of soils: Laboratory tests to determine effects of moisture content and density variations*. Vicksburg, Mississippi.
- Dreyer, C.B., et al. 2018. A new experimental capability for the study of regolith surface physical properties to support science, space exploration, and in situ resource utilization (ISRU). *Review of Scientific Instruments* 89 (6): 064502. <https://doi.org/10.1063/1.5023112>.
- Durgunoglu, H. T., and J. K. Mitchell. 1973. *Static penetration resistance of soils*.
- ElShafie, A. 2012. Penetration forces for subsurface regolith probes. Theses and Dissertations, University of Arkansas, Fayetteville. <http://scholarworks.uark.edu/etd/574%0AThis>.
- ElShafie, A., and V. F. Chevrier. 2014. Reanalysis of the penetration data provided by Lunokhod rover on the moon. In *45th lunar and planetary science conference proceedings*, February, 31–33. <http://adsabs.harvard.edu/abs/2014LPL...45.1533E>.

- ElShafie, A., V.F. Chevrier, and N. Dennis. 2012. Application of planetary analog mechanical properties to subsurface geological investigations. *Planetary and Space Science. Elsevier* 73 (1): 224–232. <https://doi.org/10.1016/j.pss.2012.09.001>.
- Gertsch, L., R. Gustafson, and R. Gertsch. 2006. Effect of water ice content on excavatability of lunar regolith. *Faculty Research & Creative Works* 10 (1063/1): 2169290.
- Gertsch, L., Rostami, J., and R. Gustafson. 2008. Review of lunar regolith properties for design of low power lunar excavators. In *6th international conference on case histories in geotechnical engineering*. Arlington, VA, 1–15.
- Green, A. J., and K.-J. Melzer. 1971. *Performance of boeing LRV wheels in a lunar soil simulant*.
- Gromov, V.V. 1998. Physical and mechanical properties of lunar and planetary soils. *Earth, Moon, and Planets* 80 (1/3): 51–72. <https://doi.org/10.1023/A:1006353510204>.
- Harr, M.E. 1977. *Mechanics of particulate media*. McGraw-Hill.
- He, C., X. Zeng, and A. Wilkinson. 2011. Geotechnical properties of GRC-3 lunar simulant. *Journal of Aerospace Engineering* 421 (July): 123. [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0000162](https://doi.org/10.1061/(ASCE)AS.1943-5525.0000162).
- Heiken, G. H., D. T. Vaniman, and B. M. French. (eds.). 1991. *Lunar sourcebook: A user's guide to the moon*. Cambridge University Press.
- Houston, W.N., and L.I. Namiq. 1971. Penetration resistance of lunar soils. *Journal of Terramechanics* 8 (1): 59–69. [https://doi.org/10.1016/0022-4898\(71\)90076-0](https://doi.org/10.1016/0022-4898(71)90076-0).
- Jaffe, L.D. 1971. Bearing strength of lunar soil. *The Moon* 3: 337–345.
- Jiang, M., et al. 2017. 'DEM simulation of soil-tool interaction under extraterrestrial environmental effects. *Journal of Terramechanics: ISTVS* 71: 1–13. <https://doi.org/10.1016/j.jterra.2017.01.002>.
- Jiang, M.J., H. Yu, and D. Harris. 2006. Discrete element modelling of deep penetration in granular soils. *International Journal for Numerical and Analytical Methods in Geomechanics* 30: 335–361. <https://doi.org/10.1002/nag.473>.
- Johnson, J. B. 2003. A statistical micromechanical theory of cone penetration in granular materials cold regions research and engineering laboratory. CRREL Report, February.
- Johnson, S. W., M. C. Koon, and W. D. I. Carrier. 1995. Lunar soil mechanics. *Journal of British Interplanetary Society*, 43–48.
- Kirkham, M. B. 2014 *Penetrometers, principles of soil and plant water relations*. <https://doi.org/10.1016/b978-0-12-420022-7.00011-2>.
- Kleinhenz, J. E., and R. A. Wilkinson. 2012. ISRU Soil mechanics vacuum facility: Soil bin preparation and simulant strength characterization. In *AIAA aerospace sciences meeting*, January, 1–8.
- Kleinhenz, J. E., and R. A. Wilkinson. 2014. *Development and testing of an ISRU soil mechanics vacuum test facility*. Cleveland, Ohio.
- Klosky, J.L., et al. 2000. Geotechnical behavior of JSC-1 lunar soil simulant. *Journal of Aerospace Engineering* 13 (4): 133–138. [https://doi.org/10.1061/\(ASCE\)0893-1321\(2000\)13:4\(133\)](https://doi.org/10.1061/(ASCE)0893-1321(2000)13:4(133)).
- Kochan, H., et al. 1989. Crustal strength of different model comet materials. In *Physics and mechanics of cometary materials*.
- Kuhn, M.R. 1987. *Micromechanical aspects of soil creep*. Berkeley: University of California.
- Lacerda, W. A., and W. N. Houston. 1973. Stress relaxation in soils. In *Proceedings of the eighth annual conference on soil mechanics and foundation engineering*, Moscow, 221–227.
- Leonovich, A.K., et al. 1971. Studies of lunar ground mechanical properties with the self-propelled Lunokhod-1. In *Peredvizhnaya Laboratoriya na Luna-Lunokhod-1*, 120–135. Moscow: Nauka.
- Leonovich, A. K., et al. 1974. The main peculiarities of the processes of the deformation and destruction of lunar soil. In *The Soviet American conference on cosmochemistry of the moon and planets*. NASA SP-370.
- Leonovich, A.K., et al. 1975. *Luna 16 and 20 investigations of the physical and mechanical properties of lunar soil*. Berlin: Akademie-Verlag.
- Leonovich, A. K., et al. 1976. Investigation of the physical and mechanical properties of the lunar sample brought by Luna-20 and along the route of motion of Lunokhod 2. In *Proceedings of*

- the XXIVth international astronomical congress*, Balu, USSR, 321–332. <https://doi.org/10.1016/B978-0-08-020365-2.50031-7>.
- Li, Y., X. Zeng, and A. Wilkinson. 2013. Measurement of small cohesion of JSC-1A lunar Simulant. *Journal of Aerospace Engineering* 26 (October): 882–886. [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0000197](https://doi.org/10.1061/(ASCE)AS.1943-5525.0000197).
- Liingaard, M., A. Augustesen, and P.V. Lade. 2004. Characterization of models for time-dependent behavior of soils. *International Journal of Geomechanics* 4 (3): 157–177. [https://doi.org/10.1061/\(ASCE\)1532-3641\(2004\)4:3\(157\)](https://doi.org/10.1061/(ASCE)1532-3641(2004)4:3(157)).
- Lunne, T., P.K. Robertson, and J.J.M. Powell. 1997. *Cone penetration testing in geotechnical practice*. New York: Spon Press.
- Mantovani, J. G., G. M. Galloway, and K. Zacny. 2016. Low force penetration of icy regolith. In *ASCE earth and space conference*, Orlando, FL.
- McKay, D. S., et al. 1993. JSC-1: A new lunar regolith simulant. *Lunar and Planetary, XXIV*: 963–964.
- Meyerhof, G. G. 1957. The ultimate bearing capacity of foundations on slopes. In *Proceedings of 4th ICSMFE*, London, England, 384–386.
- Meyerhof, G.G. 1976. Bearing capacity and settlement of pile foundations. *Journal of the Geotechnical Engineering Division* 102 (3): 197–228.
- Mitchell, J.K., et al. 1972. Mechanical properties of lunar soil: Density, porosity, cohesion, and angle of internal friction. *Proceedings of the Third Lunar Science Conference 3*: 3235–3253.
- Mitchell, J. K., and S. Kenichi. 2005. Time effects on strength and deformation. *Fundamentals of Soil Behavior*, pp. 465–522.
- Mulqueen, J., J.V. Stafford, and D.W. Tanner. 1977. Evaluation of penetrometers for measuring soil strength. *Journal of Terramechanics* 14 (3): 137–151.
- Muthukumarappan, K., and G. J. Swamy. 2017. Rheology, microstructure, and functionality of cheese. In *Advances in food rheology and its applications*, ed. J. Ahmed, P. Ptaszek, and S. Basu, 245–276 Woodhead Publishing Series in Food Science, Technology and Nutrition. <https://doi.org/10.1016/B978-0-08-100431-9.00010-3>.
- Oravec, H. A. 2009. Understanding mechanical behavior of lunar soils for the study of vehicle mobility, p. 662. [https://doi.org/10.1061/40988\(323\)12](https://doi.org/10.1061/40988(323)12).
- Oravec, H.A., X. Zeng, and V.M. Asnani. 2010. Design and characterization of GRC-1: A soil for lunar terramechanics testing in Earth-ambient conditions. *Journal of Terramechanics. ISTVS* 47 (6): 361–377. <https://doi.org/10.1016/j.jterra.2010.04.006>.
- Peleg, M., and M.D. Normand. 1983. Comparison of two methods for stress relaxation data presentation of solid foods. *Rheologica Acta* 22 (1): 108–113. <https://doi.org/10.1007/BF01679835>.
- Perkins, S.W. 1991. *Modeling of regolith structure in extraterrestrial constructed facilities*. Boulder: University of Colorado.
- Perkins, S.W., and C.R. Madson. 1996. Mechanical and load-settlement characteristics of two lunar soil simulants. *Journal of Aerospace Engineering* 9 (1): 1–9.
- Pitcher, C., et al. 2016. Investigation of the properties of icy lunar polar regolith simulants. *Advances in Space Research* 57 (5): 1197–1208. <https://doi.org/10.1016/j.asr.2015.12.030>.
- Puech, A., and P. Foray. 2002. Refined model for interpreting shallow penetration CPTs in sands. In *Proceedings of the annual offshore technology conference*, 2441–2449. <https://doi.org/10.4043/14275-MS>.
- Rao, V.N.M., D.D. Hamann, and J.R. Hammerle. 1975. Stress analysis of a viscoelastic sphere subjected to temperature and moisture gradients. *Journal of Agricultural Engineering Research* 20 (3): 283–293. [https://doi.org/10.1016/0021-8634\(75\)90064-5](https://doi.org/10.1016/0021-8634(75)90064-5).
- Ray, C. S., et al. 2010. JSC-1A lunar soil simulant: Characterization, glass formation, and selected glass properties. *Journal of Non-Crystalline Solids*. Elsevier B.V., 356(44–49): 2369–2374. <https://doi.org/10.1016/j.jnoncrysol.2010.04.049>.
- Rohani, B., and G. Y. Baladi. 1981 *Correlation of mobility cone index with fundamental engineering properties of soil*.

- Roylance, D. 2001. Engineering viscoelasticity, pp. 1–37.
- Sanchez, P., and D. J. Scheeres. 2012. Cohesion in “Rubble-Pile” asteroids’. In *Asteroids, comets, meteors*.
- Sanglerat, G. 1972. History of the penetrometer. In *The penetrometer and soil exploration: interpretation of penetration diagrams—Theory and practice*. Elsevier, pp. 1–91. <https://doi.org/10.1016/B978-0-444-40976-8.50008-1>.
- Scott, R.F. 1987. Failure. *Geotechnique* 37: 423–466. <https://doi.org/10.1680/geot.1987.37.4.423>.
- Seweryn, K., et al. 2014. Determining the geotechnical properties of planetary regolith using low velocity penetrometers. *Planetary and Space Science* 99: 70–83. <https://doi.org/10.1016/j.pss.2014.05.004>.
- Sibille, L., et al. 2006. Lunar regolith simulant materials: Recommendations for standardization, production, and usage. In *NASA technical paper*, September 2006, 142. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20060051776.pdf>.
- Snieder, R., C. Sens-Schönfelder, and R. Wu. 2017. The time dependence of rock healing as a universal relaxation process, a tutorial. *Geophysical Journal International* 208 (1): 1–9. <https://doi.org/10.1093/gji/ggw377>.
- Stoeser, D., D. Rickman, and S. Wilson. 2010 *Design and specifications for the highland regolith prototype simulants NU-LHT-1M and -2M*.
- Teale, R. 1965. The concept of specific energy in rock drilling. *International Journal of Rock Mechanics and Mining Science* 2 (1): 57–73.
- Varney, A., et al. 2001. Prediction and interpretation of pore pressure dissipation for a tapered piezoprobe. *Géotechnique* 51 (7): 601–617. <https://doi.org/10.1680/geot.2001.51.7.601>.
- Vesic, A.S. 1972. *Expansion of cavities in infinite soil mass*. Durham, NC: Duke University.
- Willman, B.M., et al. 1995. Properties of lunar soil simulant JSC-1. *Journal of Aerospace Engineering* 8 (2): 77–87.
- Wong, J.Y. 1989. *Terramechanics and off-road vehicles*. Amsterdam: Elsevier.
- Yu, H.S., and J.K. Mitchell. 1998. Analysis of cone resistance: Review of methods. *Journal of Geotechnical and Geoenvironmental Engineering* 124 (2): 140–149. [https://doi.org/10.1061/\(ASCE\)1090-0241\(1998\)124:2\(140\)](https://doi.org/10.1061/(ASCE)1090-0241(1998)124:2(140)).
- Zacny, K.A., et al. 2010. Robotic lunar geotechnical tool. *Earth and Space* 2010: 166–181.
- Zeng, X., et al. 2010a. Geotechnical properties of JSC-1A lunar soil simulant. *Journal of Aerospace Engineering* 23 (April): 111–116. [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0000014](https://doi.org/10.1061/(ASCE)AS.1943-5525.0000014).
- Zeng, X., C. He, and A. Wilkinson. 2010b. Geotechnical properties of NU-LHT-2M lunar highland simulant. *Journal of Aerospace Engineering* 23 (4): 213–218.

Part IV
Mars

Chapter 16

Ice Resource Mapping on Mars



Nathaniel E. Putzig, Gareth A. Morgan, Hanna G. Sizemore, David M. Hollibaugh Baker, Eric I. Petersen, Asmin V. Pathare, Colin M. Dundas, Ali M. Bramson, Samuel W. Courville, Matthew R. Perry, Stefano Nerozzi, Zachary M. Bain, Rachael H. Hoover, Bruce A. Campbell, Marco Mastrogiuseppe, Michael T. Mellon, Roberto Seu, and Isaac B. Smith

Abstract This chapter explains the rationale for considering shallowly buried (0 to >5 m depth) water ice in the mid-latitudes of Mars as a resource to support future human missions, and describes a NASA-funded effort to map that ice with existing orbital remote-sensing data. In recent decades, numerous studies have used various datasets to investigate the presence and stability of water ice in the Martian shallow

Putzig and Morgan are joint first authors for this work.

N. E. Putzig (✉) · M. R. Perry · Z. M. Bain
Planetary Science Institute, 405 Urban St, Suite 300, Lakewood, CO 80228, USA
e-mail: nathaniel@putzig.com

M. R. Perry
e-mail: mperry@psi.edu

Z. M. Bain
e-mail: zbain@psi.edu

G. A. Morgan (✉) · H. G. Sizemore · A. V. Pathare · S. W. Courville · I. B. Smith
Planetary Science Institute, 1700 E Fort Lowell, Suite 106, Tucson, AZ 85719, USA
e-mail: gmorgan@psi.edu

H. G. Sizemore
e-mail: sizemore@psi.edu

A. V. Pathare
e-mail: pathare@psi.edu

S. W. Courville
e-mail: swcourville@psi.edu

I. B. Smith
e-mail: ibsmith@psi.edu; ibsmith@yorku.ca

D. M. Hollibaugh Baker
NASA Goddard Space Flight Center, Mail Code: 690.1, 8800 Greenbelt Rd., Greenbelt,
MD 20771, USA
e-mail: david.m.hollibaughbaker@nasa.gov

E. I. Petersen
Geophysical Institute, University of Alaska—Fairbanks, 2156 N Koyukuk Dr, Fairbanks,
AK 99775, USA
e-mail: eipetersen@alaska.edu

This is a U.S. government work and not under copyright protection in the U.S.; foreign
copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_16

subsurface, with the aim of understanding the planet's recent climate history. As part of a renewed effort to prepare for human Mars missions, NASA has undertaken a more resource-focused approach. Here we describe the Mars Subsurface Water Ice Mapping (SWIM) team's efforts to characterize the distribution of buried water-ice resources across all longitudes from 60°S to 60°N latitude through the integration of multiple datasets. Deriving composite measures for the presence of accessible ice from a diverse range of remote sensing techniques with unique resolutions and caveats is a challenging problem. To enable data synthesis, the team developed a methodology that assigns values of ice consistency for mapped detections of hydrogen from

C. M. Dundas

U.S. Geological Survey, Astrogeology Science Center, 2255 N. Gemini Drive, Flagstaff, AZ 86001, USA

e-mail: cdundas@usgs.gov

A. M. Bramson

Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, 550, Stadium Mall Dr., West Lafayette, IN 47907, USA

e-mail: BramsonA@purdue.edu

S. W. Courville

School of Earth and Space Exploration, Arizona State University, 781 Terrace Mall, Tempe, AZ 85287, USA

S. Nerozzi

Lunar and Planetary Laboratory, University of Arizona, 1629 E University Blvd, Tucson, AZ 85721, USA

e-mail: nerozzi@arizona.edu

R. H. Hoover

Southwest Research Institute, 1050 Walnut St, Suite 300, Boulder, CO 80302, USA

e-mail: rhoover@boulder.swri.edu

B. A. Campbell

National Air and Space Museum, Smithsonian Institution, 655 Jefferson Drive SW, Washington, DC 20560, USA

e-mail: campbellb@si.edu

M. Mastrogiuseppe · R. Seu

Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Roma, RM, Italy

e-mail: marco.mastrogiuseppe@uniroma1.it

R. Seu

e-mail: roberto.seu@uniroma1.it

M. T. Mellon

Cornell Center for Astrophysics and Planetary Science, 104 Space Sciences, Building, Ithaca, NY 14853, USA

e-mail: mellon@astro.cornell.edu

I. B. Smith

Lassonde School of Engineering, York University, 101 Petrie Science and Engineering Building, 4700 Keele St., Toronto, ON M3J 1P3, Canada

a neutron spectrometer, thermal behavior from various thermal spectrometers, multi-scale geomorphology from imagery and elevation data, and surface and subsurface echoes from a radar sounder. Faced with diverse sensing depths and footprints for these datasets, the team has been pursuing an optimal approach to best represent multi-dataset ice consistency. The current formulation includes the use of weighting factors tuned to depth zones of interest for resource extraction. In the absence of dedicated ground-truth data, the validity of the team's efforts is assessed by comparing the maps to the locations of fresh, ice-exposing impacts. The highest ice-consistency values occur within discrete zones poleward of $\sim 40^\circ$ latitude, where ice is relatively shallow, but positive values extend well into the $\sim 20^\circ$ – 30° latitude zone, which is preferable for landing sites due to engineering considerations.

16.1 Ice as a Critical Resource for Human Missions

16.1.1 *The Resource Value of Ice on Mars*

In the effort to deliver humans to the surface of Mars and return them safely to Earth, current propulsion technology means that mass represents the ultimate premium for cost. Thus, any such endeavor is made much more feasible by leveraging all available in-situ resources. The most valuable Martian resource for “living off the land” is water, which, when combined with atmospheric carbon dioxide, can provide methane as a fuel to sustain an outpost and for the return to Earth (Ash et al. 1978; Zubrin et al. 1991). Water also represents one of the most important ingredients of life support, including as a source of oxygen for breathing.

Mars has plentiful surface water ice, with multi-kilometer-thick ice caps in the form of the North and South polar layered deposits (NPLD/SPLD) and widespread shallow (<1 m depth) ground ice in polar and subpolar regions. However, these sources of water are at latitudes that are not feasible for the initial human missions to the Red Planet. The higher solar radiation and corresponding manageable thermal environment and length of night offered by the lower latitudes are critical to mission success. In addition, low latitudes reduce the energy needed for landing and launch from the surface for the return trip to Earth. Lower elevations are also desired, as they provide more atmosphere to slow down a spacecraft prior to a propulsive landing. Thus, locating the lowest-elevation, lowest-latitude sites that have significant water-ice deposits is a key consideration in selecting future human landing sites on Mars.

16.1.2 *Ice Stability on Mars*

Numerical modeling of the stability of ground ice on Mars dates to the 1960s, when Leighton and Murray (1966) found that subsurface ice should be present in the higher latitudes, based on what was known at the time about Martian surface properties and temperature variations. Since that time, progressive gains in our knowledge of these properties have been made, largely due to a series of ever-more-capable spacecraft in orbit and on the surface of Mars. The arrival of Mars Global Surveyor (MGS) and its Thermal Emission Spectrometer (TES) at Mars in the late 1990s (Christensen et al. 2001) precipitated a significant improvement in the accuracy and timespan of observations of temperatures and their variation over Martian seasons. Analyses of those variations allowed researchers to more accurately ascertain the stability and presence of current-day ice beneath a veneer of soil in high-latitude zones that extend equatorwards to about 50° latitude in each hemisphere (e.g., Mellon et al. 2004; Putzig and Mellon 2007).

A few years into the MGS mission, the 2001 Mars Odyssey spacecraft arrived with the Mars Orbiter Neutron Spectrometer (MONS) onboard, allowing direct detection of hydrogen within the upper half-meter of the subsurface. While lower concentrations of hydrogen likely indicate only hydrated minerals, in places where the fraction of hydrogen detected exceeds about 25%, the only plausible explanation is water ice (e.g., Feldman et al. 2002; Pathare et al. 2018). Thus, the neutron data provided confirmation of ground ice at high latitudes.

Subsequently, images from the High Resolution Imaging Science Experiment (HiRISE) camera and Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) onboard the Mars Reconnaissance Orbiter (MRO) that arrived at Mars in 2006 have confirmed the presence of shallow, high-purity ground ice, not only at those higher latitudes but also extending down to a latitude of 39°N, via the discovery of new impact craters that expose and excavate ice within a meter of the surface (Byrne et al. 2009; Dundas et al. 2014, 2021). While the depths and latitudinal profiles observed by MONS and in the icy craters are broadly consistent with the general expectations based on theoretical models of stability, there are excursions of shallow ice extending closer to the equator than expected under current climate conditions. The existence of ice out of equilibrium with the current climate suggests that the preservation of ice deposited in the mid-latitudes during previous climatic episodes has occurred. Mid-latitude remnant ice therefore represents an important resource for future missions.

16.2 The Mars Subsurface Water Ice Mapping (SWIM) Project

16.2.1 Project Overview

In 2015, NASA held a workshop in Houston, Texas, to begin consideration of landing sites for human missions to the surface of Mars. Workshop participants were instructed to evaluate 100-km-diameter exploration zones that would encompass the actual spacecraft landing site, human habitation facilities, features of scientific interest, and resources for generating fuel to sustain on-ground activities and supply an Earth-return vehicle. One outcome of the workshop was the realization that water, in the form of either hydrated minerals or buried water ice, represents an especially critical resource that had yet to be identified and assessed at the level needed to fully support landing-site planning. To address this need, NASA held a gathering of Mars scientists studying ice and hydrated minerals during the fall 2016 American Geophysical Union meeting in San Francisco to discuss the extent to which current data could help address knowledge gaps concerning water resources on Mars. The group identified a number of datasets and methods that would be useful for mapping hydrated minerals and buried mid-latitude ice, and NASA used this information to produce a request for proposals, issued in June 2017. After reviews, NASA selected and funded four proposals, two focused on the mapping of hydrated minerals and two on the mapping of buried water ice. The latter studies were limited to a swath of $\sim 10^\circ$ longitude in the Arcadia Planitia region, extending from the equator to 60°N , as NASA intended to evaluate the success of the ice-mapping methods proposed prior to broadening the study region (Fig. 16.1). These pilot studies, which began in late 2017 and early 2018, aimed to assess ice presence and map its distribution across the swath. Putzig et al. (2017) proposed a combination of thermal data and modeling with radar observations of subsurface interfaces, whereas Morgan and Campbell (2017) proposed a combination of geomorphological data and radar observations of near-surface properties.

Realizing the highly complementary nature of the two pilot studies, the teams proposed during a joint interim review at NASA headquarters in July 2018 to merge their investigations in a larger study area. NASA agreed to the merger and to an expanded study area in a project extension entitled Subsurface Water Ice Mapping (SWIM) in the Northern Hemisphere of Mars (Fig. 16.1), which began in September 2018. For this first ‘2019’ phase of the Mars SWIM project, we added additional techniques, including the use of neutron spectrometer data (Sect. 16.2.2.1), and we expanded the combined team to meet an increased workload under a seven-month timeline.

Our team began developing a way of presenting a coherent view of ice presence as informed by the collection of data and techniques, which we termed “ice consistency.” For each technique, we came up with a means to evaluate how consistent each dataset at a given location is with either the absence or the presence of ice, assigning values between -1 (wholly inconsistent with ice) and $+1$ (wholly

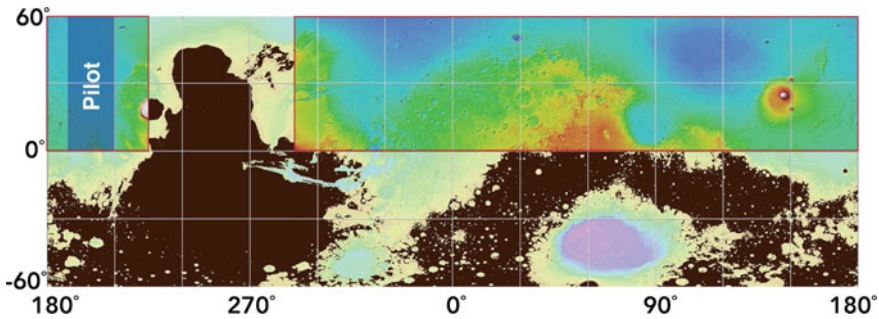


Fig. 16.1 Map of the SWIM study areas overlain on Martian topography. Blue box at 0–60°N, 190–210°E is the pilot study region. Red boxes in the northern hemisphere (–70–225°E) delineate the 2019 study region. The remainder of the map is the 2020 study region, subject to a cut-off at +1 km elevation (mask highlighted in black). MOLA basemap credit: NASA/JPL/USGS

consistent with ice). To present an overall assessment of ice consistency with all relevant data, we introduced the SWIM equation, in which the technique-specific ice-consistency values are combined into a composite value for each map pixel. In the first SWIM phase, we chose to use a simple averaging, with each technique’s ice consistency weighted equally with the others (Sect. 16.2.3). We presented results from the first study phase in a series of presentations at the 2019 Lunar and Planetary Science Conference in The Woodlands, Texas, and in a *Nature Astronomy* publication by Morgan et al. (2021).

Upon completion of the first SWIM study phase, NASA asked us to propose an extension to further refine our methods and expand the study to include all other areas below +1 km elevation and equatorward of 60° latitude in both hemispheres (Fig. 16.1). In this second ‘2020’ phase of the Mars SWIM project, we made substantial refinements to our techniques, including the incorporation of a new thermal dataset (Sect. 16.2.2.2) and finer-resolution geomorphic mapping (Sect. 16.2.2.3). In addition, we replaced the single SWIM equation with a set of equations for three distinct depth zones, weighting the terms in each equation by the sensing depths of the respective techniques. Realizing that this means of producing composite assessments of ice has some limitations, such as not allowing a quantitative treatment of uncertainties associated with each technique, we embarked on an exploration of an alternative presentation of our composite results through the use of Bayesian statistics (Sect. 16.2.3.2). Among other capabilities, this alternative approach allowed the team to consider ice presence in terms of probability, tunable for given depths or targeted concentrations of ice.

16.2.2 *Spacecraft Datasets and Processing Techniques*

16.2.2.1 Neutron Spectrometer Data

Our ice-consistency values derived from neutron spectroscopy are based on MONS observations of fast, thermal, and epithermal neutron fluxes at the top of the Martian atmosphere, which are highly sensitive to the presence of hydrogen (and, by extension, H₂O ice) within the upper half-meter of the Martian subsurface (Feldman et al. 2002). We employed the global two-layer water equivalent hydrogen (WEH) maps of Pathare et al. (2018), who refined the crossover approach of Feldman et al. (2011) to calculate the WEH abundances of an upper layer of weight fraction W_{up} with thickness D overlying a semi-infinite lower layer of weight fraction W_{dn} . For the SWIM project, we expressed the positive and negative ranges of neutron ice consistency C_N as linear functions of lower-layer WEH abundance, W_{dn} :

$$C_N = \mathbf{1} \ (W_{\text{dn}} \geq 25\%) \text{ to } \mathbf{0} \ (W_{\text{dn}} = 10\%)$$

$$C_N = \mathbf{0} \ (W_{\text{dn}} = 10\%) \text{ to } -\mathbf{1} \ (W_{\text{dn}} \leq 5\%)$$

Maximum ($C_N = 1$) values were so assigned because $W_{\text{dn}} \geq 25\%$ corresponds to excess ice (i.e., the mass fraction needed to saturate the pore volume) for the surface density and porosity assumptions of Pathare et al. (2018). The other benchmarks ($C_N = 0$ at $W_{\text{dn}} = 10\%$, and $C_N = -1$ at $W_{\text{dn}} \leq 5\%$) were chosen based on our qualitative assessment of the likelihood of near-surface ice in all terrains exhibiting such W_{dn} values (see Fig. 11b of Pathare et al. 2018).

16.2.2.2 Thermal Datasets

Our thermal analysis combines three derived data products: two maps of surface-layer thermal inertia and ice-table depth produced independently by Bandfield and Feldman (2008) and Piqueux et al. (2019) as well as a global map of subsurface layering developed within the SWIM project using the methodology of Putzig and Mellon (2007). These three datasets (designated BF08, PQ19, and SP20, hereinafter) are built on the same core physics principle. Planetary surfaces consisting of different materials and different combinations of materials exhibit distinctive seasonal and diurnal variations in surface temperature and apparent thermal inertia (ATI) (Putzig and Mellon 2007). Information about layering structure in the upper meter of the Martian subsurface can therefore be derived by forward modeling the thermal behavior of a suite of material layering scenarios and comparing those model results to spacecraft observations of surface temperature or derived ATI. This section summarizes our methods and results, and a more detailed description is provided by Sizemore et al. (2020).

BF08, PQ19, and SP20 all employ this basic strategy to search for and map shallowly buried ice, but differ in terms of procedural details. Broadly, BF08 and

SP20 apply somewhat different analysis pathways to the same spacecraft dataset, which is bolometric and surface temperatures from MGS TES, whereas BF08 and PQ19 apply nearly identical analyses to two different datasets, temperatures from MGS TES and the MRO Mars Climate Sounder (MCS), respectively.

Briefly, the BF08 and PQ19 approach is as follows. At each map pixel, a point in the orbit of Mars defined by a particular areocentric longitude of the Sun (L_S) designated L_S^* is determined where modeled surface temperatures become the least sensitive to any putative buried ice. Surface-layer thermal inertia (TI) is then derived at L_S^* using a standard look-up table. Next, a suite of two-layer forward models is run at each pixel to predict the seasonal temperature variation resulting from ice occurring at a range of depths beneath a dry soil layer with thermal properties consistent with the previously derived TI. Spacecraft-observed surface temperatures are compared directly to modeled seasonal surface temperatures. The best match at each pixel is selected based on minimizing relative standard deviation (RSD), allowing the derivation of ice-table depth at locations where a match is identified (for details see Piqueux et al. 2019; Bandfield and Feldman 2008; and references therein).

A key characteristic of the BF08 and PQ19 approach is that all forward models include a lower layer with high thermal inertia, consistent with buried ice. Both BF08 and PQ19 used the KRC software package (Kieffer 2013) for forward modeling. BF08 employed an older version that did not include temperature-dependent thermophysical properties, whereas PQ19 employed a recent version that includes temperature dependence.

The SP20 ice-mapping process is based on the Putzig and Mellon (2007) approach of identifying a wide range of material heterogeneity types on the Martian surface. It uses global TES ATI maps computed at 10° intervals of L_S spanning a full Martian seasonal cycle. At each pixel of these maps, ATI is derived from TES bolometric temperatures using a look-up table of single-layer thermal-model results. A suite of two-layer thermal models is also run and used to derive ATI as a function of season via the same pathway used in the production of the global ATI maps. The suite of forward models includes layering scenarios with seven combinations of four material types defined by their thermal properties: dust over sand (D/S), dust over duricrust (D/C), dust over rock (D/R), sand over rock (S/R), duricrust over dust (C/D), duricrust over sand (C/S), and rock over sand (R/S). Because the thermal properties of rock and ice are effectively equivalent, D/R and S/R scenarios are consistent with subsurface ice. We also consider D/C consistent with subsurface ice that does not completely fill the soil pore space, based on theoretical (Mellon et al. 1997; Piqueux and Christensen 2009) and empirical (Siegler et al. 2012) models describing the increase of soil thermal conductivity with progressive cementation. A range of upper-layer thicknesses are allowed for each material combination in our two-layer model suite. The seasonal variation at each pixel of the global maps, $ATI(L_S)$, is compared to the suite of forward models. Matches are identified based on minimizing RSD, and ice-table depth can be derived at locations of D/R, S/R, and D/C matches.

Our SP20 analysis used MARSTHERM (Putzig et al. 2013) for ATI derivation and forward modeling. This thermal model, which was originally developed for TES thermal inertia derivation (Mellon et al. 2000), includes a relatively sophisticated 13-layer one-dimensional radiative-convective atmospheric model (as compared to KRC’s single-layer radiative model). It does not include temperature dependence of thermal properties for surface and subsurface materials.

In the second SWIM phase, we modified the heterogeneity matching scheme from that of previous mapping efforts by Putzig and Mellon (2007), Putzig et al. (2014), and Morgan et al. (2021) to improve sensitivity to buried ice and derivation of ice-table depths. Specifically, we removed the shallowest depth nodes from the two-layer model suite to eliminate matches to models with extremely thin (<1 mm) upper-layer duricrust, as we deemed these models to be physically implausible. We also updated the algorithm used for interpolating between discrete depth models to improve its accuracy.

BF08, PQ19, and SP20 all identified layering ostensibly consistent with ice at low latitudes. BF08 attributed this effect to atmospheric heat transport via Hadley circulation that is not accounted for in one-dimensional thermal models. They truncated their ice-table maps equatorward of 50° in both hemispheres to compensate. PQ19 developed a more sophisticated (but still ad-hoc) algorithm for latitudinally truncating their ice depth map where ice-table depths became shallower with decreasing latitude. They applied this algorithm equatorward of 50°N and 60°S , and they also removed outliers equatorward of 35° latitude.

In the SP20 analysis, we used a similar algorithm to truncate points where apparent ice detections were shallowing towards the equator in the BF08 and SP20 maps. We generated a filtered version of the SP20 map using a 9×9 -pixel low-pass filter. Before filtering, we filled gaps in model matches by assuming that the empty pixel had the same ice depth as the next valid poleward pixel. We employed an equatorward-marching algorithm starting at 50° latitude in each hemisphere. If the local filtered pixel had an ice depth shallower than the median for that of the next five poleward pixels, we treated that pixel and all others equatorward of it as non-detections. We also applied a cut-off at 35° latitude to remove minor outliers. To facilitate comparisons, we applied a similar filter to the BF08 map, but using a 3×3 -pixel filter and three poleward pixels due to the lower spatial resolution. Our approach produced reasonable agreement between the SP20 map and predictive ice stability maps by Mellon et al. (2004) in both hemispheres, and good agreement between the equatorward extent of ice in the SP20 and PQ19 maps of the northern hemisphere.

For all three maps, we computed thermal ice consistency C_T at each pixel where ice was modeled as a function of the predicted depth d to the ice table using the following scheme:

For $d < 30$ cm, $C_T = 1$ (high consistency with ice)

For $30 \text{ cm} \leq d \leq 50 \text{ cm}$, C_T decreases linearly from **1** to **0**

For all $d > 50$ cm, $C_T = 0$ (data gives no information about the presence of ice)

We based this scheme on a suite of 1-, 2-, and 3-layer thermal models which show that it is increasingly difficult to distinguish icy and non-icy subsurfaces as ice-table depth increases beyond 30 cm (e.g., Fig. 1 of PQ19). We gave the three thermal ice-consistency maps (C_{BF08} , C_{PQ19} , and C_{SP20}) equal weighting in the combined map:

$$C_{\text{T}} = (C_{\text{BF08}} + C_{\text{PQ19}} + C_{\text{SP20}})/3$$

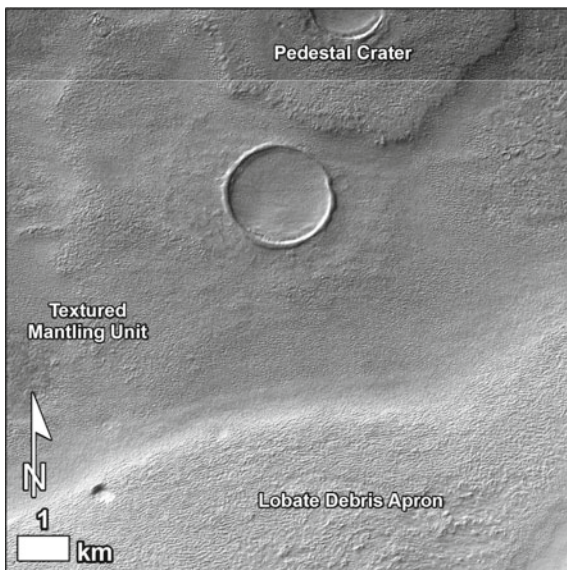
Our choice to equally weight the three thermal ice-consistency maps does not indicate a preference for the TES data over the MCS data. Rather, it reflects the complexities and differing assumptions in the generation of all three maps from their constituent input data sources. A priori, none of the dataset–methodology combinations is to be preferred. In the absence of ground truth (at multi-pixel scales), there is no clear rationale for weighting the maps differently.

16.2.2.3 Geomorphology

We determined our geomorphology ice-consistency values from previous and new mapping of periglacial and glacial features. Our new mapping included a grid-mapping survey of eight groups of landforms and terrain types that were inferred by previous workers to indicate the presence of ice (Fig. 16.2). These features included: mantle (Mustard et al. 2001; Milliken et al. 2003; Dundas et al. 2018), sublimation-type pits and textured terrain (Carr 2001; Mangold 2005; Kostama et al. 2006), scalloped terrain (Morgenstern et al. 2007), viscous flow features (VFF) (i.e., lobate debris aprons (LDA), lineated valley fill (LVF), concentric crater fill (CCF), and small-scale glacier-like forms (GLF)) (Milliken et al. 2003; Head et al. 2010; Souness and Hubbard 2012; Levy et al. 2014), pedestal craters (Kadish et al. 2009), expanded craters (Viola et al. 2015), ring-mold craters (e.g., Kress and Head 2008; Baker and Carter 2019), and terraced craters (Bramson et al. 2015). We based our identification of landforms and descriptions on previous work, including recent grid-mapping efforts (see Ramsdale et al. (2017) and references therein). Maps produced by previous work were also used as input, including pedestal craters (Kadish et al. 2009; Viola and McEwen 2018), LDA/LVF/CCF (Levy et al. 2014), expanded craters (Viola and McEwen 2018), and scalloped terrain (Viola and McEwen 2018).

In the first SWIM phase (mapping in the northern hemisphere between 0–225°E and 290–360°E longitude), we used a modified grid-mapping (Ramsdale et al. 2017) approach. Using a sampling of 4° × 4° MRO Context Camera (CTX) image mosaics within previously mapped geological units (Tanaka et al. 2005), we tallied the number of observed periglacial and glacial landforms and extrapolated the observations to the mapped unit boundaries. The CTX mosaics we used are at 5 m/pixel resolution and are beta01 versions available from the Bruce Murray Laboratory for Planetary Visualization via <http://murray-lab.caltech.edu/CTX/index.html> (Dickson et al. 2018). We refined this approach in the second SWIM phase for our southern hemisphere mapping, surveying periglacial and glacial landforms using the CTX mosaic within

Fig. 16.2 Example of landforms mapped as part of the geomorphology analysis. In this example, the following features were tallied: textured terrain, mantle, LDA and pedestal crater. Image data, CTX mosaic centered at 44.07°S, 107.45°E



1° × 1° grid cells between 24 and 38°S and within 4° × 4° cells elsewhere. For the 4° × 4° cells between 0 and 24°S where we found positive identifications of glacial/periglacial landforms, we carried out additional grid mapping at the 1° × 1° resolution.

Using binary values (1 = present, 0 = absent), we marked observed landforms (from the list of eight above) regardless of their abundance within each grid cell. To compute geomorphology ice consistency, C_G , we tallied the number (n) of periglacial features (pf) identified within a given grid and weighted them based on their likelihood of containing ice. For most features, we assigned a weight of 0.1 per feature to indicate some ambiguity in their ice content. However, we chose higher weights for several features to reflect the high ice content required to maintain such features, including scalloped terrain (0.75), pedestal craters (0.75), and VFF (1). If the nominal C_G total for a given pixel was >1, then we capped the pixel value at +1.

$$C_G = \min \left[\left(\begin{array}{l} 0.1 * n_{pf1} + 0.1 * n_{pf2} + 0.1 * n_{pf3} + \dots + 0.75 * n_{scalloped} \\ + 0.75 * n_{pedestal} + 1 * n_{VFF} \end{array} \right), 1 \right]$$

Final geomorphology ice-consistency values thus ranged from $C_G = 0$, i.e., no evidence of ice from periglacial/glacial features, to $C_G = 1$, i.e., highest evidence of ice from periglacial/glacial features. Note that values of $C_G < 0$ were not included in the geomorphology ice-consistency formulation since we did not specifically mark features that would be inconsistent with ice, for example exposed bedrock. In addition, to enable a layered approach, we split the geomorphology ice consistency into shallow (≤ 5 m) and deep (> 5 m) components, wherein mantled and textured terrains are limited to the shallow component C_{Gs} and terraced craters are limited to the deep

component C_{GD} . We included all mapped periglacial and glacial landforms in the shallow term due to the possibility that ice exists at depths ≤ 5 m for each of those landforms.

16.2.2.4 Radar Surface Power Analysis

The Shallow Radar (SHARAD) sounder onboard MRO was designed to search for subsurface structures by identifying radar-reflective interfaces ranging in depth from tens of meters to 1 km (or more) within the Martian crust (Seu et al. 2007). However, the radar echoes that are nominally from the surface also contain important information about materials within the upper ~ 5 m of the subsurface. The strength of the surface return is governed by multiple factors that include observational conditions (e.g., orientation of the spacecraft, state of the ionosphere), surface roughness, regional slope, and Fresnel reflectivity. The last of these can be estimated from the data by accounting for the others, and this technique provides a measure of near-surface density. Due to the low density of water ice relative to other geological materials, measuring reflectivity is suitable for searching for near-surface ice-rich deposits.

Broadly following a methodology first attempted with data from the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) on Mars Express (Mouginot et al. 2010) while accounting for the higher frequency of SHARAD, we derived a process to isolate Fresnel reflectivity from the SHARAD dataset for the first SWIM phase (Morgan et al. 2021).

This process applies the following steps: (1) limit the ionosphere effects by excluding all daytime tracks; (2) normalize surface power for the effects of surface roughness using the SHARAD roughness parameter developed by Campbell et al. (2013) (Fig. 16.3); (3) correct the loss of power due to regional slope using the median slope value over a Fresnel zone (3 km) as derived from Mars Orbiter Laser Altimeter (MOLA) data; and (4) take the median value of all the corrected SHARAD returns sampled within a given region to account for MRO influences (spacecraft roll, solar-panel configuration) (see Grima et al. 2012).

After isolating an approximation of reflectivity from the MARSIS surface power measurements, Mouginot et al. (2010) took the additional step of converting their results to values of the real relative dielectric permittivity (ϵ'). The ϵ' of a given geological substrate controls the speed of radar signals as they pass through the subsurface. Measurements of ϵ' can therefore be compared against laboratory and field measurements to constrain bulk subsurface compositions. In terms of the geological materials that comprise the Martian subsurface, ϵ' is related to the density and porosity of those materials, and thus ice exhibits a much lower permittivity relative to basaltic lava flows. The radar subsurface analysis described in Sect. 16.2.2.5 estimates the bulk ϵ' of the subsurface (to depths >15 m) to search for low permittivity values that could indicate the presence of buried ice.

For the radar surface analysis, we did not attempt to convert the corrected power returns to estimates of ϵ' . To achieve that would require calibration of the corrected

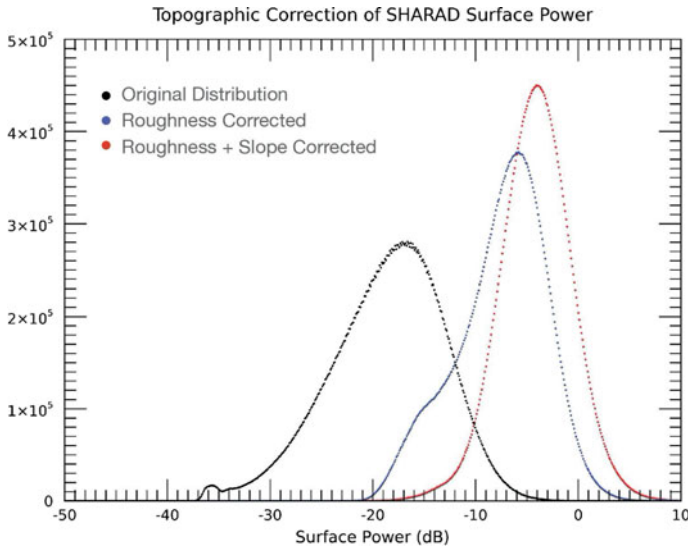


Fig. 16.3 Corrections of the SHARAD surface return (60°S – 60°N) to isolate Fresnel reflectivity. The roughness correction (blue distribution) applied the Campbell et al. (2013) roughness parameter. The slope correction (red distribution) applied MOLA-gridded data

power by comparing our measurements against terrains of a known permittivity. Mouginot et al. (2010) used the north and south polar layered deposits (NPLD and SPLD) as a reference for nearly pure water ice ($\epsilon' = 3.1$, see Plaut et al. 2007; Grima et al. 2009). However, the shallow near-surface sampling of the SHARAD surface echo (~ 5 m) makes this approach problematic. In the case of the NPLD, the fine layering of the shallow subsurface can cause positive and negative interference of the SHARAD return, whereas the surface of the SPLD is covered in a dust layer of variable thickness. As a consequence, the resulting SHARAD surface power returned over the polar caps is not equivalent to that of near-pure ice and likely drifts from region to region in response to layering changes and variations in dust content.

In lieu of estimating values of ϵ' , we opted to use the global distribution of corrected power values as a means to explore relative near-surface density. To test our approach, we compared the relationship between surface power and roughness exhibited by distinct Martian terrain types. Our analysis revealed that different units display similar curves, albeit with an offset in power. For example, high-density terrains such as young volcanics were found to sit higher on the power axis relative to the icy polar caps (Fig. 16.4). To apply such an analysis across the SWIM study region, we translated the global power distribution such that surface power $< -1\sigma$ maps to radar surface ice consistency $C_{\text{RS}} = +1$, surface power $> 1\sigma$ maps to $C_{\text{RS}} = -1$, and surface power values of -0.5σ , 0 , and 0.5σ were mapped to $C_{\text{RS}} = +0.5$, 0 , and -0.5 , respectively.

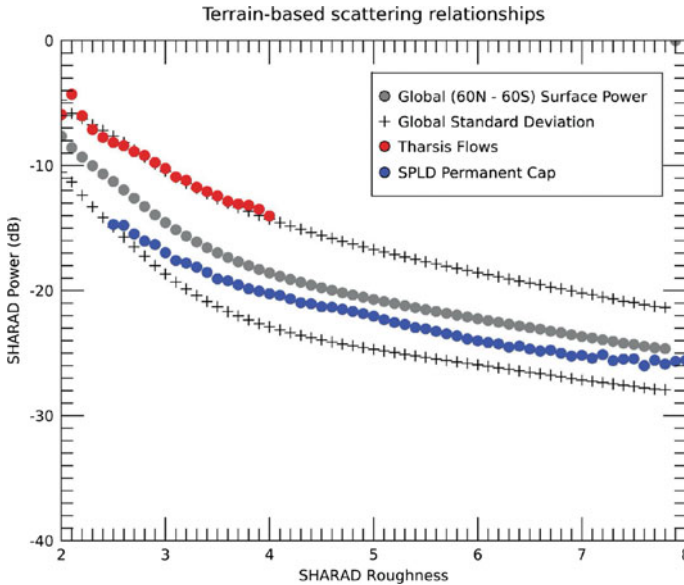


Fig. 16.4 Different Martian terrain types (CO₂ of the SPLD permanent cap in blue and basaltic substrate within Tharsis in red) exhibit similar radar scattering relationships (slopes in this figure) but with distinct offsets in power. We attribute the power offset to differences of Fresnel reflectivity, where the higher-power terrain corresponds to a denser substrate relative to that of the lower-power terrain

16.2.2.5 Radar Subsurface Structure and Composition Analysis

We incorporated an analysis of subsurface radar returns observed by SHARAD into the SWIM project through an extensive mapping and interpretation effort undertaken by five members of the team who specialize in planetary radar sounding. For subsurface analysis, SHARAD data from a particular observing run along an MRO orbit segment are presented as images of returned radar power, with along-track distance on the horizontal axis and either delay time or depth on the vertical axis (a display format known as a radargram; see Fig. 16.5).

We visually inspected all available SHARAD radargrams in the SWIM survey regions to identify candidate subsurface signals. To assess the nature of candidate subsurface signals, we compared the radargrams to cluttergrams, which simulate the radar echoes produced by the Martian surface topography. Cluttergrams allow one to determine if the candidates are true subsurface signals or merely an artifact of off-nadir surface reflections returned to the spacecraft at delay times similar to those of potential subsurface returns (Choudhary et al. 2016). Upon determining that signals are likely to be true subsurface returns (i.e., not at the same locations and delay times as features in MOLA-derived cluttergrams), we then further analyzed them to estimate a value for the real relative dielectric permittivity (ϵ') of the subsurface materials.

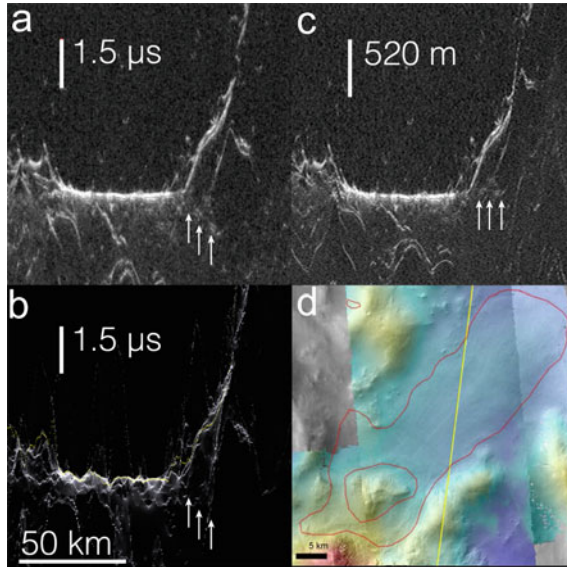


Fig. 16.5 SHARAD observation 35218-01 over an LDA in the southern hemisphere region of Argyre. **a** Delay-time radargram, with candidate subsurface signal indicated by white arrows. **b** Clutter simulation using a Mars Express High Resolution Stereo Camera digital terrain model, showing no predicted clutter at the delay time of the candidate signal. **c** Depth-corrected radargram using $\epsilon' = 3$; reflector has aligned with the surrounding plains. **d** Context image of radar ground track (yellow line) with topography and mapped extent of LDA (red line). After Berman et al. (2021)

As is explained in Sect. 16.2.2.4, ϵ' controls the speed of radar signals as they pass through a geological material, and thus places constraints on the subsurface composition. When the two-way delay time Δt is measured between SHARAD reflections bounding a geological unit of known thickness h , ϵ' can be calculated as $\epsilon' = (c \Delta t / 2 h)^2$, where c = speed of light. Δt can be measured readily using SHARAD radargrams, but it is often more difficult to estimate h .

We estimated h using different methods appropriate for different types of geological units. To enable quick application to many observations while producing reliable results in the aggregate, we optimized our techniques for each type of unit. For mantling units on the plains, we defined the base of the unit as a straight line between the plains elevation observed on either side of the unit in the radargram, with h being the depth to that line from the top of the unit. For geologic units that lie atop the plains but abut scarps and highlands on one side (such as for lobate debris aprons), we defined the base as a flat or fixed-slope interface continuous with the plains elevations observed on one side (following the methods of Petersen et al. 2018). In some special cases, we used stereo images to identify layer thickness in terraced craters, fossae, or other outcrops to provide point estimates of h . An important caveat is that the estimate of ϵ' is only as good as the estimate of h and the assumption that that thickness corresponds to the location of the interface producing the radar reflection.

Pure water ice has $\varepsilon' \sim 3$ (Ulaby et al. 1986), whereas basaltic materials typical of Martian bedrock have higher values of $\varepsilon' \sim 6\text{--}12$ (Campbell and Ulrichs 1969). Mixtures produce ε' values that are intermediate between those of the individual materials. Mixing models (e.g., Sihvola 1999; Stillman et al. 2010; Brouet et al. 2019) can predict ε' for mixed materials, but there remains an issue of non-uniqueness when translating from ε' to a specific composition.

Nevertheless, we used this knowledge of ε' values typical for our target ice deposits and Martian geology to formulate the following equation for radar subsurface (deep) ice consistency C_{RD} :

$$\begin{aligned} C_{\text{RD}} &= +1 \text{ where } \varepsilon' \leq 3 \\ C_{\text{RD}} &= 1/2(5 - \varepsilon') \text{ where } 3 \leq \varepsilon' \leq 7 \quad (C_{\text{RD}} = 0 \text{ where } \varepsilon' = 5) \\ C_{\text{RD}} &= -1 \text{ where } \varepsilon' \geq 7 \end{aligned}$$

In this framework, an ε' value of 3 is representative of pure ice, an ε' value of 7 is representative of pure basalt bedrock (no ice), and an intermediate ε' value is representative of an ice–rock mixture. While it is possible that materials such as high-porosity volcanic ash deposits could have a value of $\varepsilon' = 3$ while containing no ice, it is the combination of C_{RD} with ice-consistency values from the other instruments and techniques that help to disambiguate such materials.

16.2.3 Composite Ice Consistency from Data Integration

The greatest challenge to the SWIM project is synthesizing the diverse ice characterization techniques into congruent map products. As is outlined in Sect. 16.2.2, each of the five techniques provide contrasting perspectives on the physical properties of the subsurface, probe different depths, and exhibit separate caveats. The complexity of the problem is further compounded by the lack of ground-truth data that is sufficiently widespread to calibrate the individual techniques. For similar terrestrial problems, the integration of remote-sensing products is typically facilitated by the use of calibration maps produced through fieldwork. Remotely sensed datasets can then be calibrated via the ground-truth data. While a few very geographically limited areas, such as the Phoenix landing site (Smith et al. 2008) and fresh ice-exposing impact craters (Byrne et al. 2009; Dundas et al. 2014, 2021) provide actual and effective ground truth of buried ice on Mars, they are too few and far between to enable the production of calibration maps at the scales that would be required for the SWIM study's remote-sensing datasets.

Below, we describe two different approaches to integrating our ice characterization techniques. The first represents the evolution of the SWIM equation derived for the first SWIM phase in the northern hemisphere (Morgan et al. 2021). The second introduces a Bayesian framework to synthesize the probabilistic range of ice content measured by each technique.

16.2.3.1 The SWIM Equations

The driving motivation behind our integration efforts is to track agreements between our diverse ice characterization techniques. Within this framework, ice consistencies derived from multiple data sources are valued above that from a single methodology. There is clearly a higher potential for ice in areas where more datasets are consistent with the presence of ice and fewer are inconsistent with ice. The SWIM equations were developed to rank locations across Mars based on both the number of ice signatures and the relative strength of those signatures. For the first SWIM phase, we applied a straightforward approach that averaged the five ice-consistency values into a single value of composite ice consistency (C_i):

$$C_i = (C_N + C_T + C_G + C_{RS} + C_{RD})/5 \quad (16.1)$$

As is the case for the individual techniques, C_i can in principle range from -1 (all techniques are fully inconsistent with the presence of ice) to $+1$ (all techniques are fully consistent with the presence of ice). A value of 0 reflects no data or a balanced ambivalence among the ice characterization techniques. However, in our current implementation, ice consistency for our thermal and geomorphological techniques is restricted to the range of $0-1$, and thus the minimum C_i value is limited $-3/5$.

To interpret our C_i results for landing-site planning, we consider a value corresponding to $1/(\text{no. of techniques}) = 0.2$ as the minimum threshold for areas of interest that are likely to contain ice in quantities viable as a resource. The rationale behind this threshold assignment is that either one technique must post the maximum ice-consistency value while the others are inconclusive, or multiple datasets must provide positive indications of ice. Following this same logic, we consider values of $C_i > 0.6$ to be of the highest significance, as this would indicate a majority of techniques are strongly supportive of ice.

The known locations of mid-latitude ice, as exposed by impacts (Byrne et al. 2009; Dundas et al. 2014) and along scarps (Dundas et al. 2018), are not statistically sufficient to warrant calibration of the individual terms. Consequently, our first approach was to treat all of the terms within the SWIM equation equally. Nevertheless, the ice-exposing impacts do provide an opportunity to assess the effectiveness of our mapping approach. For example, the SWIM region from the first phase of mapping (Fig. 16.1) contains 13 ice-exposing impact sites. The average C_i value for these sites is 0.26 ± 0.16 (Morgan et al. 2021), which is above our minimum threshold, lending confidence to our methodology. Another motivation for the absence of weighting in Eq. (16.1) for the first SWIM phase was to encourage community engagement and permit SWIM product users to experiment with different formulations to produce their own integrated maps.

During the second SWIM phase, we continued to develop our data integration techniques, encouraged by community feedback that suggested we examine alternative approaches. As the datasets behind the various techniques probe different depths into the subsurface, we chose to leverage those depth sensitivities to produce

three new equations that correspond to three depth zones: <1 m, which is dominated by neutron and thermal spectrometer data; 1–5 m, which is dominated by shallow-geomorphic and radar surface-return data; and >5 m, which is dominated by deeper geomorphic data and radar subsurface dielectric permittivity estimations. To enable this layered approach, we divided the mapped geomorphic landforms into two groups corresponding to ice presence above and below a depth of 5 m, as described in Sect. 16.2.2.3. In addition, we introduce weighting factors s_M , i.e., the shallowness of method M, which is determined by dividing the depth of interest for each equation by the sensing depth of each method. Thus, for the first depth zone at <1 m, we have:

$$\begin{aligned} C_i[< 1 \text{ m}] &= (s_N C_N + s_T C_T + s_{GS} C_{GS} + s_{RS} C_{RS}) / (s_N + s_T + s_{GS} + s_{RS}) \\ &= (C_N + C_T + 0.2 * C_{GS} + 0.2 * C_{RS}) / 2.4 \end{aligned} \quad (16.2)$$

where s_N and s_T are both set equal to 1 because the neutron and thermal sensing depths of ~ 1 m are entirely within the zone of interest, whereas s_{GS} and s_{RS} are both set equal to 0.2 because only 20% of the shallow-geomorphic and radar surface sensing depths of ~ 5 m extend into the zone of interest. The deep radar consistency C_{RD} does not appear in Eq. (16.2) because the *minimum* sensing depth (taken as the SHARAD range resolution) of ~ 15 m translates to an s_{RD} for this zone of $1/15$ that makes any C_{RD} term negligible.

For the second depth zone at 1–5 m, we have:

$$\begin{aligned} C_i[1-5 \text{ m}] &= (s_{GS} C_{GS} + s_{RS} C_{RS} + s_{RD} C_{RD}) / (s_{GS} + s_{RS} + s_{RD}) \\ &= (C_{GS} + C_{RS} + 0.3 * C_{RD}) / 2.3 \end{aligned} \quad (16.3)$$

where s_{GS} and s_{RS} are both set equal to 1 because the shallow-geomorphic and radar surface sensing depths of ~ 5 m are entirely within the zone of interest, whereas s_{RD} is set equal to 0.3 because only 30% of a nominal radar subsurface sensing depth (taken as the SHARAD range resolution) of ~ 15 m extends into the zone of interest.

For the third depth zone at >5 m, we have:

$$\begin{aligned} C_i[> 5 \text{ m}] &= (s_{GD} C_{GD} + s_{RD} C_{RD}) / (s_{GD} + s_{RD}) \\ &= (C_{GD} + C_{RD}) / 2.0 \end{aligned} \quad (16.4)$$

where s_{GD} and s_{RD} are both set equal to 1 because the deep-geomorphic and radar subsurface sensing depths extend indefinitely into the zone of interest.

Our formulation of the SWIM equations is by no means the only approach possible, and other methods may be brought to bear, depending on the different goals of end users of the ice-consistency mapping products.

16.2.3.2 Bayesian Statistical Analysis

The SWIM equation approach to mapping ice consistency has two limitations: the values of ice consistency do not indicate ice concentration and there is no formal way to estimate uncertainty. A potential solution to these shortcomings is the use of Bayesian inversion. At the core of the SWIM project is an inverse problem in which we have a set of satellite remote-sensing data and we want to know the Martian subsurface composition that produces the measurements and whether that composition includes water ice. Bayesian inversion, as described by Tarantola (2005), is a well-developed procedure to solve inverse problems. Our application of the Bayesian procedure is based on the ability to represent the likelihood of all possible subsurface compositions as a probability density function (PDF). To formulate that function, we assume that the Martian subsurface is composed of ice, rock, and pore space. Thus, the set of all possible subsurface models can be described by two values: the volume percent of water ice and the volume percent of rock (with the pore volume given by the remaining percentage). In the absence of data or prior information, all composition models are equally likely. Our purpose in using the Bayesian method is to combine multiple sets of data to whittle down the set of possible model solutions and constrain the set of possible subsurface compositions at each location on Mars. Additionally, because the Bayesian approach is based on PDFs, the approach allows for the incorporation of measurement uncertainty.

Although some measurements may be ambiguous toward water ice, all the datasets we use allow us to place constraints on the composition of subsurface materials. For example, a low permittivity determined from radar reflections could either indicate water ice or highly porous material, but not solid rock. Conversely, a high thermal inertia determined from temperature measurements would suggest either solid rock or the presence of water ice, but not a porous material. Separately, the two measurements do not uniquely indicate ice, but together they do. Bayesian inversion allows us to handle such nuanced cases by probabilistically expressing the subsurface compositions that can explain each measured property.

We discretize the space of all composition models and link each model to potential data measurements via theoretical or empirical formulas. Radar reflections obtained by the SHARAD instrument result in estimates of the subsurface dielectric permittivity. By simplifying the Maxwell Garnett mixing model (Koledintseva et al. 2006), we can express the effective permittivity ϵ_{eff} as a function of the subsurface composition as:

$$\epsilon_{eff} = \frac{\epsilon_r}{1 - f_i \frac{\epsilon_i - \epsilon_r}{\epsilon_i} - f_p \frac{\epsilon_p - \epsilon_r}{\epsilon_p}},$$

where ϵ_i , ϵ_r , and ϵ_p are the permittivities of water ice, rock, and pore space, respectively, and f_i and f_p are the volume fractions of water ice and pore space, respectively. In reality, all types of rock do not have the same permittivity, so we assign ϵ_r to a distribution of values covering the range of plausible rock types on Mars. This means

that rather than a given composition model mapping to one ϵ_{eff} , it maps to a distribution of ϵ_{eff} values. Likewise, a given estimation of ϵ_{eff} maps to a distribution of composition models.

The water equivalent hydrogen (WEH) values derived from MONS data by Pathare et al. (2018) can be expressed as a function of the subsurface composition:

$$WEH\% = \frac{M_i + M_h WEH_{hmax}}{M_{total}} \times 100\%$$

where water–ice mass $M_i = f_i \times \rho_i$ is the water–ice volume fraction f_i multiplied by the water–ice density ρ_i ($= 920 \text{ kg/m}^3$), hydrated mineral mass $M_h = f_h \times \rho_h$ is the hydrated mineral volume fraction f_h multiplied by the density of hydrated minerals ρ_h ($\approx 1680 \text{ kg/m}^3$), $WEH_{hmax} = 50\%$ is the assumed maximum WEH% that could be recorded by a subsurface composed entirely of hydrated minerals, and the total mass $M_{total} = M_i + M_h + M_{rock} + M_{air}$. The density of the rock and the fraction of the rock that is composed of hydrated minerals are assigned to a distribution of values, so a given subsurface composition model maps to a distribution of WEH% values.

Analysis of TES data provides an estimate of the thermal inertia of the subsurface. The thermal inertia is related to the subsurface composition by:

$$TI = \sqrt{k\rho c_h}$$

where k is the thermal conductivity of the subsurface, ρ is the density of the subsurface, and c_h is the heat capacity of the subsurface. The densities and heat capacities are simply the volume-fraction weighted averages of corresponding values for the rock, ice, and pore-space components, but the thermal conductivity is more nuanced. We use the approach described by Mellon et al. (1997) to write the thermal conductivity as a function of ice, rock, and pore-space composition. Following the rules outlined by Tarantola (2005), the set of theoretical relationships among datasets from the radar, neutron spectrometer, and thermal spectrometer can be used to create PDFs that relate a given measurement to a distribution of composition models.

Geomorphological features do not have strict mathematical relationships to water–ice content. However, geomorphology can be added to the Bayesian approach if treated as prior knowledge. For example, recent impact events in some regions of Mars expose water ice in and around the craters they form. Although the exact concentration of the ice may not be determined from images, the fact that ice is visible adds a constraint to the minimum amount of ice that must be present in the subsurface in the general vicinity of the crater (Dundas and Byrne 2010; Dundas et al. 2014). Thus, we can assign qualitatively justified PDFs for all geomorphological features that are related to ice.

Once a PDF has been created for each dataset, we can combine the PDFs in a similar fashion to combining consistency values. For the near surface, 0–5 m depth, we combine PDFs for the geomorphology, radar surface data, thermal data, and neutron data. For the deeper subsurface, >5 m depth, we combine PDFs for the

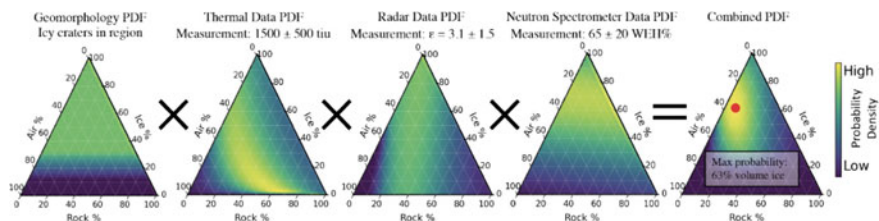


Fig. 16.6 Illustration of the Bayesian SWIM method applied to a single location on Mars that is in the vicinity of an icy crater and has a thermal inertia of 1500 ± 500 tiu, an effective permittivity of 3.1 ± 1.5 , and a WEH of $65 \pm 20\%$. Each observation and measurement has a PDF relating it to a set of permissible composition models. Upon multiplying the PDFs together, the result is a more limited set of models that can explain all measurements and indicates the most likely ice volume percentage (the red dot). In this case, the most likely concentration of water ice is 63%, and the overall result is consistent with porous dusty ice

geomorphology and radar subsurface data. We combine the PDFs through multiplication, following the rules of Tarantola (2005). By definition, multiplying PDFs results in a new PDF that is always a more refined state of information. Figure 16.6 shows an example combination of PDFs for the 0–5 m depth zone at one hypothetical location on Mars. The result gives us a PDF that indicates the most likely ice volume percent and also how certain that value is. By repeating this process for every location on Mars, we can generate a probability map of water–ice content across the Martian globe.

The Bayesian approach provides multiple advantages over the ice-consistency approach. First, it allows us to estimate ice percentage versus simply whether ice may or may not be present. Second, the method allows us to address quantitatively how measurement uncertainty and interpretation ambiguity lead to uncertainty about the ice concentration. Lastly, the procedure provides a robust set of rules for how to incorporate each element of knowledge we have about water ice on Mars. New data and information can easily be included in the framework to provide updated probability results, and by definition, new information can only refine the results and further constrain the composition. However, the SWIM Bayesian approach is still an area of active research that needs refinement and testing before the results are formally used. Specifically, we plan to refine the approach by adding compositional variation with depth to the models and by adding the correct depth dependence to each theoretical relation between measurement and composition when funding allows.

16.3 SWIM Results

16.3.1 *Non-Layered Ice Consistency*

We present the results of our predicted composite ice consistency (Eq. 16.1) in Fig. 16.7a. In addition to extending the mapping beyond the northern hemisphere study area (Morgan et al. 2021), we revised the methods used in generating the dataset-specific ice-consistency maps as described in Sect. 16.2. This product provides an overall view of where the collection of datasets points toward a high likelihood of finding buried ice. However, it does not provide a clear sense of the distribution of that ice with depth nor of its concentration relative to other materials in the subsurface.

16.3.2 *Ice Consistency for Depths <1 m*

We present the results of our predicted ice consistency for the upper meter of the Martian surface in Fig. 16.7b. Ice consistency in this depth range is driven by the weighted averages of the consistencies with ice from neutron and thermal spectrometers, surface radar, and shallow geomorphology observations (Eq. 16.2). Not surprisingly, the results are broadly consistent with those of prior mapping efforts based on data from thermal and neutron spectrometers, which are the dominant components of Eq. (16.2). To apply the SWIM 1.0 approach for determining the minimum threshold of an area of interest— $1/(\text{no. of techniques})$ (Morgan et al. 2021)—we need to also factor in the weighting incorporated within Eq. (16.2) when designating the denominator. For C_i [<1 m], this equates to $1/2.4 = 0.42$. In this case, either one technique (with a maximum weighting of 1) is recording a maximum positive ice consistency and the others are uncertain or, alternatively, multiple techniques are providing positive ice-consistency values. Considering this threshold of 0.42, we find that substantially high ice consistency occurs in three regions between 40° and 45° latitudes, specifically in Arcadia Planitia (170 – 220°E) and eastern Utopia Planitia (120 – 160°E) in the north and Promethei Terra (100 – 140°E) in the south. Also apparent in Fig. 16.7b are regions of slightly positive ice consistency ranging between 0.1 and 0.3 at low latitudes in Arabia Terra (0 – 50°E), Aeolis Mensae (165 – 200°E), and Medusae Fossae (165°E – 200°E). Two analyses of epithermal neutron flux—one using MONS data (Wilson et al. 2018) and one using Fine Resolution Epithermal Neutron Detector (FREND) data (Malakhov et al. 2020)—modeled a simplified uniform surface layer and argued for the presence of excess ice “oases” in these near-equatorial regions. However, Pathare et al. (2018) modeled a more realistic two-layer near surface that was constrained by epithermal, thermal, and fast neutrons, and they concluded that the MONS data are much more consistent with the

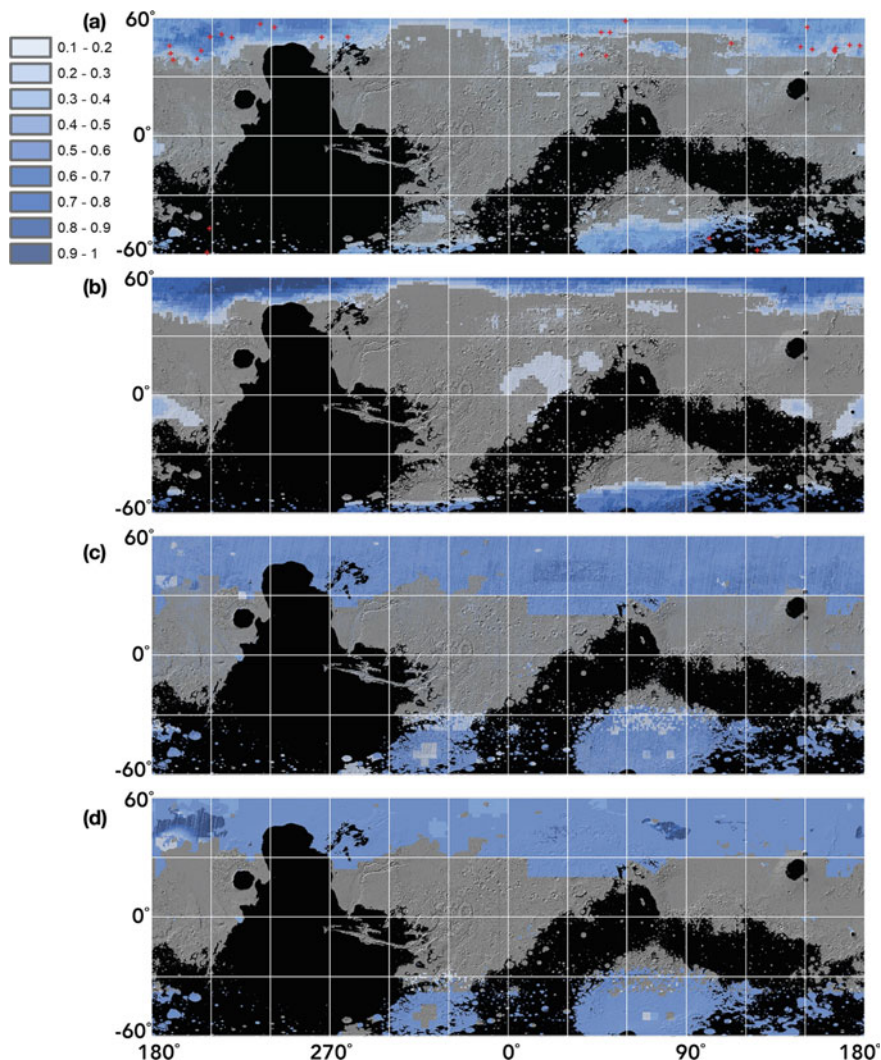


Fig. 16.7 Composite ice-consistency maps. **a** For all depths (C_i) using Eq. (16.1). **b** For depth zone <1 m (C_i [<1 m]) using Eq. (16.2). **c** For depth zone 1–5 m (C_i [1–5 m]) using Eq. (16.3). **d** For depth zone >5 m (C_i [>5 m]) using Eq. (16.4). Basemaps are MOLA-shaded relief, with black masking of elevations above +1 km. Red crosses in **(a)** represent ice-exposing impacts

widespread presence of hydrated minerals in these low-latitude regions. This conclusion is supported by a lack of highly elevated abundances of observed atmospheric water vapor that should result from the sublimation of such shallow excess ice at equatorial temperatures.

16.3.3 Ice Consistency for Depths of 1–5 m

We present the results of our predicted ice consistency for depths between 1 and 5 m in Fig. 16.7c. Ice consistency in this depth range is driven by the weighted averages of the consistencies with ice from surface radar, deep radar, and shallow geomorphology observations (Eq. 16.3). The neutron and thermal spectrometer ice-consistency values are not included, as their measurements do not extend into this depth range. Compared with results for the top 1 m of the subsurface (Fig. 16.7b), the results for 1–5 m depths show a substantial equatorward expansion of areas containing evidence of ice, with moderate ice-consistency values extending to $<30^\circ$ latitude in some locations, most notably in Arabia Terra south of Deuteronilus Mensae ($10\text{--}60^\circ\text{E}$) and within the Hellas basin ($\sim 45^\circ\text{--}90^\circ\text{E}$). The equatorward extent of continuous C_i [1–5 m] in both the northern and southern hemispheres is largely driven by the spatial occurrence of mantling units, which are included in the shallow geomorphology term (C_{GS}). The highest values of C_i [1–5 m] are in localized regions dispersed throughout Arcadia Planitia and the broader northern plains and in association with locations that have a high density of glacial features (LDA, LVF, and CCF) such as Deuteronilus Mensae and eastern Hellas. Intriguingly, several locations near the equator show low to moderate C_i [1–5 m]. Many of these locations are associated with the Medusae Fossae Formation, where pedestal craters are observed and where layers of dust might be contributing to the surface-radar signal. Near Schiaparelli crater in western Terra Sabaea ($\sim 5\text{--}10^\circ\text{S}$, 15°E), we mapped a number of landforms resembling CCF and other glacial features, consistent with previous work in this area (Shean 2010).

16.3.4 Ice Consistency for Depths >5 m

We present the results of our predicted ice consistency for depths >5 m in Fig. 16.7d. Ice consistency in this depth range is driven only by the deep radar and deep geomorphology terms (Eq. 16.4), since the other terms all have sensing depths ≤ 5 m. The deep geomorphology term excludes mantled and textured terrains due to their typical thickness <5 m, especially at the most equatorward locations where mantle appears as isolated patches and highly dissected. The mapping results for this depth zone are similar in spatial extent to those for C_i [1–5 m] (Fig. 16.7c) with notable exceptions in the occurrences of the highest ice-consistency values. The deep radar term includes detections of subsurface reflectors associated with widespread units in Arcadia Planitia and Utopia Planitia and glacial landforms across the northern and southern latitudes. The permittivities calculated from these reflectors and their elevation constraints show high consistency with ice. Combined with the corresponding presence of periglacial and glacial landforms, these yield the highest ice consistency in the map (Fig. 16.7d). The northeast-to-southwest gradient of ice consistency in Arcadia Planitia is likely an artifact of the limited number of elevation tie points available for calculating dielectric permittivity values from radar data in this region.

The excursions of moderate ice consistency into regions $<30^\circ$ latitude are still present (Fig. 16.7d) and are due to the occurrences of glacial landforms extending to these latitudes. These glacial landforms appear mostly as isolated CCF or other “icy” fill confined within craters. Unfortunately, the small size of these craters relative to the SHARAD radar footprint limits the ability to detect subsurface reflectors due to the pervasive clutter resulting from the steep crater walls.

Radar reflectors are not observed in many locations where there is geomorphic evidence of ice. Their absence may be due to several factors, including a potential lack of dielectric contrast between ice and regolith, strong attenuation of the radar signal (e.g., from surface roughness), limitations of the vertical and horizontal resolutions of SHARAD radar data, and obscuration of reflectors by clutter. Reflectors are most prevalent in association with LDA, LVF, and CCF in Deuteronilus and Protonilus Mensae and in eastern Hellas, as well as in the Amazonian units in Utopia and Arcadia, resulting in high C_i [>5 m] values, where subsurface ice may be thickest and/or most pure.

16.3.5 *SWIM Products*

All products from the Mars SWIM projects are being made available via the publicly accessible website at <https://swim.psi.edu>. The team also intends to archive the products with NASA’s Planetary Data System. The website contains PNG, TIFF, and GeoTIFF versions of all the composite C_i maps presented here as well as of the dataset-specific ice-consistency maps that went into creating the composites. Other information and ancillary data are also available, such as a map of the depth to the base of ice produced from SHARAD subsurface data.

16.4 Discussion

16.4.1 *Comparison of SWIM Results with Ice-Exposing Impacts*

As was discussed in Sect. 16.2.3, a major challenge to producing synthesized maps of ice presence on Mars is the lack of sufficient ground-truth data with which to calibrate our data products. Nevertheless, the detection of ice-exposing fresh impacts across Mars (Byrne et al. 2009; Dundas et al. 2014, 2021) provides an opportunity to evaluate our mapping approach.

The first fresh impact to be recognized on Mars was identified in Mars Orbiter Camera (MOC) data (Malin et al. 2006). Since the arrival of MRO to Mars, CTX has been actively employed to search for new impacts. To determine the validity of the detections, each potential impact is followed up by at least one HiRISE observation.

With an order-of-magnitude higher resolution relative to CTX and a multi-channel capability (one infrared and two visible color channels), HiRISE enables a search for evidence of ice excavated by the impact (either within the crater or in its ejecta blanket).

During the 2019 phase of the SWIM project, 14 ice-exposing impact sites had been identified within the northern hemisphere study area (Fig. 16.1) (Dundas et al. 2014). Comparisons between the impact locations and the corresponding values on the ice-consistency map of Morgan et al. (2021) showed good agreement, with 13 of the sites registering positive C_i values. The average C_i value of map pixels containing impacts was 0.26 ± 0.16 and the single negative value was close to zero, and thus considered negligible. As of the publication of Dundas et al. (2021), CTX and HiRISE have identified 48 fresh ice-exposing impact sites within the mid- and high latitudes of Mars, and 30 of these sites are located within the SWIM study region. Comparing the location of these sites with the updated 2020 SWIM mapping using Eq. (16.1) (Fig. 16.7a; to be concordant with the 2019 analysis), we found that all of the impact sites are within positive C_i pixels that have a mean value of 0.29 ± 0.14 .

At most of these sites, there are multiple craters due to bolide break-up in the Martian atmosphere prior to impacting the surface. Regardless of the number of craters at a given site, the diameters of the majority of the ice-exposing impacts are <10 m, with the largest at 48 m. From empirical studies of the relationship between the diameter and depth of impact craters, we expect most of the ice to have been excavated from the upper ~ 1 m of the subsurface (Dundas et al. 2021). As a result, the ice sampled by the impacts is most relevant to C_i [<1 m] map product (Fig. 16.7b). In this case 27 of the 30 impacts correspond to positive C_i [<1 m] values, with a mean of 0.39 ± 0.32 . One of the three impacts with negative C_i [<1 m] values is the largest of the ice-exposing impacts with a diameter of 48 m that likely excavated ~ 4 m into the subsurface. It is therefore possible that the upper meter of the subsurface at this location is ice-free. The corresponding C_i [1–5 m] value for this location, $+0.5$, is consistent with the presence of deeper ice.

The strong correlation between ice-exposing impacts and the SWIM maps lends weight to the effectiveness of integrating multiple datasets. Within the northern hemisphere, the clustering of low-latitude impacts within longitudes that correspond to broad regions of elevated C_i values (140–220°E) provides further encouragement. It is important to note that a hemispherical discrepancy exists between the number of ice-exposing craters observed in the north relative to the south. The distinct lack of fresh crater detections (icy or not) in the south is likely an observation bias in part related to the lower dust cover (Dundas et al. 2021) and should therefore not be attributed to a lack of buried ice. Additionally, imaging of the clusters of craters indicates that there are local-scale variations in the depth to and concentration of buried ice (Dundas et al. 2021).

16.4.2 Constraints on Ice Content

To enable the use of buried ice as a resource at any given location, one must understand its geographic distribution, thickness, depth of burial, and purity. The results of the SWIM study contribute to this understanding and place important constraints on each of these factors. A key result is new, more complete mapping of the geographic distribution of where ice is mostly likely to be found. Where subsurface radar returns related to buried ice are constrained by elevation data, they have allowed us to not only determine the depth to the base of ice but also to assess the bulk concentration of ice within the column extending to the surface. Where ice is shallow, the neutron and thermal spectrometer data allowed an estimation of the burial depth within the upper meter of the subsurface, with the neutron data also placing limits on the concentration of the ice. Detection of hydrogen via neutron-spectrometer data is inherently limited to a sensing depth less than about 0.5 m below the Martian surface. Similarly, the ability to detect buried ice by its thermal effects is also limited to about 0.5 m by the thermal skin depth of geological materials that form the overburden above the ice table. While the SHARAD surface-return data enable some measure of material properties in the upper 5 m, that assessment is of bulk properties and does not allow one to identify variations or distinct interfaces in this zone. The subsurface radar sounding by SHARAD cannot resolve interfaces shallower than 15 m deep due to a combination of the inherent vertical resolution of the radar and its band-limited nature that leads to interferences with the surface return. These considerations leave some shortcomings in our ability to resolve ice content between depths of 0.5 and 15 m, although the radar surface return does contribute toward understanding properties down to 5 m depth.

16.4.3 Future Considerations

As noted above, a primary motivation for the SWIM project has been mapping of buried water ice that may serve as a potential resource for future human missions. While the work presented here represents a major advance in the integration of datasets and their broad application across Mars, there is much more that can be done with the existing data to further evaluate the resource potential of buried ice. For example, due to the practical limitations of time and personnel available for this study, the grid-mapping technique employed did not yield an exhaustive catalog of individual geomorphological indicators of buried ice. More generally, complete mapping at the highest resolutions of all available data was beyond the scope of the project. In addition, a complete scientific assessment of the analyzed data, such as the age, nature and climatological implications of the mapped buried ice, was not part of the funded effort, and the Bayesian statistical approach was limited to a preliminary

analysis. The products of this study are intended to guide choices for more in-depth studies by future workers to support their evaluation of potential human landing sites. The SWIM products and methods may also serve to inform future scientific studies related to Martian climate history.

Given the limitations of the currently available data (see Sect. 16.4.2), obtaining a more thorough understanding of buried ice on Mars will also require new instrumentation. The gap in sensing depth between existing radar and other methods is especially limiting, as one cannot confidently map the depth of the ice table where it extends to depths >0.5 m using currently available data. To resolve these depths, new instrumentation that can build on the capabilities of previous and current instruments is needed. From a global perspective, a high-frequency radar sounder (L–P band) and/or synthetic-aperture radar imager would be extremely complementary to SWIM and other ice-detection studies by bridging the gap between thermal and neutron spectroscopy data and the SHARAD data. In this regard the initial plans for an International Mars Ice Mapper (I-MIM) mission (Watzin 2021) are particularly timely. Ultimately, it would be best to obtain actual ground truth at a prospective human landing site using a landed robotic mission with a drilling system capable of reaching ice within a few meters of the surface.

16.5 Acronyms and Mathematical Symbols

ATI	Apparent thermal inertia
BF08	Bandfield and Feldman (2008)
c	Speed of light in vacuum
c_h	Bulk heat capacity
C	Ice consistency
C_i	Ice consistency derived from multiple data sets
C_G	Ice consistency derived from geomorphological data
C_{GD}	Ice consistency derived from deep geomorphological data
C_{GS}	Ice consistency derived from shallow geomorphological data
C_N	Ice consistency derived from neutron spectrometer data
C_{RD}	Ice consistency derived from radar subsurface dielectric estimations
C_{RS}	Ice consistency derived from radar surface returns
C_T	Ice consistency derived from thermal spectrometer data
CTX	Context Camera on MRO
CRISM	Compact Reconnaissance Imaging Spectrometer for Mars on MRO
d	Depth below surface
f_i	Volume fraction of ice
f_p	Volume fraction of pore space

(continued)

(continued)

f_r	Volume fraction of rock
f_h	Volume fraction of hydrated minerals within rock
h	Thickness (height) of a geologic layer
HiRISE	High Resolution Imaging Science Experiment on MRO
I-MIM	International Mars Ice Mapper
JPL	Jet Propulsion Laboratory
k	Bulk conductivity
LDA	Lobate Debris Apron
L_S	Season (areocentric longitude of the Sun)
L_S^*	Season of buried ice insensitivity
MARSIS	Mars Advanced Radar for Subsurface and Ionospheric Sounding on Mars Express
MCS	Mars Climate Sounder on MRO
MGS	Mars Global Surveyor
MOLA	Mars Orbiter Laser Altimeter on MGS
MONS	Mars Odyssey Neutron Spectrometer
MRO	Mars Reconnaissance Orbiter
NASA	National Aeronautics and Space Administration
NPLD	North polar layered deposits
PDF	Probability density function
pf	Periglacial feature
PQ19	Piqueux et al. (2019)
s_Y	Shallowness of method Y (subscripts as for C_Y listed above)
SP20	SWIM Project (2020)
SPLD	South polar layered deposits
SHARAD	Shallow Radar on MRO
SWIM	Subsurface Water Ice Mapping
TES	Thermal Emission Spectrometer on MGS
TI	Thermal inertia
WEH	Water equivalent hydrogen
W_{dn}	WEH weight fraction of a lower (“down”) model layer
W_{up}	WEH weight fraction of an upper model layer
W_{hmax}	Maximum WEH for a subsurface composed of the most hydrated mineral
Δt	Radar two-way delay time
ε'	Real part of the relative dielectric permittivity
ε_{eff}	Effective dielectric permittivity
ε_i	Ice dielectric permittivity
ε_p	Pore-space dielectric permittivity
ε_r	Rock dielectric permittivity

(continued)

(continued)

ρ	Bulk density
ρ_i	Ice density
ρ_r	Rock density
ρ_w	Water density
ρ_h	Hydrated-minerals density

Acknowledgements The SWIM project was funded by NASA through JPL subcontracts 1589197, 1595721, 1611855, and 1639821. The authors are grateful to NASA, the Italian Space Agency, and the Mars Reconnaissance Orbiter Project for their ongoing support of the MRO mission, which has been invaluable to the work described here. We thank Joseph MacGregor, Laszlo Kestay, Timothy Titus, Kevin Jones, and an anonymous reviewer for insightful comments and suggestions that greatly improved the manuscript.

References

- Ash, R.L., W.L. Dowler, and G. Varsi. 1978. Feasibility of rocket propellant production on Mars. *Acta Astronautica* 5: 705–724. [https://doi.org/10.1016/0094-5765\(78\)90049-8](https://doi.org/10.1016/0094-5765(78)90049-8).
- Baker, D.M.H., and L.M. Carter. 2019. Probing supraglacial debris on Mars 2: Crater morphology. *Icarus* 319: 264–280. <https://doi.org/10.1016/j.icarus.2018.09.009>.
- Bandfield, J.L., and W.C. Feldman. 2008. Martian high latitude permafrost depth and surface cover thermal inertia distributions. *Journal of Geophysical Research: Planets* 113. <https://doi.org/10.1029/2007JE003007>.
- Berman, D.C., F.C. Chuang, I.B. Smith, and D.A. Crown. 2021. Ice-rich landforms of the southern mid-latitudes of Mars: A case study in Nereidum Montes. *Icarus* 355: 114170. <https://doi.org/10.1016/j.icarus.2020.114170>.
- Bramson, A.M., S. Byrne, N.E. Putzig, S. Sutton, J.J. Plaut, T.C. Brothers, and J.W. Holt. 2015. Widespread excess ice in Arcadia Planitia, Mars. *Geophysical Research Letters* 42: 6566–6574. <https://doi.org/10.1002/2015GL064844>.
- Brouet, Y., P. Becerra, P. Sabouroux, A. Pommerol, and N. Thomas. 2019. A laboratory-based dielectric model for the radar sounding of the martian subsurface. *Icarus* 321: 960–973. <https://doi.org/10.1016/j.icarus.2018.12.029>.
- Byrne, S., C.M. Dundas, M.R. Kennedy, M.T. Mellon, A.S. McEwen, S.C. Cull, I.J. Daubar, D.E. Shean, K.D. Seelos, and S.L. Murchie. 2009. Distribution of mid-latitude ground ice on Mars from new impact craters. *Science* 325: 1674–1676.
- Campbell, M.J., and J. Ulrichs. 1969. Electrical properties of rocks and their significance for lunar radar observations. *Journal of Geophysical Research* 74: 5867–5881. <https://doi.org/10.1029/JB074i025p05867>.
- Campbell, B.A., N.E. Putzig, L.M. Carter, G.A. Morgan, R.J. Phillips, and J.J. Plaut. 2013. Roughness and near-surface density of Mars from SHARAD radar echoes. *Journal of Geophysical Research* 118: 436–450. <https://doi.org/10.1002/jgre.20050>.
- Carr, M.H. 2001. Mars Global Surveyor observations of Martian fretted terrain. *Journal of Geophysical Research* 106: 23571–23593. <https://doi.org/10.1029/2000JE001316>.
- Choudhary, P., J.W. Holt, and S.D. Kempf. 2016. Surface clutter and echo location analysis for the interpretation of SHARAD data from mars. *IEEE Geoscience and Remote Sensing Letters* 13 (9): 1285–1289. <https://doi.org/10.1109/LGRS.2016.2581799>.

- Christensen, P.R., J.L. Bandfield, V.E. Hamilton, S.W. Ruff, G.L. Mehall, N. Gorelick, K. Bender, K. Murray, H.H. Kieffer, and T.N. Titus. 2001. Mars Global Surveyor thermal emission spectrometer experiment: Investigation description and surface science results. *Journal of Geophysical Research* 106: 23823–23871.
- Dickson, J.L., L.A. Kerber, C.I. Fassett, and B.L. Ehlmann. 2018. A global, blended CTX mosaic of Mars with vectorized seam mapping: a new mosaicking pipeline using principles of non-destructive image editing. In *Proceedings of 49th Lunar and planetary science conference abstract*, 2480.
- Dundas, C.M., and S. Byrne. 2010. Modeling sublimation of ice exposed by new impacts in the Martian mid-latitudes. *Icarus* 206: 716–728. <https://doi.org/10.1016/j.icarus.2009.09.007>.
- Dundas, C.M., S. Byrne, A.S. McEwen, M.T. Mellon, M.R. Kennedy, I.J. Daubar, and L. Saper. 2014. HiRISE observations of new impact craters exposing Martian ground ice. *Journal of Geophysical Research* 119: 109–127. <https://doi.org/10.1002/2013JE004482>.
- Dundas, C.M., A.M. Bramson, L. Ojha, J.J. Wray, M.T. Mellon, S. Byrne, A.S. McEwen, N.E. Putzig, D. Viola, S. Sutton, E. Clark, and J.W. Holt. 2018. Exposed subsurface ice sheets in the Martian mid-latitudes. *Science* 359: 199–201. <https://doi.org/10.1126/science.aao1619>.
- Dundas, C.M., M.T. Mellon, S.J. Conway, I.J. Daubar, K.E. Williams, L. Ojha, J.J. Wray, A.M. Bramson, S. Byrne, A.S. McEwen, L.V. Posiolova, G. Speth, D. Viola, M.E. Landis, G.A. Morgan, and A.V. Pathare. 2021. Widespread exposures of extensive clean shallow ice in the mid-latitudes of mars. *Journal of Geophysical Research: Planets*. <https://doi.org/10.1029/2020JE006617>.
- Feldman, W.C., W.V. Boynton, R.L. Tokar, T.H. Prettyman, O. Gasnault, S.W. Squyres, R.C. Elphic, D.J. Lawrence, S.L. Lawson, S. Maurice, G.W. McKinney, K.R. Moore, and R.C. Reedy. 2002. Global distribution of neutrons from Mars: Results from Mars Odyssey. *Science* 297: 75–78. <https://doi.org/10.1126/science.1073541>.
- Feldman, W.C., Pathare, A., Maurice, S., Prettyman, T.H., Lawrence, D.J., Milliken, R.E., Travis, B.J., 2011. Mars Odyssey neutron data: 2. Search for buried excess water ice deposits at nonpolar latitudes on Mars. *Journal of Geophysical Research: Planets*, 116(E11).
- Grima, C., W. Kofman, J. Mouginot, R.J. Phillips, A. Hérique, D. Biccari, R. Seu, and M. Cutigni. 2009. North polar deposits of Mars: Extreme purity of the water ice. *Geophysical Research Letters* 36: L03203. <https://doi.org/10.1029/2008GL036236>.
- Grima, C., W. Kofman, A. Herique, R. Orosei, and R. Seu. 2012. Quantitative analysis of Mars surface radar reflectivity at 20 MHz. *Icarus* 220: 84–99.
- Head, J.W., D.R. Marchant, J.L. Dickson, A.M. Kress, and D.M. Baker. 2010. Northern mid-latitude glaciation in the Late Amazonian period of Mars: Criteria for the recognition of debris-covered glacier and valley glacier landsystem deposits. *Earth and Planetary Science Letters* 294: 306–320. <https://doi.org/10.1016/j.epsl.2009.06.041>.
- Kadish, S.J., N.G. Barlow, and J.W. Head. 2009. Latitude dependence of Martian pedestal craters: Evidence for a sublimation-driven formation mechanism. *Journal of Geophysical Research* 114: E10001. <https://doi.org/10.1029/2008JE003318>.
- Kieffer, H.H. 2013. Thermal model for analysis of Mars infrared mapping. *J. Geophys. Res. Planets* 118: 451–470.
- Koledintseva, M., et al. 2006. A maxwell garnett model for dielectric mixtures containing conducting particles at optical frequencies. *Progress in Electromagnetics Research* 63: 223–242.
- Kostama, V.-P., M.A. Kreslavsky, and J.W. Head. 2006. Recent high-latitude icy mantle in the northern plains of Mars: Characteristics and ages of emplacement. *Geophysical Research Letters* 33: L11201. <https://doi.org/10.1029/2006GL025946>.
- Kress, A.M., and J.W. Head. 2008. Ring-mold craters in lineated valley fill and lobate debris aprons on Mars: Evidence for subsurface glacial ice. *Geophysical Research Letters* 35: L23206. <https://doi.org/10.1029/2008GL035501>.
- Leighton, R.B., and B.C. Murray. 1966. Behavior of carbon dioxide and other volatiles on Mars. *Science* 153: 136–144.
- Levy, J.S., C.I. Fassett, J.W. Head, C. Schwartz, and J.L. Waters. 2014. Sequestered glacial ice contribution to the global Martian water budget: Geometric constraints on the volume of remnant,

- midlatitude debris-covered glaciers: Buried Martian glaciers. *J. Geophys. Res. Planets* 119: 2188–2196. <https://doi.org/10.1002/2014JE004685>.
- Malakhov, A.V., I.G. Mitrofanov, M.L. Litvak, A.B. Sanin, D.V. Golovin, M.V. Djachkova, S.Y. Nikiforov, A.A. Anikin, D.I. Lisov, N.V. Lukyanov, M.I. Mokrousov. 2020. Ice Permafrost “Oases” Close to Martian equator: planet neutron mapping based on data of FREND instrument onboard TGO orbiter of Russian-European ExoMars mission. *Astronomy Letters* 46: 407–421. <https://doi.org/10.1134/S1063773720060079>.
- Malin, M.C., K.S. Edgett, L.V. Posiolova, S.M. McColley, and E.Z.N. Dobreá. 2006. Present-day impact cratering rate and contemporary gully activity on Mars. *Science* 314: 1573–1577. <https://doi.org/10.1126/science.1135156>.
- Mangold, N. 2005. High latitude patterned grounds on Mars: Classification, distribution and climatic control. *Icarus* 174: 336–359.
- Mellon, M.T., B.M. Jakosky, and S.E. Postawko. 1997. The persistence of equatorial ground ice on Mars. *Journal of Geophysical Research* 102: 19357–19369.
- Mellon, M.T., B.M. Jakosky, B.M. Kieffer, H.H. Kieffer, and P.R. Christensen. 2000. High-resolution thermal inertia mapping from the Mars Global Surveyor thermal emission spectrometer. *Icarus* 148: 437–455.
- Mellon, M.T., W.C. Feldman, and T.H. Prettyman. 2004. The presence and stability of ground ice in the southern hemisphere of Mars. *Icarus* 169: 324–340.
- Milliken, R.E., J.F. Mustard, and D.L. Goldsby. 2003. Viscous flow features on the surface of Mars: Observations from high-resolution Mars Orbiter Camera (MOC) images. *Journal of Geophysical Research* 108: 5057. <https://doi.org/10.1029/2002JE002005>.
- Morgan, G.A., and B.A. Campbell. 2017. Local subsurface ice mapping through the integration of SHARAD derived data products with other datasets: A proposal to the Jet Propulsion Laboratory request for proposals no: KM-2691-947266 on Mapping of Water Deposits to Support NASA Mars Exploration Program Studies.
- Morgan, G.A., N.E. Putzig, M.R. Perry, H.G. Sizemore, A.M. Bramson, E.I. Petersen, Z.M. Bain, D.M.H. Baker, M. Mastrogiuseppe, R.H. Hoover, I.B. Smith, A. Pathare, C.M. Dundas, and B.A. Campbell. 2021. Availability of Martian water-ice resources to support human missions. *Nature Astronomy*. <https://doi.org/10.1038/s41550-020-01290-z>.
- Morgenstern, A., E. Hauber, D. Reiss, S. van Gasselt, G. Grosse, and L. Schirrmeyer. 2007. Deposition and degradation of a volatile-rich layer in Utopia Planitia and implications for climate history on Mars. *Journal of Geophysical Research* 112: E06010. <https://doi.org/10.1029/2006JE002869>.
- Mouginot, J., A. Pommerol, W. Kofman, P. Beck, and B. Schmitt. 2010. The 3–5 MHz global reflectivity map of Mars by MARSIS/Mars express: Implications for the current inventory of subsurface H₂O. *Icarus* 210: 612–625.
- Mustard, J.F., C.D. Cooper, and M.K. Rifkin. 2001. Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* 412: 411–414.
- Pathare, A.V., W.C. Feldman, T.H. Prettyman, and S. Maurice. 2018. Driven by excess? Climatic implications of new global mapping of near-surface water-equivalent hydrogen on Mars. *Icarus* 301: 97–116. <https://doi.org/10.1016/j.icarus.2017.09.031>.
- Petersen, E.I., J.W. Holt, and J.S. Levy. 2018. High ice purity of Martian Lobate Debris Aprons at the regional scale: Evidence from an orbital radar sounding survey in Deuteronilus and Protonilus Mensae. *Geophysical Research Letters* 45 (21): 11595–11604. <https://doi.org/10.1029/2018GL079759>.
- Piqueux, S., and P.R. Christensen. 2009. A model of thermal conductivity for planetary soils 2. Theory for cemented soils. *Journal of Geophysical Research* 114. <https://doi.org/10.1029/2008JE003309>.
- Piqueux, S., J. Buz, C. S. Edwards, J.L. Bandfield, A. Kleinböhl, D.M. Kass, P.O. Hayne, and the MCS and THEMIS Teams. 2019. Widespread shallow water ice on Mars at high latitudes and midlatitudes. *Geophysical Research Letters* 46: 14290–14298.

- Plaut, J.J., A. Ivanov, A. Safaeinili, S.M. Milkovich, G. Picardi, R. Seu, and R. Phillips. 2007. Radar sounding of subsurface layers in the south polar plains of Mars: Correlation with the Dorsa Argentea Formation. Presented at the Lunar Planetary Science.
- Putzig, N.E., and M.T. Mellon. 2007. Apparent thermal inertia and the surface heterogeneity of Mars. *Icarus* 191: 68–94. <https://doi.org/10.1016/j.icarus.2007.05.013>.
- Putzig, N.E., H.G. Sizemore, and I.B. Smith. 2017. Mapping buried water ice in Arcadia & beyond with radar & thermal data: A proposal to the Jet Propulsion Laboratory Request for Proposals No: KM-2691-947266 on Mapping of Water Deposits to Support NASA Mars Exploration Program Studies.
- Putzig, N.E., E.M. Barratt, M.T. Mellon, and T.I. Michaels. 2013. MARSTHERM: A web-based system providing thermophysical analysis tools for Mars research. *AGU Fall Meeting*, abstract P43C-2023.
- Putzig, N.E., M.T. Mellon, K.E. Herkenhoff, R.J. Phillips, B.J. Davis, K.J. Ewer, and L.M. Bowers. 2014. Thermal behavior and ice-table depth within the north polar erg of Mars. *Icarus* 230: 64–76.
- Ramsdale, J.D., M.R. Balme, S.J. Conway, C. Gallagher, S.A. van Gasselt, E. Hauber, C. Orgel, A. Séjourné, J.A. Skinner, F. Costard, A. Johnsson, A. Losiak, D. Reiss, Z.M. Swirad, A. Kereszturi, I.B. Smith, and T. Platz. 2017. Grid-based mapping: A method for rapidly determining the spatial distributions of small features over very large areas. *Planetary and Space Science* 140: 49–61. <https://doi.org/10.1016/j.pss.2017.04.002>.
- Seu, R., R.J. Phillips, D. Biccari, R. Orosei, A. Masdea, G. Picardi, A. Safaeinili, B.A. Campbell, J.J. Plaut, L. Marinangeli, S.E. Smrekar, and D.C. Nunes. 2007. SHARAD sounding radar on the Mars Reconnaissance Orbiter. *Journal of Geophysical Research* 112: E05S05. <https://doi.org/10.1029/2006JE002745>.
- Shean, D.E. 2010. Candidate ice-rich material within equatorial craters on Mars. *Geophysical Research Letters* 37: L24202. <https://doi.org/10.1029/2010GL045181>.
- Siegler, M., O. Aharonson, E. Carey, M. Choukroun, T. Hudson, N. Schorghofer, and S. Xu. 2012. Measurements of thermal properties of icy Mars regolith analogs. *Journal of Geophysical Research: Planets* 117. <https://doi.org/10.1029/2011JE003938>.
- Sihvola, A.H. 1999. *Electromagnetic mixing formulas and applications*, 47. London, UK. Institution of Electrical Engineers, Electromagnetic Waves Series.
- Sizemore, H. G., A. Pathare, R.H. Hoover, N.E. Putzig, Z. Bain, C.M. Dundas, M.T. Mellon, and the SWIM Team. 2020. Subsurface Water Ice Mapping (SWIM) on Mars: Thermal and neutron datasets. In: *51st Lunar and Planetary Science Conference*, Abstract 2529.
- Smith, P.H., L. Tamppari, R.E. Arvidson, D. Bass, D. Blaney, W. Boynton, A. Carswell, D. Catling, B. Clark, T. Duck, E. DeJong, D. Fisher, W. Goetz, P. Gunnlaugsson, M. Hecht, V. Hipkin, J. Hoffman, S. Hviid, H. Keller, S. Kounaves, C.F. Lange, M. Lemmon, M. Madsen, M. Malin, W. Markiewicz, J. Marshall, C. McKay, M. Mellon, D. Michelangeli, D. Ming, R. Morris, N. Renno, W.T. Pike, U. Staufer, C. Stoker, P. Taylor, J. Whiteway, S. Young, and A. Zent. 2008. Introduction to special section on the Phoenix Mission: Landing site characterization experiments, mission overviews, and expected science. *Journal of Geophysical Research* 113: E00A18. <https://doi.org/10.1029/2008JE003083>.
- Souness, C., and B. Hubbard. 2012. Mid-latitude glaciation on Mars. *Progress in Physical Geography: Earth and Environment* 36: 238–261. <https://doi.org/10.1177/0309133312436570>.
- Stillman, D.E., R.E. Grimm, and S.F. Dec. 2010. Low-frequency electrical properties of ice–silicate mixtures. *The Journal of Physical Chemistry B* 114 (18): 6065–6073. <https://doi.org/10.1021/jp9070778>.
- Tanaka, K.L., J.A. Skinner, and T.M. Hare. 2005. Geologic map of the northern plains of Mars, USGS Scientific Investigations Map 2888. US Geological Survey.
- Tarantola, Albert. 2005. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.
- Ulaby, F.T., R.K. Moore, and A.K. Fung. 1986. *Microwave remote sensing: active and Passive. Vol. 2. Radar remote sensing and surface scattering and emission theory*, 962–966. Chapter 12. Norwood: Artech House Publishers.

- Viola, D., A.S. McEwen, C.M. Dundas, and S. Byrne. 2015. Expanded secondary craters in the Arcadia Planitia region, Mars: Evidence for tens of Myr-old shallow subsurface ice. *Icarus* 248: 190–204.
- Viola, D., and A.S. McEwen. 2018. Geomorphological evidence for shallow ice in the southern hemisphere of Mars. *Journal of Geophysical Research: Planets* 123: 262–277. <https://doi.org/10.1002/2017JE005366>.
- Watzin, J. 2021. NASA Perspectives on future exploration science at Mars. Presentation to the Mars Exploration Program Analysis Group (MEPAG), Virtual Meeting 11, January 27.
- Wilson, J.T., V.R. Eke, R.J. Massey, R.C. Elphic, W.C. Feldman, S. Maurice, and L.F.A. Teodoro. 2018. Equatorial locations of water on Mars: Improved resolution maps based on Mars Odyssey Neutron Spectrometer data. *Icarus* 299: 148–160. <https://doi.org/10.1016/j.icarus.2017.07.028>.
- Zubrin, R., D. Daker, and O. Gwynne. 1991. Mars direct: A simple, robust, and cost effective architecture for the Space Exploration Initiative. In *Proceedings of 29th Aerospace Sciences Meeting*, AIAA.

Chapter 17

Design and Modeling of an Electrochemical Device Producing Methane/Oxygen and Polyethylene from In-Situ Resources on Mars



Jeffery B. Greenblatt

Abstract Preliminary engineering designs were developed for two devices that work in tandem with an electrochemical device that converts Martian carbon dioxide (CO_2) and water (H_2O) into methane (CH_4), ethylene (C_2H_4) and oxygen (O_2). Major components of each device along with component masses, energy consumption, and mass flows of major constituents were modeled. The first device converts CO_2 and H_2O into CH_4 and O_2 propellant for a human-rated Mars Ascent Vehicle based on NASA's 2009 Design Reference Architecture 5.0 (DRA 5.0) for producing 7 metric tonnes (t) CH_4 and 23 t O_2 over 480 days. We conclude that a 412-kg device, with stowed dimensions of $0.7 \text{ m} \times 0.7 \text{ m} \times 0.6 \text{ m}$, drawing 30 kW, can produce the required amount of propellant (14.6 kg/day CH_4 and 47.9 kg/day O_2), along with 0.8 kg/day hydrogen (H_2) and 77 kg/day of additional O_2 for other uses (life support, etc.). The second device converts C_2H_4 into high-density polyethylene (HDPE), a versatile, high-strength polymer. We find that a 36-kg device consuming 750 W can convert a mixture of CH_4 and C_2H_4 from the first device into 9 kg/day of HDPE, while also increasing the purity of the unreacted CH_4 to 97%.

17.1 Background

Opus 12 (now Twelve) is developing a technology that will enable the “one-pot” synthesis of methane (CH_4) and/or ethylene (C_2H_4) from Martian CO_2 and H_2O using electricity, with O_2 as the other major product. Minor products such as H_2 , CO , etc. are also synthesized and must be separated. CH_4 and/or C_2H_4 can be used along with O_2 as a propellant for spacecraft or surface applications. C_2H_4 can also be fed into a secondary reactor to make common plastics such as polyethylene. It was assumed that existing equipment on the surface of Mars would supply pure CO_2 , H_2O and electricity as inputs to our device.

J. B. Greenblatt (✉)

Emerging Futures, LLC, 721 NW 9th Ave., Suite 195, OR 97209 Portland, USA

e-mail: jeff@emerging-futures.com

© Springer Nature Switzerland AG 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,

https://doi.org/10.1007/978-3-030-97913-3_17

617

This work was conducted on behalf of Opus 12 as part of a Phase II NASA SBIR titled “In-Situ Ethylene and Methane Production from CO₂ as Plastic Precursors.” Emerging Futures, LLC, as a subcontractor to Opus 12, was tasked with performing system analysis and design of a thermal management system, identification of uses of rejected heat for separations, and design of a full CO₂-to-plastic reactor. This chapter outlines the results of this 18-month research project.

17.1.1 Context

For any Mars mission with long-term aspirations, the use of in-situ resource utilization (ISRU) is paramount, as mass brought from Earth is very expensive and constrains the useful payload; the ability to produce fuel, water, oxygen, or chemicals for surface use could be game-changing. Fortunately, Mars is well-endowed with both CO₂ in its atmosphere as well as H₂O found in many locations on the surface or subsurface, including in the midlatitudes and, especially, polar regions. Most missions planned to Mars include, at minimum, atmospheric CO₂ extraction to produce breathing O₂ or a portion of propellant mass, and H₂O extraction and purification for astronaut use. Some missions also include production of CH₄ (typically via the Sabatier reaction). NASA’s DRA 5.0 specifies production of CH₄ and O₂ propellants for its MAV, which we used as the target production rate for our project.

Less important but still potentially very useful is the production of materials such as hydrocarbon polymers, which have a variety of uses including structural rigidity, durability, liquid and gaseous materials storage, and food preservation. No specific material or quantitative target was established for the project, so the goal was instead to determine how much of a high-quality polymer could be produced while also fulfilling the propellant requirements outlined above. Moreover, it was assumed that significant quantities of C₂H₄ would be co-produced along with CH₄ in the Opus 12 reactor. Since C₂H₄ is the monomer used to make various types of polyethylene, this general polymer class was chosen as the target.

The two devices described here address the needs of both propellant and polymer production, providing important enabling technologies that can add considerable capabilities to both robotic and human missions.

17.1.2 System Boundary Assumptions

For the purposes of our study, we assumed that other system components being developed by other teams would produce the required quantities of purified CO₂, H₂O and electricity on the Martian surface, including the extraction and separation of CO₂ from other atmospheric gases, and the extraction, transport and separation of H₂O from regolith with which it is likely to be mixed. Moreover, we assumed that an electrical system would provide continuous power at any level required, irrespective

of time of day, seasonal changes, or disruptions, for example, from dust storms; while we strove to minimize our use of electricity, it was not a limiting factor. These conditions represented the assumed input boundaries for our devices.

On the output side, we assumed that the desired products (CH_4 , O_2 , C_2H_4 , and H_2) would be compressed (and for gases, possibly liquefied) by other systems, so the additional energy required to do so was not included. Moreover, undesired co-products such as carbon monoxide (CO), ethanol, or unreacted C_2H_4 were assumed to be discarded (by venting to the atmosphere).

We also made various assumptions about the surface location of the devices to provide estimates of daytime and nighttime temperatures, average wind speeds, insolation, etc.

17.1.3 Sabatier System

Kleinhenz and Paz (2017) developed a schematic overview of a $\text{CH}_4 + \text{O}_2$ production device using Martian resources based on the Sabatier reaction (see Fig. 17.1).

Figure 17.1 represents two possible configurations:

- Using both Martian CO_2 and H_2O , the electrolyzer produces all the H_2 needed as an input to the Sabatier reactor, plus all the O_2 required for a stoichiometric $\text{CH}_4:\text{O}_2$ (1:2 mol) ratio. The chemical reactions involved are:

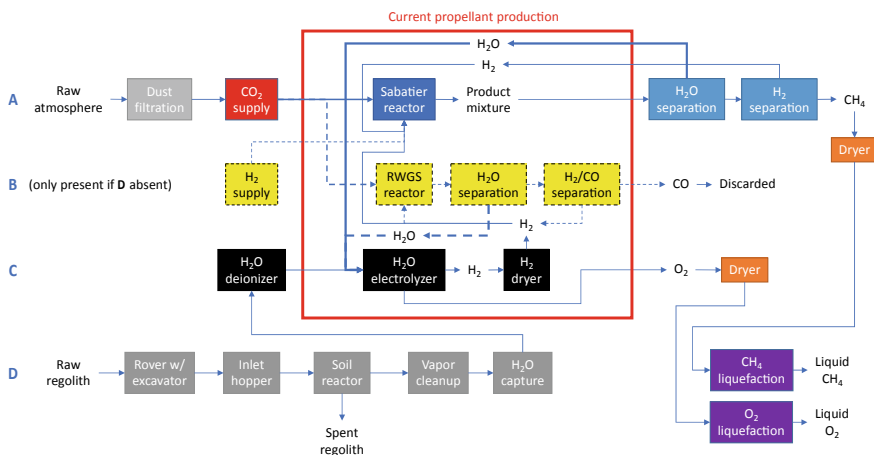
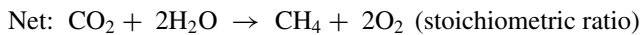
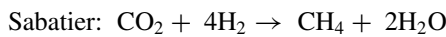
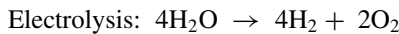
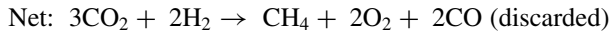
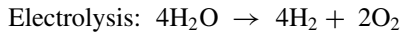
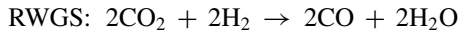
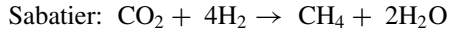


Fig. 17.1 Diagram of conventional Sabatier approach elements

- Using only Martian CO₂, the H₂ must be supplied from Earth; about 1,750 kg would be required to meet the DRA5.0 CH₄ production target. In this case, a reverse water–gas shift (RWGS) reactor is also needed to produce sufficient O₂ to meet the stoichiometric ratio. The chemical reactions are:



Sabatier reactors operate at temperatures of 200–550 °C and pressures of 1–100 bar (Götz et al. 2016). Heat produced in the reaction must be removed. Higher pressures are more favorable, allowing higher conversion efficiency and production of high-grade heat that may be useful for other system elements. Several approaches, including fixed-bed, fluidized-bed, three-phase, and structured reactors have been explored. Many Sabatier devices (e.g., El Sherif and Knox 2005; Murdoch et al. 2005; Junaedi et al. 2012, 2014; Kappmaier et al. 2016) are only designed to convert CO₂ to O₂, such as for air revitalization in the International Space Station and other spacecraft, and the CH₄ produced is typically discarded. These Sabatier devices should be contrasted with reactive CO₂ “scrubbing” devices intended for short-duration missions (e.g., James and Macatangay 2009).

According to Junaedi et al. (2012), “Nickel is the traditional Sabatier catalyst that has been extensively investigated, while ruthenium was reported as the most active catalyst with the highest selectivity toward CH₄.” Thompson (2015) reports that low-cost nickel (Ni), cobalt (Co), iron (Fe), and molybdenum (Mo) catalysts can be used in the Sabatier reactor, “although these metals generally require higher temperatures to achieve reasonable kinetics.” Much more expensive catalysts include ruthenium (Ru), palladium (Pd), rhodium (Rh) and platinum (Pt), which can operate at lower temperatures. For space applications, performance rather than cost is of prime consideration (A. Meier, pers. commun. 2017), so designs typically incorporate these more expensive catalysts.

Catalyst deactivation is a significant concern for some Sabatier reactor designs, with the microchannel device of Brooks et al. (2005) experiencing deactivation after only a few hundred hours (A. Meier, pers. commun. 2017). By comparison, negligible deactivation must occur after ~12,000 h to fulfill the requirements of propellant production over 480 days as specified for the MAV in DRA 5.0 (Kleinhenz and Paz 2017).

Propellant production designed for space applications using a microchannel Sabatier reactor has much lower mass (~10 kg) compared with the water electrolyzer (~50 kg) (Brooks et al. 2005). This system was sized for a robotic sample-return mission requiring 16% of the DRA 5.0 propellant mass. A larger device, sized for DRA 5.0, requires 200 kg for the water electrolyzer, but an unreported (i.e., negligible) mass for the Sabatier reactor (Kleinhenz and Paz 2017). The microchannel Sabatier device also required a RWGS reactor, whose mass at half-scale for a robotic

mission (~8% of DRA 5.0) was reported to be 0.5 kg (Holladay et al. 2007); thus, a full-scale RWGS reactor for DRA 5.0 would be ~6.0 kg, much smaller than either the Sabatier reactor or electrolyzer.

As an exothermic reaction, the Sabatier reactor consumes essentially no energy, whereas the water electrolyzer consumes 24 kW (Kleinhenz and Paz 2017). The RWGS reactor, being endothermic, does require some thermal energy; Muscatello and Santiago-Maldonado (2012) reported that thermally coupling it with a Sabatier reactor reduces net heat generated by 37%, a clever way to recycle excess heat.

17.1.4 *Opus 12 Device*

The Opus 12 reactor is a proton exchange membrane (PEM, also known as a polymer electrolyte membrane) CO₂ electrolyzer; it operates at a temperature range of 20–55 °C and pressure range of 1–400 bar. Cooling the system to be within the operating temperature range can be done by the circulating water in the system and a heat exchanger within the water reservoir. The system tends to operate at the higher end of the pressure range to support downstream processes such as gas separations. The PEM electrolyzer is made of a polymer ionically conductive membrane as the electrolyte that separates the two distinct catalytic regions called the anode and cathode. The Opus 12 PEM CO₂ electrolyzer has many of the same parts as a PEM water electrolyzer, the main exception being the cathode. The stack within a PEM CO₂ electrolyzer can be changed out to produce one of many carbon products, including CO, CH₄, C₂H₄, and ethanol. In each case, there is a stoichiometric amount of O₂ produced at the anode.

The system does not currently experience catalyst deactivation in the way the Sabatier reactor does with coking. Nonetheless, according to Opus 12, it does experience deactivation under certain conditions that are still being investigated.

PEM water electrolyzers can run for 50,000 h with 97% uptime. Annual maintenance is recommended but not required. This represents Opus 12's target for the PEM CO₂ electrolyzer that is still under development.

Figure 17.2 shows an image of the Opus 12 team with a complete EC system.

17.2 CO₂-to-Propellant Device Design

17.2.1 *Overview*

The first part of the project was to develop the balance-of-system components needed to convert the raw output from the Opus 12 EC cell into usable products. The EC cell produces a variety of reduced products at the cathode, including CH₄, C₂H₄, H₂, CO, ethanol and small amounts of other light (C₁–C₃) hydrocarbons such as formate,



Fig. 17.2 The Opus 12 team posing *with* their commercial electrochemical device

acetaldehyde, etc.¹ Opus 12 expects that once in a commercial system, there would be steady-state operation with a stable product yield distribution, though start-up and shut-down operations would likely have a different and less stable distribution.

The initial approach was intended to produce nearly pure CH_4 as the primary fuel, but our research revealed that it is quite difficult technically to separate CH_4 and C_2H_4 , so our design centered on producing a well-defined mixture of CH_4 and C_2H_4 that could be used with O_2 as a rocket propellant, with nearly the same specific impulse (I_{sp}) as $\text{CH}_4 + \text{O}_2$, around 360 s (see Sect. 17.2.6).

Figure 17.3 shows the major components of the CO_2 -to-propellant device, consisting of a pressurization pump, EC cell, water-gas shift (WGS) reactor (including heater, reactor stages and heat exchangers), gas dryers to remove H_2O , membrane separators, and heat radiators. Of note is the WGS reactor, which converts $\text{CO} + \text{H}_2\text{O}$ into $\text{H}_2 + \text{CO}_2$. CO is difficult to separate using membrane techniques, but relatively straightforward to convert to $\text{H}_2 + \text{CO}_2$ thermochemically, and H_2 is much easier to separate from other gases. Since CO_2 and H_2 are already present in

¹ Opus 12 currently has one set of catalysts that achieves >50% CH_4 yield (expressed by molar ratio), and another set that achieves ~40% C_2H_4 yield. For modeling purposes, we assumed a representative snapshot product distribution of 51.6, 19.6, 9.6, 9.7 and 9.4% respectively (expressed by molar ratio) or 50% CH_4 , 30% C_2H_4 , 3% H_2 , 3% CO and 14% ethanol (expressed by energy).

the product mixture and must be separated, an increase in the quantities of these gases does not add any significant complexity to the overall design.

Note that the gas dryers are designed to be cycled periodically (about once per hour) to liberate captured H₂O; the resulting ~2.5 kg/d can be either reinjected into the cell, or discarded. The physical design layout of the WGS subsystem is given in Fig. 17.4.

Approaches for removing H₂O from product gases (fuel mixture and O₂) are discussed in Sect. 17.2.7. The separation of gases is assumed to take place in at least two stages. MTR, Inc. developed a membrane-based separator design for us that simultaneously removes H₂ and CO₂, leaving most of the CH₄ and C₂H₄ intact

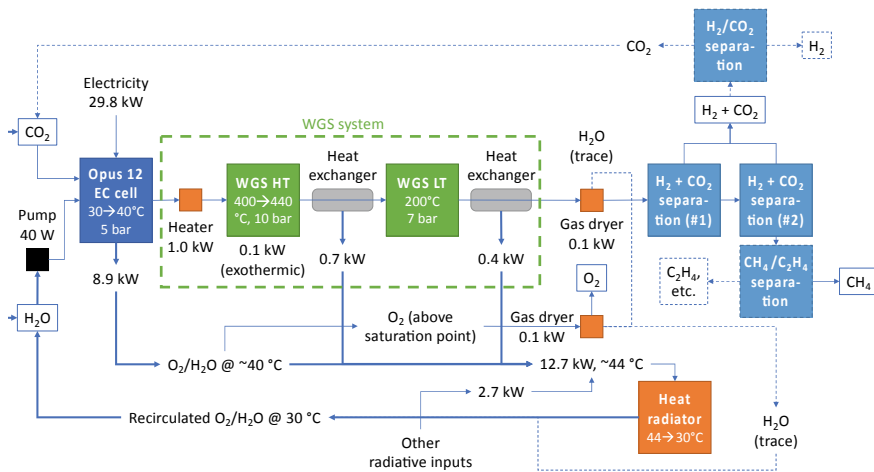


Fig. 17.3 Major components of the CO₂-to-propellant reactor design

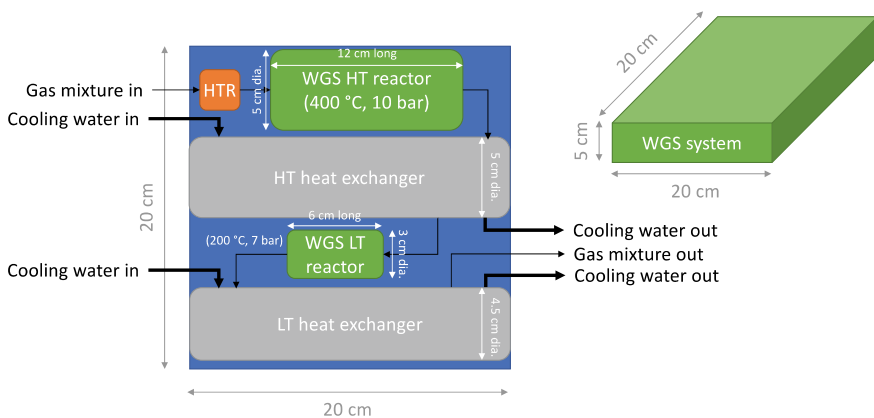


Fig. 17.4 Water-gas shift design layout

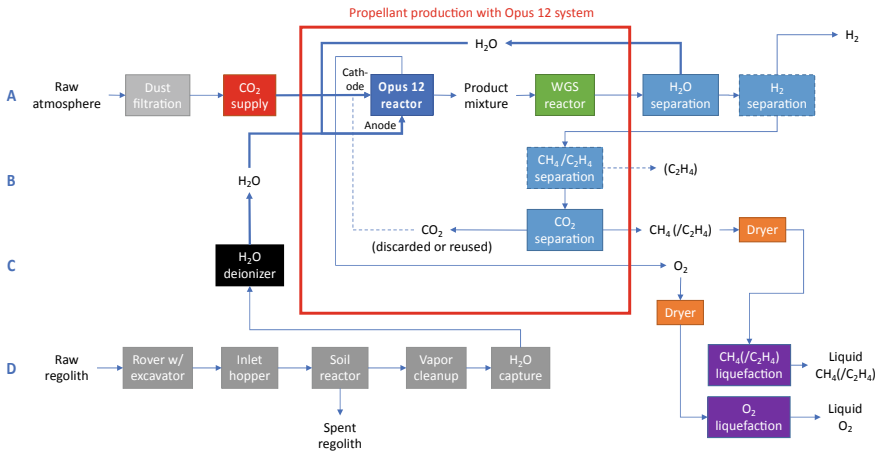


Fig. 17.5 Diagram of CO₂-to-propellant overall device design

(Tim Merkel, pers. commun. 2017; see Sect. 17.2.4). A single-stage configuration loses approximately 25% of the CH₄ and C₂H₄, whereas a two-stage configuration, requiring two compression steps, loses only 10%. We opt for the two-stage configuration due to its much lower loss rate. We assume, based on parameters supplied by MTR, that each stage requires ~1 m² of membrane area and occupies a cylinder volume of 20 cm in diameter and 2.2 cm in height.

H₂ and CO₂ can be subsequently separated with high efficiency using a standard membrane. Room for additional separation stages (e.g., H₂ from CO₂, or CH₄ from C₂H₄) with up to ~4 m² of membrane area is feasible, and is included as a contingency. These extra stage(s) would in total occupy a cylinder 20 cm in diameter and 9 cm in height. Each stage would be driven by a compact scroll-type compressor 20 cm diameter whose total height for three compression stages is ~26 cm.

Figure 17.5 shows the device integrated into a complete system that includes regolith and atmospheric extraction processes. The elements inside the red box are unique to our system, whereas elements outside it are also present in a conventional Sabatier reactor-based design (see Fig. 17.1). However, because gas separation is intrinsic to producing the desired products, we include these separation steps in our design along with the other components inside the red box.

17.2.2 Heat Rejection

The physical system layout, including the large heat radiators that are required to reject the waste heat generated by the system, is shown in Fig. 17.6. The fully stowed device including radiator panels is 66 cm long, 66 cm wide, and 60 cm high. There are four symmetrically folded panels against the sides of the central “device core”

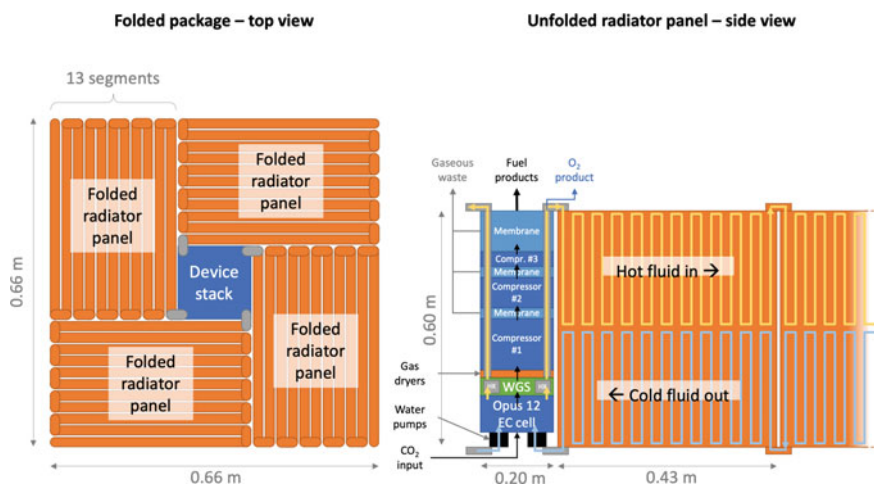


Fig. 17.6 Overall device physical layout

containing a water pump, EC cells, WGS reactor, gas dryers, and multiple separation stages. The fully extended radiator panel area is 26.9 m^2 (double-sided), configured as four panels each measuring 5.59 m long (folded into 13 segments each 43 cm long) by 60 cm high. Heat is removed from the EC and WGS reactors using water circulated through the anode side of the device. The fuel gas stream is dried prior to separation, and the O_2 gas stream is separated from the water and dried just prior to use. The H_2O stream sent to the radiator panels remains saturated in O_2 (about 0.65% by mass at 40°C), but proper coating of materials should prevent any degradation. Schedule 5 aluminum pipe is used for the radiator panel tubing, as it is the least massive yet has a maximum burst pressure of $\sim 400 \text{ bar}$ (Engineering Toolbox, no date; aluminum has a burst pressure $\sim 50\%$ that of stainless steel; TubeWeb, no date), many times higher than the expected maximum pressure of 10 bar . An outer diameter of 1.72 cm ($3/8\text{-in.}$ nominal pipe size) was chosen to keep the total pressure drop in each panel at $\sim 0.5 \text{ bar}$ including bends. The total pipe length is $1,570 \text{ m}$ (sum of all panels).

17.2.3 Water–Gas Shift System

The WGS reaction proceeds more rapidly at higher temperatures, but products are thermodynamically more favored at lower temperatures (Lima et al. 2012; Wikipedia 2017). Therefore, industrial-scale reactions are generally designed in two steps, starting with a high temperature (HT) reaction (at $\sim 400^\circ\text{C}$) which converts $\sim 80\%$ of the CO into CO_2 (and H_2O into H_2), followed by a lower temperature (LT) reaction (at $\sim 200^\circ\text{C}$) to remove all but $\sim 0.1\%$ of the CO . Pressures range from ~ 1 up to $\sim 80 \text{ bar}$;

in the system we modeled, 10 bar was assumed, following Ma et al. (2009). If the CO concentration is sufficiently low (less than ~2–4%), it may be possible to only employ the LT reactor; however, more research would be required to determine ideal reaction conditions for such a system. In our system, 5% CO by mole is assumed in the overall raw product gas stream.

A typical HT catalyst is “ferrochrome” (74–95% Fe_2O_3 or Fe_3O_4 + 5–10% Cr_2O_3 , with small amounts of Cu or other oxides, e.g., K, Mg, Zn, Pb, Al), whereas typical LT catalysts are Cu/Zn/Al or Cu/Ce/La oxides (Callaghan 2006; Ma et al. 2009; Lima et al. 2012; Wikipedia 2017). Scaling down from an industrial-sized design (Ma et al. 2009) handling 1,500 mol/s CO to ~2.3 mmol/s implied by our parameter assumptions (total product flow rate of 24 mmol/s and ~10 mol% CO), we calculate required dimensions of a 3.8 cm (inner diameter) by 11 cm (length) cylinder for the HT reactor, and a 2.7 cm (inner diameter) by 7 cm (length) cylinder for the LT reactor. Ma et al. assumed catalyst densities of ~1 g/cm³ for both reactors, implying available volumes of ~80% of total (empty) cylinder volumes. With assumed temperatures of 400 and 200 °C, and pressures of 10 and 7 bar, we calculate residence times of 7.9 and 2.6 s in the HT and LT reactors, respectively.

To manage the temperature changes of the WGS reactor, we include an electric resistance heater to raise the temperature of the raw product gas mixture from ~40 to 400 °C (see Fig. 17.3). The mass of this heating unit was estimated to be ~0.5 kg, based on compact metal-ceramic heaters (Thorlabs 2018) also used in the gas dryer design. Using the estimated enthalpies for this temperature change of each gas in the mixture (including the vaporization of H_2O), the required power is ~990 W. Because the reaction is exothermic, an additional ~100 W are generated by the reaction, with ~80 W in the HT reactor and ~20 W in the LT reactor. Therefore, a total of ~1.1 kW must be removed from the system to return it to 40 °C. This is accomplished by circulating H_2O from the anode side of the EC reactor through both the HT and LT reactors with sufficient contact areas to lower the temperatures by the desired amounts. Two heat exchangers are included to remove the heat after each reactor stage, whose dimensions are based on necessary contact areas and an assumed overall system heat coefficient of 200 W/m²-K.

At these elevated temperatures, we needed to ensure that the maximum pressure ratings would not be exceeded. Note that the additional heat provided by the exothermic reaction raises the HT reactor temperature to ~450 °C. For the HT reactor at this temperature, our assumed 1.5-in. Schedule 80 stainless steel pipe has a maximum pressure rating of 21.6 bar. For the LT reactor at 200 °C, our assumed 1-in. Schedule 40 stainless steel pipe has a maximum pressure rating of 25.9 bar. Because of concern over the use of stainless steel in contact with H_2 , we opt for carbon steel, whose maximum rated pressures are 80% those of stainless steel (TubeWeb, no date). The resulting modified pressure ratings still exceed the expected maximum 10 bar pressure by more than 70%.

17.2.4 $H_2 + CO_2$ Membrane Separation Subsystem

MTR, Inc. provided Emerging Futures with calculations of a preliminary membrane separation of $CH_4 + C_2H_4$ from other product gases, assuming CO was first shifted to $H_2 + CO_2$ via the WGS reactor (Tim Merkel, pers. commun. 2017) (see Fig. 17.7).

The design separates nearly 100% of H_2 , CO_2 , ethanol and H_2O , but did not focus on separation of CH_4 from C_2H_4 . The figure summarizes the design, which included a compression stage that also removed 91% of the ethanol and 96% of the water prior to membrane separation. Starting from an initial gas mixture of 50.7% CO_2 , 23.7% CH_4 and 9.5% C_2H_4 by mole, the resulting product stream contained 64.7% CH_4 and 30.5% C_2H_4 , with most of the remaining gas being CO_2 (3.7%). The total membrane area required for our assumed 14.6 kg/day CH_4 flow rate (about 3.8 times the modeled flow) was $\sim 1.9 \text{ m}^2$; see Fig. 17.8. However, about 25% of the CH_4/C_2H_4

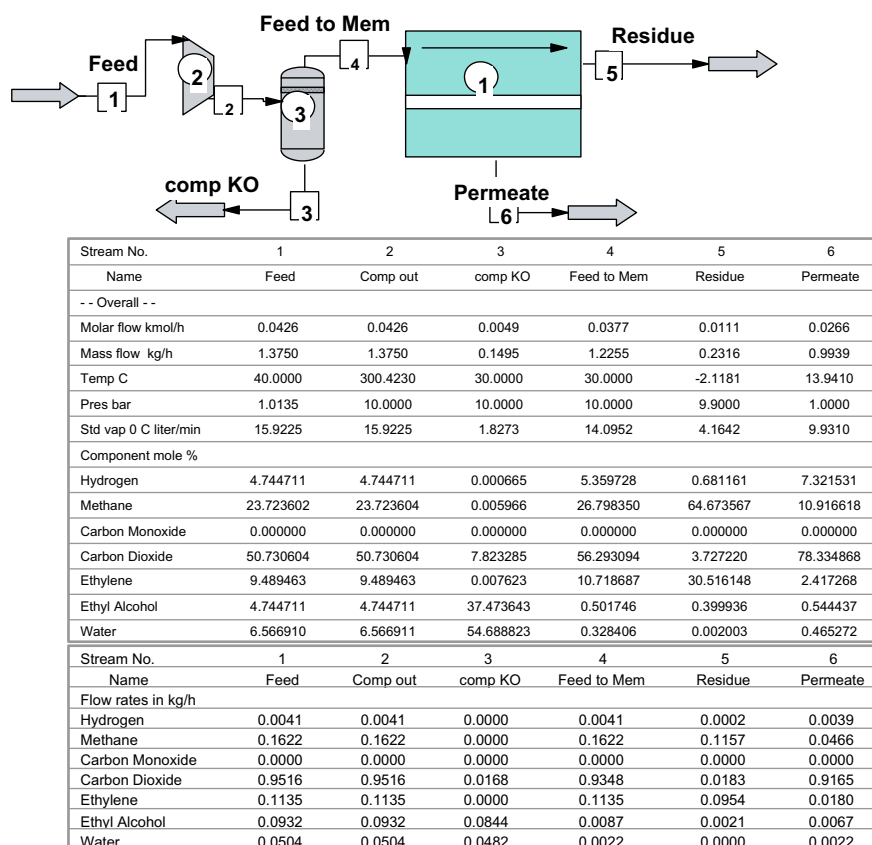


Fig. 17.7 Design of product gas separation provided by Tim Merkel (MTR, Inc.). Used with permission

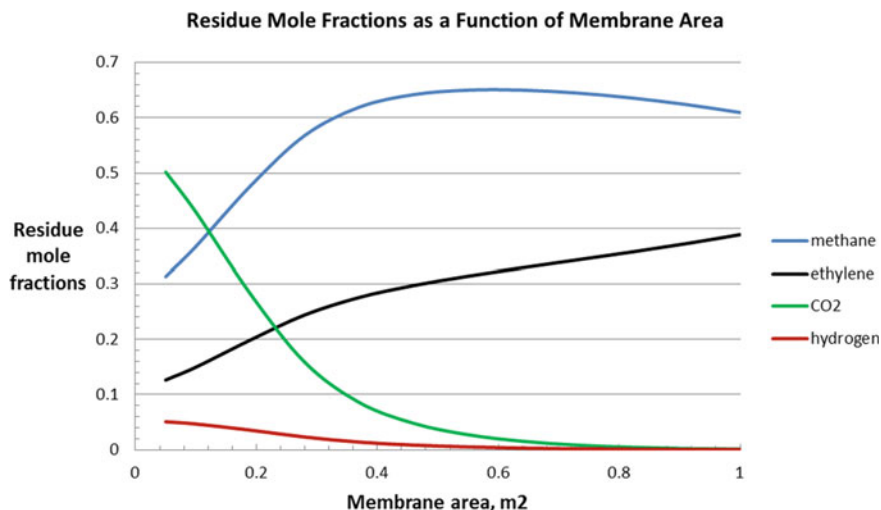


Fig. 17.8 Effect of membrane area on residue mole fractions

mixture is lost in the permeate stream; MTR indicated that a two-stage separation design could reduce loss to ~10%.

17.2.5 Challenges of CH₄/C₂H₄ Separation

Beginning with the MTR design presented above, which separates nearly 100% of H₂, CO₂, ethanol and water, but did not separate CH₄ from C₂H₄, we investigated options for subsequently separating these key products. We have identified at least three membrane-based approaches that could accomplish the required separations:

1. We found that olefin/paraffin (e.g., C₂H₄/C₂H₆) separations are quite challenging, with fluorinated polyimide membranes offering the best performance for C₃H₆/C₃H₈, with mixed-gas relative permeabilities (e.g., maximum separation efficiency) of ~6 (Sanders et al. 2013) to ~15 (Yoshino et al. 2003). However, worse performance is expected for C₂H₄/C₂H₆ due to the smaller size of the molecules, e.g., the trend is toward increased separation efficiency as molecular size increases. Non-fluorinated polyimide membranes offer pure-gas CH₄/C₂H₆ relative permeabilities of ~5 (Baker 2012). We were not able to find estimates of CH₄/C₂H₄ separation performance, but expect it to be low, since both gases are more permeable than C₂H₆ by about the same ratio.
2. CH₄/C₂H₄ relative permeabilities of ~10 may be possible using a hydrated slurry approach: by cooling an aqueous solution containing both CH₄ and C₂H₄ to nearly 0 °C, both gases form hydrates (nonstoichiometric crystalline compounds composed of water and gas) but C₂H₄ hydrate formation is favored over CH₄

hydrates, enabling effective separation, “no matter whether the mole fractions of methane in the mixture were low or high” (Pan et al. 2015). Members of this research group have found improved results using a wet zeolitic imidazolate framework (ZIF-8 or ZIF-67) to separate CH₄ from C₂H₄ hydrates (e.g., Zhang et al. 2015; Pan et al. 2016), with >68% C₂H₄ recovery (Pan et al. 2016). However, the use of such an approach is complicated by the introduction of a liquid solvent and the need to cool the system.

3. Kang et al. (2006) have explored the use of Ag ions to reversibly coordinate olefins such as C₂H₄ or C₃H₆, with estimated relative permeabilities for C₂H₄/C₂H₆ of ~1,000. While still at a research stage, this suggests that more effective membrane separation approaches may be possible in the next few years. As the Ag ions selectively bind olefins, we expect permeabilities of non-olefins such as CH₄ and C₂H₆, and therefore separation efficiencies with respect to C₂H₄, to be similar.

Starting with the assumed product stream from the MTR system (by mole ~65% CH₄, ~30% C₂H₄ and ~5% CO₂ and other gases), using a membrane with a separation efficiency of ~10 or less will require multiple stages of separation to achieve 95% CH₄ purity. Note that each separation stage also requires significant pressurization (with an upstream/downstream pressure ratio that is ~4–5 times the relative permeability; Baker 2012) to maximize separation efficacy. Lower overall pressures can be used, but the trade-off is larger membrane area and hence system mass. Clearly, a high-efficiency separation such as described in (3) above would offer superior performance, provided the permeability of the target gas is sufficiently high.

In the absence of such a membrane, we concluded that we must use the resulting gas mixture (primarily CH₄/C₂H₄ with small amounts of other fuel gases and CO₂) in lieu of pure CH₄, as rocket engines can be modified to use such mixtures with little impact on performance, so long as the gas mixture is consistent over time; see Sect. 17.2.6. Alternatively, the CH₄/C₂H₄ mixture could first be passed through the polymerization reactor (e.g., to produce polyethylene), since CH₄ is inert in this reaction (Hiemenz and Lodge 2007) and so C₂H₄ would be preferentially stripped from the input gas, resulting in an almost pure CH₄ output stream.

Assuming a target gas permeability of 1 Barrer ($\sim 3.34 \times 10^{-16}$ mol Pa⁻¹ s⁻¹ m⁻¹) and a separation efficiency of 20, the required pressure ratio would be ~100. Therefore, with an assumed initial pressure of 10 bar, inlet gas flow rate of 54 mmol/s, and active membrane thickness of 100 nm, the required membrane area is 16 m². The total membrane thickness is much larger, typically 100 μm; assuming a fill factor of 50% (e.g., 5,000 m² per 1 m³ volume), the required volume of this system is therefore 0.0033 m³. Assuming similar dimensions for the H₂ and CO₂ separations, the total membrane volume required would be ~0.0099 m³. Assuming polyimide membranes with density of 1,420 kg/m³, the required mass of each separation is 4.7 kg, for a total mass of 14.1 kg.

Note that the MTR base-case design loses about 25% of the CH₄/C₂H₄ mixture in the permeate stream, but a two-stage separation design could reduce this loss to ~10% and so would be preferable. Also of note is that the MTR design does not attempt to

Table 17.1 Calculated vacuum specific impulse (I_{sp}) for various $\text{CH}_4/\text{C}_2\text{H}_4$ and fuel: O_2 mass ratios

Fuel: O_2 mass ratio	100 wt% CH_4	90 wt% CH_4 10 wt% C_2H_4	75 wt% CH_4 25 wt% C_2H_4	57 wt% CH_4 43 wt% C_2H_4
	I_{sp} (s)			
2.0	316.89	319.96	324.63	330.29
2.5	338.99	341.50	345.10	349.10
3.0	352.31	353.80	355.74	357.53
3.5	357.47	357.70	357.56	356.44
4.0	353.56	352.31	350.31	347.78

Assumptions: 1000 psi, 93% combustion efficiency. Source: Gerald Sanders, NASA Johnson

separate the H_2 from CO_2 ; if it is important to recover the H_2 for use elsewhere in the system, or to recycle the CO_2 , a different separation approach would have to be employed.

17.2.6 $\text{CH}_4/\text{C}_2\text{H}_4 + \text{O}_2$ I_{sp} Calculations

We utilized a NASA model to calculate the specific impulse (I_{sp}) of combustion of various mixtures of CH_4 and C_2H_4 with different fuel: O_2 ratios, assuming 1,000 psi and 93% combustion efficiency. Gerald Sanders (NASA Johnson), in a private communication in September 2016, provided the results which are summarized in Table 17.1.

For a 3.5 fuel: O_2 mass ratio that maximizes the I_{sp} for pure CH_4 , we found almost no difference (<0.1%) in I_{sp} between a 100% CH_4 and 90 wt% $\text{CH}_4/10$ wt% C_2H_4 mixture. Even for a 57 wt% $\text{CH}_4/43$ wt% C_2H_4 mixture, the difference in I_{sp} was <0.3%. Therefore, there is no real performance impact of using $\text{CH}_4/\text{C}_2\text{H}_4$ mixtures. In fact, for lower fuel: O_2 mass ratios, the addition of C_2H_4 appears to improve the I_{sp} ; for instance, for a fuel: O_2 mass ratio of 3.0, the highest I_{sp} is achieved with a 57 wt% $\text{CH}_4/43$ wt% C_2H_4 mixture.

17.2.7 Moisture Removal from Product Gas Stream

The natural gas industry employs at least two methods of removing residual water vapor from gases (Hoskins 2017). The first method involves lowering the gas temperature to cause the water to condense into a liquid or solid, whereas the second method passes the gas mixture through a desiccant that absorbs the moisture; the material is then recharged through heating, driving off the water vapor. Solid desiccants

include silica gel, molecular sieves, activated alumina or activated carbon, whereas the preferred liquid desiccant is triethylene glycol.

Both methods require cycling to liberate the captured H₂O. The advantage of the first method is that it does not require a specialized material that could degrade over time; moreover, on Mars it is straightforward to achieve the required temperatures to condense water simply by passing the gas loop outside of the device into the ambient cold atmosphere. However, such a system is cumbersome, requiring additional piping and heat exchange contact area with the atmosphere.

The second method, while not without drawbacks, has the advantage of being relatively compact, and many desiccants can be regenerated at temperatures of 175–315 °C. It is estimated that such systems require an energy input of ~6.0 MJ per kg of H₂O removed, assuming 70% heating efficiency (Sigma-Aldrich 2017). These desiccants are not known to degrade, so can be used for many years of service.

While either method will work in principle, we have opted for the latter approach, utilizing a molecular sieve desiccant such as Type 3A (0.6 K₂O: 0.4 Na₂O: 1 Al₂O₃: 2 SiO₂) from Sigma-Aldrich. This material, with a dry bulk density of 720 kg/m³, can absorb ~20% H₂O by weight and is regenerated at a temperature between 175 and 260 °C. We have estimated the following H₂O flow rates and required desiccant parameters for the fuel and O₂ sides of the system. The H₂O content of each gas stream was calculated as follows. For fuel, we make a conservative estimate of the remaining H₂O after the WGS reaction. For O₂, we use the vapor pressure of H₂O at 40 °C (0.073 bar) compared with the pressurized O₂ gas stream (5 bar) (see Table 17.2).

An hourly recharge cycle will greatly reduce the mass of desiccant required, so we opt for this configuration in our design. Approximately 0.5 kg of material is needed in total. In addition, the design requires heaters to liberate the H₂O during an hourly recharge cycle lasting 2.4 min., resulting in 96% uptime. Due to the short cycling time, the peak power draw of the heaters is sizable, about 4,400 W. We have identified a commercially available, flat metal-ceramic heater that provides the required energy input across the desiccant area when arranged in a double-layer (top and bottom) configuration, with a maximum ambient temperature of 400 °C (Thorlabs 2018). The heating elements add 2.5 kg, bringing the total mass of the moisture removal system to 3.0 kg. We estimate the combined physical dimensions of the two dryers to be 20 cm by 20 cm by 2.1 cm, arranged side by side with the fuel dryer occupying 11.6 cm width and the O₂ dryer the remaining 8.4 cm width (see Fig. 17.9).

17.2.8 Mass and Energy Budgets

A mass breakdown of the complete system is summarized in Table 17.3.

Note that this design produces the target CH₄ output rate of 14.6 kg/day or 7.03 t after 480 days (plus sufficient O₂) to resupply propellant to the NASA DRA 5.0 MAV (Kleinhenz and Paz 2017). It also produces 9.7 kg/day C₂H₄, 0.77 kg/day H₂

Table 17.2 Desiccant requirements to remove moisture from fuel and O₂ gas streams

Parameter	Fuel	O ₂	Total	Units
Gas flow rate	1.71×10^{-3}	1.45×10^{-3}	3.16×10^{-3}	kg/s
H ₂ O content	1.0%	0.8%	0.9%	% mass
H ₂ O flow rate	1.71×10^{-5}	1.22×10^{-5}	2.93×10^{-5}	kg/s
	1.47	1.06	2.53	kg/d
Required desiccant (24 h cycle time)	7.4	5.3	12.7	kg
Required desiccant (1 h cycle time)	0.31	0.22	0.53	kg
Required power (average)	102	73	176	W
<i>1-h cycle time design</i>				
Required power (2.4 min. peak)	2,558	1,835	4,394	W
Desiccant volume	4.26×10^{-4}	3.06×10^{-4}	7.32×10^{-4}	m ³
Heating element volume	7.25×10^{-5}	5.20×10^{-5}	1.25×10^{-4}	m ³
Total volume	4.99×10^{-4}	3.58×10^{-4}	8.57×10^{-4}	m ³
Desiccant mass	0.31	0.22	0.53	kg
Heating element mass	1.45	1.04	2.49	kg
Total mass	1.76	1.26	3.02	kg
<i>Dimensions</i>				
Length	0.200	0.200	0.200	m
Width	0.116	0.084	0.200	m
Height	0.021	0.021	0.021	m

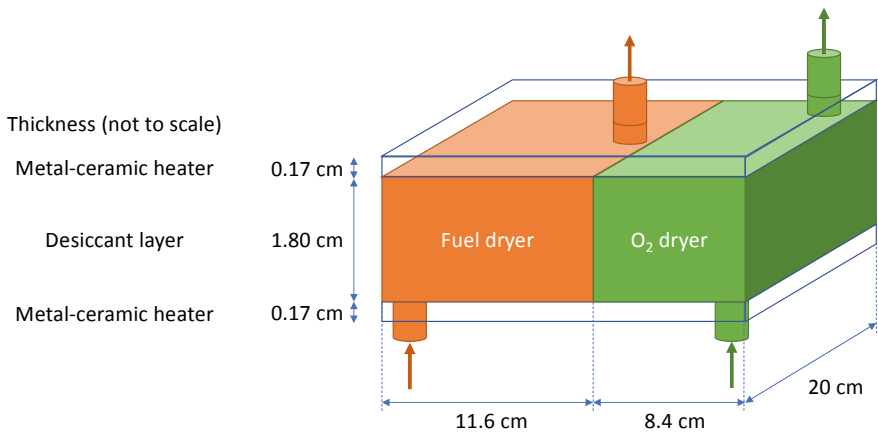


Fig. 17.9 Gas dryer design

Table 17.3 Mass breakdown of CO₂-to-propellant system

Component	Mass (kg)	Component	Mass (kg)
Water pump	0.9	Gas dryers	3.0
EC stack	124.0	Gas compressors	11.6
WGS reactors	1.1	Membrane separation stages	5.9
Heat exchangers	0.3	Radiator panels	264.9
WGS heater	0.5	Total	412.2

Table 17.4 Energy breakdown of CO₂-to-propellant system

Component	Energy (kW)
Electricity in	29.75
Products out (70% efficient)	20.83
Thermal energy from electricity	8.93
Incident from Sun, ground and sky on device	0.01
WGS energy to remove	1.13
Gas dryer energy	0.18
Pump energy	0.00
Incident from Sun, ground and sky on radiators	1.92
Total energy radiated	12.69

and 77.25 kg/day additional O₂. If a CH₄/C₂H₄ mixture were used instead, the mass could in principle be reduced to 249 kg while still providing sufficient propellant. We have also explored the sensitivity of the system mass to several key parameters (see Sect. 17.3.4).

The energy breakdown of the system is summarized in Table 17.4.

17.3 CO₂-to-Plastics Reactor Design

17.3.1 Overview

The second part of the project was to develop a design to convert the C₂H₄ produced in the Opus 12 EC cell into polyethylene. HDPE as well as low-density polyethylene (LDPE) can be made from C₂H₄, and both have a number of desirable attributes. However, HDPE requires far milder pressures (10–80 atm) than LDPE, which requires >1,000 atm. Reaction temperatures for making HDPE are also modest (80–150 °C),² and the polymer is produced below its melting point. As a result, we decided

² Note that both CH₄ and C₂H₄ become supercritical at temperatures well below this range above ~50 bar. The impact of potentially supercritical CH₄ on the C₂H₄ polymerization process is unclear, but should be investigated in follow-on work.

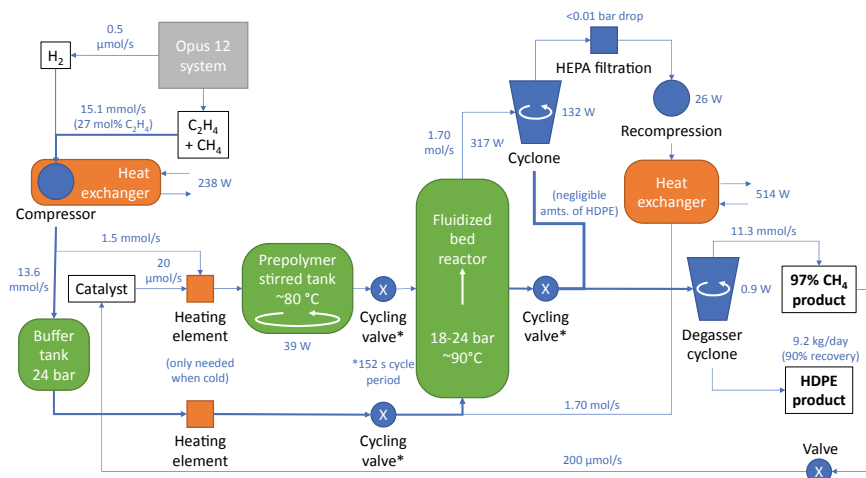


Fig. 17.10 Diagram of CO₂-to-plastics reactor design

to focus exclusively on HDPE, a common and versatile plastic that can be used to fabricate many useful objects including gas and liquid containers, tools, furniture, etc., as well as high-quality radiation protection.

Figure 17.10 shows the major components of our CO₂-to-plastics reactor design. While outside the scope of our design, the resulting HDPE product would likely be stored as small granules in bags, or perhaps formed into solid blocks to avoid the need for containers. To make a desired material, the polymer could be re-melted before being cast, extruded, powderized, or drawn into filaments for 3D printing.

We assume that the Opus 12 PEC cell produces a mixture of C₂H₄ and CH₄; the latter species does not participate in the polymerization reaction and so can be tolerated even if it constitutes >50% of the gas volume. Small amounts of CO₂ may also be present, but as this gas also does not react with other chemical species under the conditions of the polymer reactor, it is considered (along with CH₄) to be an inert buffer gas. The reactor also requires a small amount of H₂ (0.01 mol% of C₂H₄) which acts to terminate the polymerization reaction, as well as a polymerization catalyst (0.5 mol% of C₂H₄).

The resulting polymer product removes nearly all the C₂H₄ from the input gas stream, resulting in a final gas mixture that is 97% CH₄. A small portion (~2%) is used as an inert buffer gas to entrain catalyst particles into the reactor; the rest can be used as a nearly pure CH₄ product, potentially for use as propellant in a rocket engine.

The additional heat load that must be removed from the CO₂-to-plastics reactor is modest, about 750 W (see Sect. 17.3.3). If integrated with the same heat radiator system as described earlier, the radiator area would increase 7% to 28.7 m², add an additional 19 kg to the system mass, and increase the device footprint to 72 cm by 72 cm.

17.3.2 Reactor Design Details

The input gas mixture is compressed to ~24 bar, with the majority sent to fill a buffer tank that is periodically injected into the fluidized-bed reactor every ~150 s. A portion of this flow (~10% or ~1.5 mmol/s) is injected into a prepolymerization stage, along with catalyst (~20 $\mu\text{mol/s}$) and 97% CH_4 (~200 $\mu\text{mol/s}$) that acts as a carrier gas for the catalyst. The prepolymerization stage is important to maintain a consistent molecular mass of product, though a range of molecular masses is inevitable and can actually enhance polymer mechanical properties. A very small amount of catalyst (0.005–0.5 mol% of polymer product according to Mun 2002; Evertz et al. 1992 reported <0.0001 mol%) is needed to produce small polymer “seed” particles. To increase the accessible surface area, the prepolymer can be treated with *n*-hexane to remove low molecular weight polymer, but given the challenges of providing additional reagents on Mars, we assume this step is skipped at the expense of lower porosity and subsequently greater catalyst mass needed; here we assume the upper end (0.5 mol%) in our model. The recommended temperature is between 40 and 115 °C; we assume 80 °C in our model. Reagents are heated with a small resistive heat element prior to entry into the prepolymer reactor.

When the buffer tank has reached full capacity of 0.57 mol C_2H_4 (total gas mixture is 2.08 mol) after 152 s, the cycling valve is opened, emptying the contents of the buffer tank into the fluidized-bed reactor. If the reactor is starting from cold conditions, the gas mixture is heated to ~90 °C before entering the reactor; however, the reaction itself is exothermic, producing heat in excess of what is needed to maintain the reaction conditions, so the heating element is not needed under normal operational conditions. Prepolymer product is meanwhile being continuously fed into the fluidized-bed reactor to attain full-sized polymer chains with additional C_2H_4 and ~70 μmol of hydrogen (H_2) gas fed through the bottom of the reactor. H_2 must be included as a co-reagent to control polymer molecular mass: with an average chain length for HDPE of ~7,000 to ~100,000 C_2H_4 units (Reusch 2013), and one H_2 molecule needed per HDPE molecule to terminate the chain, the $\text{H}_2/\text{C}_2\text{H}_4$ ratio is ~0.001 to ~0.014%; we assume the upper end of this range in our model. Note that because H_2 reacts much less frequently with the catalyst than C_2H_4 , the concentration of H_2 in the reactor is ~5 mol% (Fernandes and Lona 2000) despite its much lower flow rate.

There are a number of catalyst choices available for making HDPE. For gas-phase reactions, the Phillips catalyst (chromium(II) bis(cyclopentadienyl) oxide or variants) supported on silica/alumina is preferred. Other alternatives are Ziegler–Natta catalysts, which have many variations but are generally TiCl_4 supported on MgCl_2 , with triethyl (or other trialkyl) aluminum co-catalyst added separately to the reactor feed. However, Ziegler–Natta catalysts are generally used in liquid solvent systems. For the reactor design specified here, we assume a pure chromium-based Phillips catalyst; though Mun (2002) indicates that triethylaluminum is also used in the prepolymer stage of their design, it was not employed here. Clearly, due to the complexity of these catalysts, they must be brought from Earth, but fortunately, only

very small amounts are required: a five-year supply for a system producing 9.2 kg/day of HDPE with 90% uptime requires only 140 kg.

As mentioned above, unlike the continuous flow design described in Mun (2002), we opt for a batch mode design in order to achieve the secondary objective of high CH₄ purity in the product gas stream. Because CH₄ does not participate in the polymerization reaction, it is enriched with each pass of the mixture through the reactor, as C₂H₄ is converted into solid polymer. By allowing the reaction to proceed without injecting more C₂H₄/CH₄ mixture into it, the C₂H₄ (and very small amount of H₂) is progressively depleted, leaving almost pure CH₄. In continuous flow designs, the reactor gas flow rate is ~2–8 times the minimum flow required for fluidization (Fernandes and Lona 2000; Mun 2002) and ~50 times (1.70 mol/s) that of the reagent input rate. We assume a similar flow rate in our batch design, and estimate that the per-pass conversion rate is 2% of the amount of C₂H₄ remaining in the gas stream. Therefore, as C₂H₄ is depleted, the reaction rate slows. We calculate that 114 passes are required to achieve 90% C₂H₄ conversion. At this point, the total pressure has dropped to 18 bar due to consumed C₂H₄.

The reactor dimensions are estimated based on ratios suggested in Mun (2002) to be 7.4 cm in diameter and 61 cm long, for a total internal volume of 2,612 cm³. Assuming 24 bar initial pressure and 90 °C operating temperature (see below), the gas mixture will contain 0.57 mol C₂H₄, 1.51 mol CH₄, trace amounts of other gases, as well as the 0.07 mol of solid prepolymer. (It may also contain a ~ few mol% of CO₂ depending on the upstream separation process, but the presence of CO₂ will not have any effect on the reaction and so can be considered part of the inert fraction with CH₄). The total gas mixture is 2.08 mol initially. Moreover, the prepolymerization reactor consumes an additional 0.23 mol of gas, for a total of 2.31 mol. Therefore, at the Opus 12 reactor output rate of 4.15 mmol/s C₂H₄, it requires 152 s to refill the reactor. However, rather than taking this entire period to refill the reactor after each batch, we assume that the reactor is purged and refilled quickly from a buffer tank, which then slowly fills again over the next period. We assume that it takes ~14 s to accomplish this cycling, leaving 139 s left to run the batch reaction itself. With 114 required passes, each pass takes 1.219 s, which is consistent with the calculated cycle rate of the reactor.

The temperature of the fluidized-bed reactor must operate between 70 and 115 °C; we choose 90 °C in our model. Both the prepolymerization and fluidized-bed reactor stages generate considerable heat (396 W in total) liberated as a byproduct of polymerization, which is more than sufficient to heat the incoming gases to the desired working temperatures. The heat remaining as well as waste heat from other processes (gas compression, cyclonic separation—see below) is dissipated in the heat exchanger after the reactor.

The unreacted gas mixture exits the top of the fluidized-bed reactor where it passes through a cyclone, a type of centrifuge designed to remove any fine polymer particles suspended in the gas stream. Solid particles fall to the bottom of the cyclone, where they are combined with the bulk of the polymer solids coming out of the reactor.

While it is more efficient to send the gas flow through the heat exchanger before the compressor, newer designs have switched the order to reduce fouling of the cooler by

fine polymer particles. However, such designs still require periodic servicing, which is not a realistic option for our design because it is not expected that astronauts will be present to perform maintenance; the plant must run for several years without a shutdown. Therefore, we have included a secondary high-efficiency particulate air (HEPA) filtration stage after the cyclone in order to remove all remaining particle fines and eliminate the need for maintenance. HEPA filters are designed to remove >99.97% of particles $\geq 0.3 \mu\text{m}$ in diameter, and they perform even better than that at smaller particle sizes (Wikipedia 2018). It is not expected that the HEPA filter will need frequent replacement, but it could be easily replaced by an astronaut or even, conceivably, a robotic system.

After the batch cycle concludes, a second cycling valve opens to purge the resulting mixture from the reactor tank (adding the small amounts of solid particles recovered from the cyclone) and passing it through a degasser (basically a solid/gaseous separator) to produce the HDPE and enriched CH_4 products. The production rates are 9.2 kg/d of HDPE and 11.3 mmol/s of gas mixture (97% CH_4 , 3% C_2H_4). Small amounts of this enriched CH_4 mixture are used as an inert buffer gas for entraining the powdered catalyst during injection into the prepolymer phase, as described above. We assume 200 $\mu\text{mol/s}$ of 97% CH_4 is used.

Note that the polymer production rate is more than twice the maximum requirement (2–4 kg/d). This is because we assumed a scaled-up reactor sufficient to provide 15.2 kg/d CH_4 , slightly higher than the 14.6 kg/d required by the DRA 5.0 MAV. If the device is not also required to produce propellant, it can be scaled down significantly.

17.3.3 Mass and Energy Budgets

The mass budget for the CO_2 -to-plastic reactor is summarized in Table 17.5. The total mass of this system (17 kg) plus the additional required radiator panel mass (19 kg) is <10% of the mass of the CO_2 -to-propellant system (412 kg). While also easily fabricated on Earth, the tremendous mass leverage afforded by making HDPE on Mars (producing its own mass in HDPE in 3.7 days) can provide valuable cost savings and design flexibility. Therefore, for a small increment in total mass, the system could be enhanced to provide not only ~9 kg/day of high-quality HDPE plastic, but also >95% CH_4 for use as rocket propellant.

The overall heat rejection budget for the polymerization system is summarized in Table 17.6.

Fernandes and Lona (2000) claim that a sufficiently high gas flow rate in the fluidized-bed reactor will prevent polymer melting and hence build-up of solid material on reactor walls. However, according to Mun (2002), large chunks of polymer can be still be formed within the fluidized-bed reactor, potentially blocking the recycle gas flow or even the entire polymerization zone, unless the reactor is shut down and the sheets are removed. Although our design mostly avoids build-up of polymer fines that could coat the inside surfaces of critical equipment (such as valves, heat exchanger

Table 17.5 Mass budget of CO₂-to-plastics system

Component	Mass (kg)	Component	Mass (kg)
Initial compressor	1.88	HEPA filter	1.64
Prepolymer tank	0.30	Precompression heat exchanger	0.84
Buffer tank	1.93	Main heat exchanger	0.84
FB reactor	2.35	Recompressor	0.21
Cyclone	7.03	Degasser	0.05
		Total	17.06

Table 17.6 Heat budget of CO₂-to-plastics system

Item	Heat load (W)
Initial gas compression from 1 to 24 bars	238
Dissipated by first heat exchanger	-238
Prepolymerization stage	
• Heat liberated during reaction	43
• Heat absorbed in raising temperature of C ₂ H ₄ /CH ₄ mixture	-4
Fluidized-bed reactor stage	
• Heat liberated during polymerization	352
• Heat absorbed in raising temperature of C ₂ H ₄ /CH ₄ mixture	-35
Main cyclone	132
Gas recompression from 18 to 24 bars	26
Dissipated by second heat exchanger	-514
Degasser cyclone	1
Total heat generated by CO₂-to-plastics reactor	754

or compressor), periodic human maintenance may still be necessary, though it is as yet unclear how frequently such maintenance may be required.

17.3.4 Sensitivity Analysis

We have explored the sensitivity of the system mass to several key parameters shown in Table 17.7. As expected, higher EC efficiency and/or CH₄ share lowers scaled system mass. Moreover, raising the device operating temperature can result in a substantial reduction in system mass.

We also explored the effect of changing the average surface and sky temperatures on Mars, since these are dependent on location as well as season. Note that the

Table 17.7 Sensitivity analysis

Variable	Value	System mass (kg)	Radiator panel area (m ²)	Total pressure drop (bar)	Pipe inner diameter (m)	Scaled system mass ^a (kg)
EC efficiency (% electrical energy in products)	50%	676.91	40.96	0.573	0.0180	950.79
	55%	631.55	37.44	0.433	0.0180	806.44
	60%	586.00	33.88	0.317	0.0180	685.92
	65%	540.65	30.34	0.224	0.0180	584.15
	70% ^b	412.19	26.86	0.509	0.0147	413.55
	75%	377.62	23.28	0.319	0.0147	353.60
	80%	343.19	19.71	0.181	0.0147	301.28
	85%	306.33	16.16	0.401	0.0112	253.10
CH ₄ share (% energy of total products) (sum of CH ₄ + C ₂ H ₄ is constant at 80%)	30%	411.73	26.85	0.509	0.0147	688.47
	40%	411.91	26.85	0.509	0.0147	516.58
	50% ^b	412.19	26.86	0.509	0.0147	413.55
	60%	412.27	26.85	0.509	0.0147	344.69
	70%	412.45	26.85	0.509	0.0147	295.58
Device operating temperature (K)	293	512.06	36.95	0.701	0.0147	513.74
	303	456.94	31.38	0.595	0.0147	458.44
	313 ^b	412.19	26.86	0.509	0.0147	413.55
	323	376.94	23.29	0.442	0.0147	378.19
	333	347.23	20.29	0.385	0.0147	348.37
Mars surface temperature (K)	175	389.89	24.60	0.467	0.0147	391.17
	200	398.87	25.51	0.484	0.0147	400.18
	224 ^b	412.19	26.86	0.509	0.0147	413.55
	250	435.87	29.25	0.555	0.0147	437.30
	275	473.73	33.07	0.627	0.0147	475.29
Mars sky temperature (K)	100	405.30	26.16	0.496	0.0147	406.63
	125	407.43	26.37	0.500	0.0147	408.77
	153 ^b	412.19	26.86	0.509	0.0147	413.55
	175	418.24	27.47	0.521	0.0147	419.62
	200	429.47	28.60	0.543	0.0147	430.88
Pipe schedule (schedule number and wall thickness in cm)	5, 0.12 ^b	412.19	26.86	0.509	0.0147	413.55
	10, 0.17	581.92	26.80	0.193	0.0171	583.83
	40, 0.28	698.86	26.80	0.284	0.0158	701.16
	80, 0.37	854.76	26.86	0.536	0.0139	857.57
Pipe inner diameter (cm)	0.85	376.33	29.40	14.685	0.0085	377.56
	1.12	413.35	27.17	2.393	0.0112	414.71
	1.47 ^b	412.19	26.86	0.509	0.0147	413.55

(continued)

Table 17.7 (continued)

Variable	Value	System mass (kg)	Radiator panel area (m ²)	Total pressure drop (bar)	Pipe inner diameter (m)	Scaled system mass ^a (kg)
	1.80	495.39	26.79	0.150	0.0180	497.02
	2.34	501.04	26.79	0.035	0.0234	502.69
	3.01	505.70	26.79	0.008	0.0301	507.37

^aRelative to CH₄ target production rate of 14.6 kg/day

^bBase case

freezing points of CH₄ and C₂H₄ lie below the lowest Mars surface temperature explored in these sensitivity runs, but not below the lowest Mars sky temperature. While we assume that all parts of the device would be maintained at or above 40 °C (the operating temperature of the EC cell), detailed thermal modeling will be required to ensure that no component is exposed to cold ambient temperatures to the degree that they would adversely affect device performance.

Changing the Mars surface temperature by ±50 K had a larger effect than changing the sky temperature by a similar amount. Because the average Mars sky temperature in the base case is ~75 K lower than the surface temperature, radiator panel absorption of infrared radiation from the sky is much lower than from the Martian surface, resulting in a lower sensitivity when this parameter is varied.

In two other sets of sensitivity runs, the pipe inner diameter was varied along with the EC efficiency sensitivity runs while keeping the total pressure drop at a reasonable level (<1 bar). In other sensitivity runs, the pipe diameter was kept fixed, except for the two sensitivities that explicitly explored its variation. The effect of increasing the pipe schedule (which increases pipe thickness) along with a simultaneous variation in pipe diameter to maintain acceptable pressure drop simply served to increase system mass. Changing just the pipe inner diameter while keeping pipe schedule fixed changed system mass and pressure drop in opposite directions. For the two smallest pipe diameters, the pressure drop was unacceptably high (>1 bar).

17.4 Conclusions and Future Directions

We have developed a preliminary device design that can produce 14.6 kg/day of CH₄ and 47.9 kg/day of O₂ from CO₂ and H₂O Martian resources, sufficient to refuel a NASA DRA5.0 MAV in 480 days. In fact, there is 77.3 kg/day of O₂ left over for other purposes. The device also produces 9.7 kg/day of C₂H₄ that can either be converted into HDPE plastic in a separate reactor, or mixed with CH₄ to provide 66% more rocket propellant with almost no change in I_{sp}. In addition, the device produces 0.77 kg/day of H₂. The overall mass of the combined propellant + plastic product system is 449 kg. A five-year supply of catalyst adds another 140 kg.

There is a possibility of using the Martian atmosphere for convective cooling similar to forced-air cooling on Earth; other related devices (e.g., a Mars nuclear reactor; Morrison 2018) have successfully incorporated this concept into their design. Based on conversation with their designer, we estimate that such an approach would require an additional ~ 1.3 kW of power but may shrink the radiator cooling area significantly. Additional work is needed to fully validate this approach.

Acknowledgements We would like to thank the following people for contributing to the final version of this paper: Dr. John Hogan (NASA Ames Research Center), Dr. Anne Meier (NASA Kennedy Space Center), Dr. Tim Merkel (MTR, Inc.) and Dr. Daniel Miller (Lawrence Berkeley National Laboratory). We also gratefully acknowledge Dr. Etosha Cave (Opus 12) for selecting Emerging Futures for this project, and for working closely with us throughout the 18-month duration, as well as NASA for providing funding through contract number NNX17CJ02C. Finally, we are grateful to Dr. Angel Abbud-Madrid (Colorado School of Mines) for accepting a version of this work for presentation at the 2019 Space Resources Roundtable—Planetary and Terrestrial Mining Sciences Symposium (SRR-PTMSS), and to Viorel Badescu, Kris Zacny, and Yoseph Bar-Cohen for providing us with the opportunity to include an account of the work in this *Springer Handbook of Space Resources*.

References

- Baker, Richard W. 2012. Chapter 8: Gas separation. In: *Membrane technology and applications*, 3rd ed., 325–378. West Sussex, UK: Wiley
- Brooks, K.P., S.D. Rassat, and W.E. TeGrotenhuis. 2005. Development of a microchannel in situ propellant production system, final report, Pacific Northwest National Laboratory, U.S. Department of Energy Grant No.: DE-AC05-6RL01830, September.
- Callaghan, C. 2006. Kinetics and catalysis of the water-gas-shift reaction, PhD thesis. <http://www.che.utah.edu/~ring/DesignII/Articles/Water%20Gas%20Shift%20Kinetics-c%20callaghan%20hesis.pdf>.
- El Sherif, Dina, and James C. Knox. 2005. International Space Station Carbon Dioxide Removal Assembly (ISS CDRA) concepts and advancements, 2005-01-2892. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050210002.pdf>.
- Engineering Toolbox. No date. Bursting and collapsing pressures of ASTM A312 stainless steel pipes. https://www.engineeringtoolbox.com/stainless-steel-pipes-bursting-pressures-d_463.html. Accessed 6 Sept 2018.
- Evertz, Kaspar, Roland Saive, Guido Funk, Peter Koelle, Rainer Konrad, and Hans Gropper. 1992. Phillips catalyst and its use for the preparation of ethylene homopolymers and copolymers. U.S. patent number US5352658A, 13 February. <https://patents.google.com/patent/US5352658>.
- Fernandes, F.A.N., Lona, L.M.F. 2000. Fluidized bed reactor for polyethylene production. The influence of polyethylene prepolymerization. *Brazilian Journal of Chemical Engineering* 17 (2). <https://doi.org/10.1590/S0104-66322000000200004>, <http://ref.scielo.org/n6pzs6>.
- Götz, M.J., F. Lefebvre, A.M. Mörs, F. Graf. Koch, et al. 2016. Renewable power-to-gas: A technological and economic review. *Renewable Energy* 85: 1371–1390.
- Hiemenz, Paul C., and Timothy P. Lodge. 2007. *Polymer chemistry*, 2nd ed. CRC Press.
- Holladay, J.D., K.P. Brooks, R. Wegeng, J. Hu, J. Sanders, and S. Baird. 2007. Microreactor development for Martian in situ propellant production. *Catalysis Today* 120 (2007): 35–44.
- Hoskins, David. 2017. How to remove moisture from natural gas. *Sciencing*, 25 April. <https://sciencing.com/remove-moisture-natural-gas-8756813.html>. Accessed 14 Sept 2018.

- James, John T., and Ariel Macatangay. 2009. Carbon Dioxide—Our common “Enemy.” <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090029352.pdf>.
- Junaedi, Christian, Kyle Hawley, Dennis Walsh, Subir Roychoudhury, Morgan B. Abney, and Jay L. Perry. 2012. Compact and lightweight sabatier reactor for carbon dioxide reduction. In *42nd international conference on environmental systems (ICES)*. American Institute of Aeronautics and Astronautics, AIAA 2012-3482, 15–19 July, San Diego, CA. <https://arc.aiaa.org/doi/abs/10.2514/6.2012-3482>.
- Junaedi, Christian, Kyle Hawley, Dennis Walsh and Subir Roychoudhury, Morgan B. Abney, and Jay L. Perry. 2014. CO₂ reduction assembly prototype using microlith-based sabatier reactor for ground demonstration. In *44th international conference on environmental systems, ICES-2014-090*, 13–17 July 2014, Tucson, AZ.
- Kang, Yong Soo, John Hak Kim, Jongok Won, and Hoon Sik Kim. 2006. Chapter 16: Solid-state facilitated transport membranes for separation of olefins/paraffins and oxygen/nitrogen. In: *Materials science of membranes for gas and vapor separation*, ed. Y. Yampolskii, I. Pinnau, and B.D. Freeman, 391–410. Wiley.
- Kappmaier, Fabian, Carsten Matthias, and Johannes Witt. 2016. Carbon dioxide reprocessing subsystem for loop closure as part of the regenerative life support system ACLS, 2016. In *46th international conference on environmental systems, ICES-2016-391*, 10–14 July 2016, Vienna, Austria.
- Kleinhenz, Julie E., and Aaron Paz. 2017. *An ISRU propellant production system to fully fuel a mars ascent vehicle*. American Institute of Aeronautics and Astronautics, *AIAA SciTech Forum*, 9–13 January, Grapevine, TX, Paper number 20170001421. <https://ntrs.nasa.gov/search.jsp?R=20170001421>.
- Lima, D.F.B., et al. 2012. Modeling and simulation of water gas shift reactor: An industrial case (full ref at end of doc). http://cdn.intechopen.com/pdfs/34193/InTech-Modeling_and_simulation_of_water_gas_shift_reactor_an_industrial_case.pdf.
- Ma, Haotian, Li Lan, Liu Yi Fei, Luo Yu, and Shan Yu. 2009. Water gas shift reactor design. Presentation for CN2116 Group 28, AY0809 Semester 2. <https://www.slideshare.net/116cn/water-gas-shift-reactor-design>. Accessed 28 Sept 2009
- Morrison, Chris. 2018. The Pylon: Near-term commercial LEU nuclear fission power for Mars and the Moon. In *Mars society 2018 convention*, Pasadena, CA, 25 August 2018.
- Mun, Tham Chee. 2002. Production of polyethylene using gas fluidized bed reactor. National University of Singapore. <https://pdfs.semanticscholar.org/ea8/47ce6cbad3a55f6e6fa7d9e9307181d1b54b.pdf> or <http://www.klmtechgroup.com/PDF/Articles/articles/Fluidized-Bed-Reactor.pdf>. Accessed 2 May 2018.
- Murdoch, Karen, Joel Goldblatt, Robyn Carrasquillo, and Danny Harris. 2005. Sabatier methanation reactor for space exploration. In: *Ist space exploration conference: Continuing the voyage of discovery*, AIAA 2005-2706, 30 January–1 February 2005, Orlando, Florida.
- Muscattello, A., and E. Santiago-Maldonado. 2012. Mars in situ resource utilization technology evaluation. In *50th AIAA aerospace sciences meeting*, Nashville, Tennessee. NASA Kennedy Space Center, Cape Canaveral, FL; AIAA-2012-360, 9–12 January. <https://ntrs.nasa.gov/search.jsp?R=20160005057>.
- Pan, Yong, Huang Liu, Zhi Li, Bei Liu, Qinglan Ma, Guangjin Chen, Changyu Sun, Lanying Yang, and Xueting Gao. 2015. Ethylene separation via hydrate formation in W/O emulsions. *Energies* 8: 4871–4881. <https://doi.org/10.3390/en8064871>.
- Pan, Yong, Chongzhi Jia, Bei Liu, Zhe Zhang, Xiongshi Tong, Hai Li, Zhi Li, Ronald Ssebadduka, Changyu Sun, Lanying Yang, and Guangjin Chen. 2016. Separation of methane/ethylene gas mixtures efficiently by using ZIF-67/water-ethylene glycol slurry. *Fluid Phase Equilibria* 414: 14–22. <https://doi.org/10.1016/j.fluid.2016.01.003>.
- Reusch, W. 2013. Polymers. Last updated 5 May. <https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/polymers.htm>. Accessed 30 May 2018.
- Sanders, David F., Zachary P. Smith, Ruilan Guo, Lloyd M. Robeson, James E. McGrath, Donald R. Paul, and Benny D. Freeman. 2013. Energy-efficient polymeric gas separation membranes for

- a sustainable future: A review. *Polymer* 54: 4729–4761. <https://doi.org/10.1016/j.polymer.2013.05.075>.
- Sigma-Aldrich. (2017). Molecular sieves—Technical information bulletin. <http://www.sigmaaldrich.com/chemistry/chemical-synthesis/learning-center/technical-bulletins/al-1430/molecular-sieves.html>. Accessed 14 Sept 2018.
- Stainless Steel (ASTM-A-312). <http://www.tubeweb.com/Non-Ferrous-Tube/theoretical-bursting-collapsing-pressures-for-pipes.html>. Accessed 23 Dec 2017.
- Thompson, John O. 2015. Scaleable, High Efficiency Microchannel Sabatier Reactor. 45th International Conference on Environmental Systems, 12–16 July 2015, Bellevue, Washington. Paper ICES-2015-240.
- Thorlabs. 2018. HT24S-24 W, 20 mm × 20 mm Metal Ceramic Heater. <https://www.thorlabs.com/thorproduct.cfm?partnumber=HT24S> and https://www.thorlabs.com/_sd.cfm?fileName=CTN002042-S01.pdf&partNumber=HT24S (specification sheet). Accessed 22 Sept 2018.
- TubeWeb. No date. Non-ferrous tube: theoretical bursting & collapsing pressures for pipe.
- Wikipedia. 2017. Water-gas shift reaction. Last updated 20 November. https://en.wikipedia.org/wiki/Water-gas_shift_reaction.
- Wikipedia. 2018. HEPA. Last edited 3 May 2018. <https://en.m.wikipedia.org/wiki/HEPA>. Accessed 7 May 2018.
- Yoshino, Makoto, Satoshi Nakamura, Hidetoshi Kita, Ken-ichi Okamoto, Nozomu Tanihara, and Yoshihiro Kusuki. 2003. Olefin/paraffin separation performance of asymmetric hollow fiber membrane of 6FDA/BPDA-DDBT copolyimide. *Journal of Membrane Science* 212 (1): 13–27. [https://doi.org/10.1016/S0376-7388\(02\)00434-9](https://doi.org/10.1016/S0376-7388(02)00434-9).
- Zhang, Xiao-Xin, Peng Xiao, Chang-Hua Zhan, Bei Liu, Rui-Qin Zhong, Lan-Ying Yang, Chang-Yu Sun, Huang Liu, Yong Pan, Guang-Jin Chen, and Nan Li. 2015. Separation of methane/ethylene gas mixtures using wet ZIF-8. *Industrial & Engineering Chemistry Research* 54 (32): 7890–7898. <https://doi.org/10.1021/acs.iecr.5b00941>.

Chapter 18

Mobile Mars Habitation



Kürşad Özdemir and Süheyla Müge Halıcı

Abstract This chapter focuses on the concept of mobile habitation on Mars. A description of Mars' surface features is followed by a review of early concepts of crewed mobility for the Moon and Mars. Wheeled concepts for crew mobility continue to be based on the success of the Lunar Roving Vehicle, and predominantly take the form of a pressurized rover on wheels. With the help of architectural diagrams, the chapter introduces a range of habitable and mobile Mars structures, and the technologies used, taking into account mission requirements.

Abbreviations

ATHLETE	All-Terrain, Hex-Limbed, Extra-Terrestrial Explorer
EMC	Evolvable Mars Campaign
ESA	European Space Agency
EVA	Extra-vehicular Activity
LSG	Liquifer Systems Group
MCO	MarsCruiserOne
MDRM	Mars Design Reference Mission
MORPHLAB	Modular Roving Planetary Habitat, Laboratory and Base
MPL	Mobile Pressurized Laboratory
NASA	National Aeronautics and Space Administration
SEED	Spherical Environment Exploration Device
SEV	Space Exploration Vehicle
SPE	Solar Particle Events

K. Özdemir (✉)
MEF University, Istanbul, Turkey
e-mail: office@fezai.com

S. M. Halıcı
Istanbul Technical University, Istanbul, Turkey
e-mail: halicis@itu.edu.tr

SPRITE Small Pressurized Rover for Independent Transport and Exploration
WW2 The Second World War

This chapter focuses on mobile habitation systems for Mars surface missions, which involves habitable structures configured to house crewed surface mission activities in different locations

The first section of the chapter introduces the concepts of mobile living, while Sect. 18.2 covers mobility on Mars and its various surface mobility systems/mechanisms. The cases for each surface mobility system are covered in Sects. 18.3 and 18.4 presents some conclusions.

18.1 The Concept of Mobile Living

Mobility is essential to human development, and has been a key asset in our ability to explore, utilize and innovate (De Haas 2009). Ranging from Conestoga wagons to nuclear submarines, mobile habitation has enabled us to project our ambitions in different fields of human endeavor. Although perfectly suited for efficiency, this type of living has its drawbacks: limited space, limited capabilities (remember the basic flushing system of your holiday trailer?), the need for uninterrupted attention and maintenance for the habitat and all its sub-systems. Nevertheless, people never fail to see and enjoy the silver lining: freedom as to your home's location and an ever-changing view from your porch. Furthermore, some people travel to exist. Nomadic groups change their location periodically. Terrestrial mobile habitation systems exist on the ground, on and under the water, and finally in the air. The configuration, quality and service intervals of terrestrial mobile habitation systems are continuously evolving due to changes in scale, capabilities and purpose: for example, compare the cramped bunk beds in a WW2 submarine to the fancy interiors of private submarines in development (Fig. 18.1).

18.2 Mars and Mobility

Mars is an obvious target for exploration because it is relatively close to us, but there are many more reasons to explore the Red Planet. The scientific reasons for going to Mars can be summarized as the search for life, understanding the surface and the planet's evolution, and preparing for future human exploration (ESA 2020).

The temperature on Mars is estimated to fall to $-143\text{ }^{\circ}\text{C}$ at the poles or reach a high of about $27\text{ }^{\circ}\text{C}$ at noon at the equator in the summer (Tillman 1997). The low density of the Mars atmosphere results in temperature mainly being governed by solar heating and infrared cooling. With a lower-density atmosphere, lack of clouds, ozone

Fig. 18.1 Conestoga wagon
(*Credit* Encyclopædia
Britannica)



and a magnetosphere, Mars is exposed to higher impact levels from solar particle events (SPE) and galactic cosmic radiation (GCR) than Earth. Periodic exposure to radiation from SPE is hazardous for humans even on short missions. Regarding threats to human health from high-energy GCR particles, the decisive parameter is mission duration (Rapp 2006). Thus, the primary reason for the protective shell of any kind of habitat is to mitigate radiation exposure for the crew (see Table 18.1). In addition, on Mars surface radiation can pose a problem for shield, electronics and instrumentation. In broad terms, Mars surface habitats are envisaged as equipped with hull-integrated shielding and in some cases with an inner “storm shelter” for SPEs.

One other factor in Mars’ surface environment relevant for habitats is dust. Dust can pose health risks to the crew beside its hazards for the habitat’s equipment and systems. According to the International Agency Working Group of ASI, CSA, ESA, JAXA and NASA for Dust Mitigation Gap, the following components of the surface exploration for the Moon and Mars need to be studied with reference to dust mitigation (IAWG 2016):

- Life support systems (LSS)
- Extra-vehicular activity (EVA) systems (including suits, airlocks, suitport, tools)
- Human health and human-system performance
- Robotics and mobility systems
- In-situ resource utilization (ISRU)
- Ascent/descent vehicles

Table 18.1 Exposure times currently accepted by NASA for manned vehicles in low earth orbit (NASA 2007)

Exposure interval	Dose equivalent (mSv)		
	Blood forming organs (BFOs)	Ocular Lens	Skin
30 days	250	1000	1500
Annual	500	2000	3000
Career	1000 4000	4000	5000

- Surface power systems
- Thermal control systems

Suitports or suitlocks as well as robotic arms and sampling boxes are included in designs for the mitigation of dust in surface outpost designs, including mobile habitats (Liquifer Systems Group 2008). Onboard analysis capability is a key enhancement in surface exploration architecture, at the cost of adding the complexity of a sample airlock and a robotic arm component to the pressurized rover system.

As far as the physical surface conditions of Mars are concerned, topographic relief is much greater than on Earth or the Moon. Highlands occur mainly in the southern hemisphere; lowland mainly in the northern hemisphere. The southern highlands are heavily cratered. Based on Viking lander and Mars Pathfinder imagery, the surface material on Mars can be divided into rocks, soil and drift material. Across the seven landing sites, the rocks that cover 8–16% of the surface range in size from pebbles to over 1 m in diameter. Moderately dense terrestrial soil with significant clay or silt-sized particles covers 80–90% of the landing sites. Drift material is very fine-grained and porous (Zakrajsek et al. 2005).

Experience from the Lunar Roving Vehicle (LRV), which has the capabilities of extended EVAs, provided a clear picture of the benefits for exploration (Connors et al. 1994). Expansion in the coverage of exploration on the lunar surface is clear (NASA 2019) from the start to the finishing line. While Apollo 11 could range up to roughly 100 m from the lander, Apollo 17 surface crew traversed a total distance of 30.5 km using their LRV. Apollo 17 astronauts gathered 110.4 kg of lunar material. Projection of the LRV experience onto mobility concepts for the Red Planet suggests that a simple wheeled concept could also be viable for Mars, with necessary adaptations taking account of the differences between the two environments (gravity, terrain conditions).

Short- and medium-term exploration of Mars' surface is likely to involve mobility of crews and hardware. In later missions, aerial vehicles may also be required (Genta 2017). These aerial vehicles will require suitable propulsion, since ballooning heavy payloads on Mars might be challenging (NASA JPL 2012). Ballooning in the Martian atmosphere is complicated by the fact that the Martian carbon dioxide atmosphere is very cold (-73°C), and it is very thin at 0.006 bar, with 1 bar = 1 atmospheric surface pressure on Earth. Balloons to be flown on Mars must be made of very lightweight material.

These ground features provide a suitable set of conditions for the use of wheeled platforms for surface mobility on Mars. Apart from the failed Mars-1 and Mars-2 landing missions and the associated Prop-M surface rovers (Perminov 1999), four-wheeled mobile platforms for extended surface exploration have conducted operations on the Red Planet. The Mars Pathfinder Mission's Sojourner rover, Mars Exploration Rover Missions' Spirit and Opportunity, and Mars Science Laboratory Mission's Curiosity rover (active as of October 2020) have all featured six-wheeled configurations with progressively alternative designs for the wheels. The design for

the NASA Mars 2020 Mission's Perseverance rover utilizes this 6-wheeled configuration with a similar suspension system and enhanced wheel profiles (NASA 2020). Accumulated knowledge and the ongoing development of these platforms suggests wheeled locomotion will be the most viable option for future surface mobility on Mars, and this approach is likely to dominate future design concepts for habitable structures traveling on Mars' surface.

The year 1964 saw one of the earliest milestones in a line of mobile habitation systems presented by Grumman (Stoff 2004). Originally designed for the lunar environment, a crewed rover concept called the pressurized Mobile Laboratory (MOLAB) also showed potential for Mars surface missions (von Braun 1969). Further designs for use on both the Moon and Mars, with the associated cabin and trailer set-up, appeared in various studies of surface mobility assets in the following phase of space exploration. The pressurized rover concept, featured in the reference mission of the Mars exploration study team (Hoffman and Kaplan 1997), is the first shirt-sleeve environment mobile asset to be included in an overall design for the planetary infrastructure, even though it was an adapted version of a lunar crewed rover design. More recent studies on Mars' surface exploration architecture present visibly different designs for mobile Mars habitation systems rather than adaptations of lunar designs (Fig. 18.2).

Designs for habitats with limited mobility have emerged in the line of NASA Mars surface architecture studies (Cohen 2015). Kent Joosten's 1993 concept for the Mars Design Reference Mission (MDRM) 1.0 Mars exploration habitat, for instance, featured wheels for a short transport ride between the landing zone and the habitat final location (Hoffman and Kaplan 1997) (Fig. 18.3).

Similar to the first study, a large pressurized rover for two is included in the Mars Design Reference Mission (MDRM) 5.0. This time, one of the depictions of the rover implies an inflatable body with a rocker-bogie suspension arrangement, indicating rough terrain traverse capability (Fig. 18.4).



Fig. 18.2 NASA MOLAB mock-up (*Credit Grumman*)

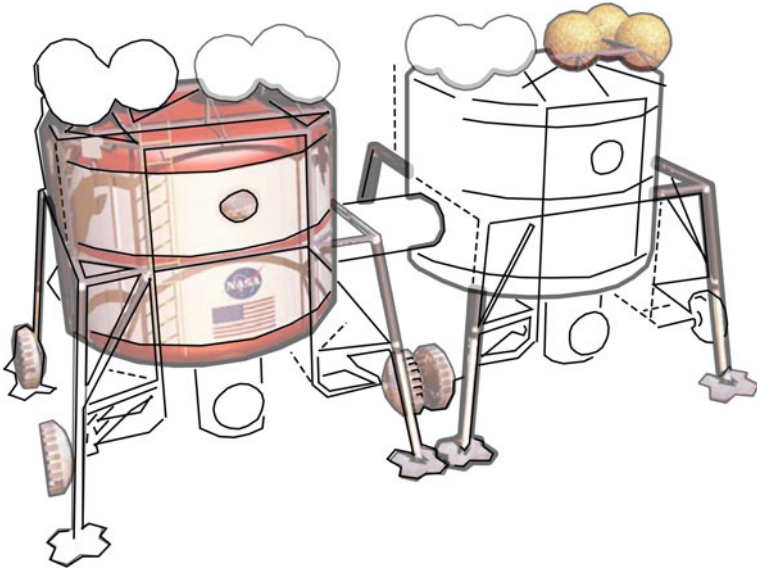
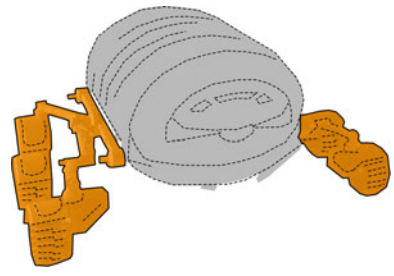


Fig. 18.3 Kent Joosten’s 1993 concept for the Mars Design Reference Mission 1.0 surface exploration habitat. Note the integrated wheels for limited mobility. (Root Image: John Frassanito and Associates/NASA; redrawn by Halıcı)

Fig. 18.4 NASA MRDM 5.0 Pressurized Rover (Root Image: Rawlings; redrawn by Halıcı)



A pressurized rover is, by definition, a mobile surface habitat capable of performing various types of missions, ranging from exploration activities to complicated construction operations (Liquifer Systems Group 2008). It serves the crew as a habitat, a refuge and a research laboratory/workshop. Although the majority of design cases in this chapter belong to the crewed rover category, mobile habitat concepts such as Mobitat2 are also included (Howe and Gibson 2006).

The range of mission parameters used in mobile habitat designs span distances, duration, terrain, type of work, and anticipated EVA activities, with crew safety as a priority factor. Various excursion scenarios, including single or multiple units, are evaluated to determine operational parameters.

Table 18.2 Design challenges for vehicles on Mars surface (Source Zakrajsek et al. 2005)

Factors	Crew	Mobility	Mass and size range	Power	Sub-system	Overall
Design challenges	Safety in extreme thermal conditions	Maneuverability in majority of terrain types	Minimal level for packaging and landing	Efficient utilization transmission	Reliable long-term operation	Ability to survive and operate in abrasive and dusty environments. Modularity for adaptation

Following an assessment of planetary rover concepts (Zakrajsek et al. 2005), a team at NASA’s Glenn Research Center produced the set of design challenges for planetary rovers shown in Table 18.2.

Any rover design also has the following benefits:

- The dual-use rover concept can significantly reduce the amount of equipment required to be placed on a planetary surface. A rover that can be both operated remotely during unmanned missions and, with some modifications, crewed in manned missions can be very efficient and economical.
- Placing rovers on the surface without a lander can save a large amount of weight that can be used to carry other supplies.
- Rovers with articulated motion capabilities will be capable of negotiating terrain in which even tracked vehicles can get stuck.
- Highly controllable manipulating arms on a rover in conjunction with a small sample airlock can reduce EVA needs and dust infiltration problems.
- Modularity, even if only within a mission scenario, can efficiently increase redundancy (Fig. 18.5).

The final beneficial factor in the list, modularity, together with a commonality approach, has become one of the key factors in recent NASA studies on Mars surface architecture. More specifically, EVA interface components such as airlocks, suitlocks and suitports appear as elements of this modularity and commonality approach in habitat designs for both mobile and fixed configurations. Unlike typical bulky airlock components, suitlocks and suitports are efficient in terms of pressurized volume and use of breathable air, speed in egress and ingress, and high level of dust mitigation. Both systems utilize rear entry spacesuits. The development of cabin pressure suits featuring hard shells and articulated joints parallels recent and upcoming advances in

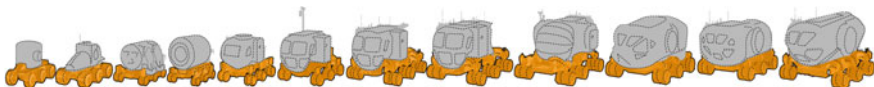


Fig. 18.5 Evolution of NASA’s Space Exploration Vehicle (SEV) (Root Image: NASA; redrawn by Halıcı)

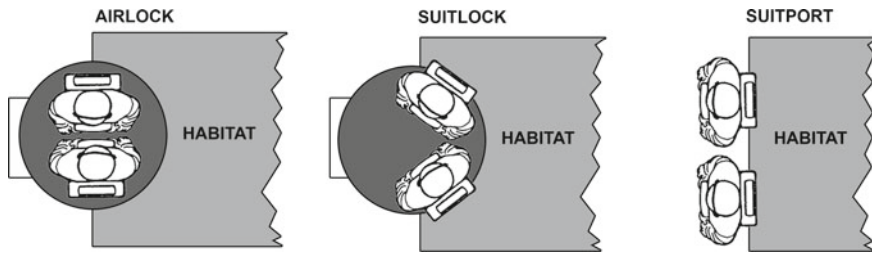


Fig. 18.6 EVA interface components (Root image: Florida Institute for Human & Machine Cognition/NASA; redrawn by Halıcı)

the EVA interfaces mentioned above (note that contemporary EVA suits have lower pressure levels than the cabin environment). While suitlocks include a pressurized or unpressurized suit compartment, suitports feature docking ports exposed to the outer environment of the vehicle. In NASA's Evolvable Mars Campaign this suitlock compartment takes the form of an inflatable module that can be expanded and retracted depending on the operational scenario. Inflatable add-on units of this kind are also featured in other design studies (Schreiner et al. 2015) as a viable option to enhance the capabilities of mobile systems for planetary exploration missions (Fig. 18.6).

Several case studies that have been carried out to assess powering of mobility assets assume two crew, four wheels, an aluminum structure and chassis, and seven-day off-base exploration missions on the Moon and Mars (Hong and Hoffman 2008). According to Hong and Hoffman (Table 18.3), fuel-cell single power systems and photovoltaic/fuel-cell hybrid systems were acceptable for short missions of only a few days. For long-range mobility for human explorers on the surface of Mars, studies suggest using combustion engines to power Martian ground vehicles with indigenous materials (Zubrin 1992). However, photovoltaic/fuel-cell hybrid systems are still required for long-duration missions (Hong, Hoffman 2008). The power requirement on Mars is higher than on the Moon because of Mars' higher gravity. Storage space for extended solar panels and technology to deploy and fold solar panels therefore would always be a better design option for Mars missions.

18.3 Mobile Habitation System Designs for Mars Missions

This section discusses habitat design concepts, including spatial design qualities and basic technical layout, and compares ten designs featuring different types of mobility.

The final report of an ESA design study (LSG 2008) for a mobile Mars habitat stated, "*Habitability* is among the important determinants for the design of any human element, including a pressurized rover for Moon and Mars exploration. Habitability largely depends on the characteristics of the physical, social and psychological environment. Habitability is also directly influenced by the design of the

Table 18.3 Power requirements for vehicles on Mars’ surface (Hong and Hoffman 2008)

	Vehicle	W (Watt)
Driving	Pressurized V	10,718
	Unpressurized V	1108
Science	Pressurized V	1993
	Unpressurized V	0
Inactivity	Pressurized V	1418
	Unpressurized V	0
Exploration	Pressurized V	547
	Unpressurized V	1108
Base	Pressurized V	0
	Unpressurized V	0

Table 18.4 Selected design concepts with context (Data Source see Sect. 18.3.)

	Vehicle
MDRM 1.0	1997 / NASA Mars Design Reference Mission 1.0
SEV	Origins in Constellation Systems, current development
EMC-Chariot	2016/NASA Evolvable Mars Campaign UPV
RAMA	2008/ESA Surface Architecture Study
MCO	2007/ESA Mars Mission Architecture Study follow-up
MSTS	1999/Design study at The University of Texas at Austin
EMC-ATHLETE	2016/NASA Evolvable Mars Campaign
MOBITAT2	2006/Design study
SEED	2015/Preliminary design work
EMC-Hopper	2016/NASA Evolvable Mars Campaign

mission elements, such as by the determination of habitable volume, configuration of habitable spaces, lighting, and other factors” (Table 18.4).

18.3.1 On Wheels, Featuring HOFFMAN, SEV, EMC, RAMA, MARS CRUISER 1

The wheel has been used in combination with an axle component specifically for transportation since the copper age (Gasser 2003). Wheeled mobility systems are believed to have appeared in Mesopotamia and Europe simultaneously. The solid board wheels of early vehicles gave way to the spoked wheels of chariots and to the Conestoga wagon, a mobile habitat on wheels. This iconic vehicle of settlement

expansion, with its simple design combining a four-wheeled wagon and two horses with a lightly covered space and a river crossing-proof hull, has also been taken as a role model for planetary surface mobility systems, with the symbolic power of the wheel appearing as a fundamental practical element of mobility.

MRDM 1.0 Pressurized Rover

By 1997, the Mars Exploration Study Team at NASA had structured a reference design mission utilizing elements that had already appeared in previous design studies. One of these designs is a pressurized rover for Mars' surface. The basic design, which is also part of NASA's lunar surface architecture, features a cylindrical pressure vessel with a cockpit and an EVA airlock as end caps on two ends of the 16.5 mt vehicle. It is designed to house a crew of two on a 20-day mission with a 500 km range (Hoffman and Kaplan 1997) (Fig. 18.7).

The mobility system of the MRDM 1.0 rover is based on four individually electric-driven cone-shaped wheels with integrated transmission. The system is designed to be fed by a dynamic isotope power system that can produce 10 kW continuous electric power for the rover. The power system is expected to be mounted on a separate trailer to be towed by the rover during the missions (Fig. 18.8).

The Conestoga wagon-style configuration of the pressurized rover contains an axial arrangement of a cockpit, a work station, galley, hygiene facilities, sleeping

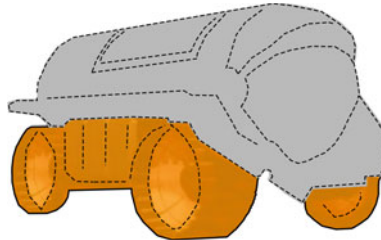


Fig. 18.7 NASA MRDM 1.0 Pressurized Rover (Root Image: John Frassanito and Associates / NASA; redrawn by Halıcı)

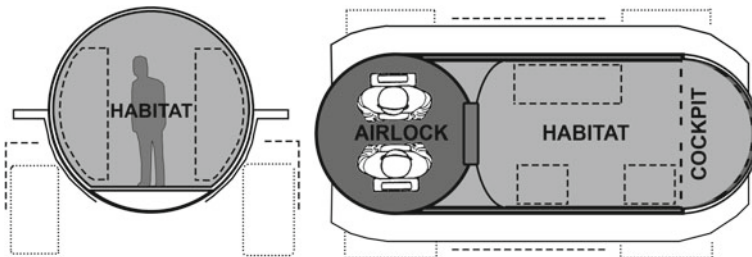


Fig. 18.8 NASA MRDM 1.0 Pressurized rover, interior layout (Credit Halıcı)

stations and an airlock at the other end of the vehicle. The chassis, featuring manipulator arms, is planned to be used during the uncrewed construction phase of Mars exploration.

SEV

NASA's Space Exploration Vehicle (SEV) concept represents a milestone in the line of vehicle designs, beginning with the Small Pressurized Rover (SPR). SEV is intended to be used both for in-space missions and for surface exploration of planetary bodies, including near-Earth asteroids, the Moon and Mars. The vehicle combines a cabin, chassis and different work package units. The cabin contains the shirt-sleeve habitable space, along with the ice-shielded lock for SPE protection, suitports and docking hatches to mate with other vehicles. The power system of SEV is thought to rely on fuel cells (Fig. 18.9).

The surface exploration version of the SEV has the cabin mounted on a chassis, with 12 omni-wheels that can pivot 360 degrees and provide the vehicle with eased lateral mobility. Unlike the Conestoga wagon-style big rovers, the SEV approach is low mass and low volume, with multiple rovers in motion instead of a big all-in-one mobile laboratory. The 3,000 kg vehicle plus the 1,000 kg payload configuration for the lunar environment can be expected to increase for the Mars version, with a heavier power system. The SEV concept features suitports providing quick egress-ingress with minimal gas loss and a chariot-style external steering post.

EMC Rover (chariot type)

NASA set a clear goal for the Evolvable Mars Campaign (EMC): "Define a pioneering strategy and operational capabilities that can extend and sustain human presence in the solar system including a human journey to explore the Mars system starting in the mid 2030s" (Moore 2016). Among other elements, EMC proposes a modular exploration system, including small pressurized cabins and mobility platforms, to provide increased exploration capabilities on Mars' surface. These mobile and habitable structures can mate with the fixed-surface habitats and function as extensions when not traveling (Fig. 18.10).

Fig. 18.9 NASA SEV pressurized rover (Root Image: NASA; redrawn by Halıcı)

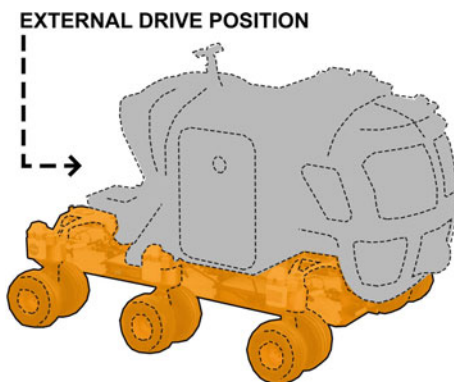
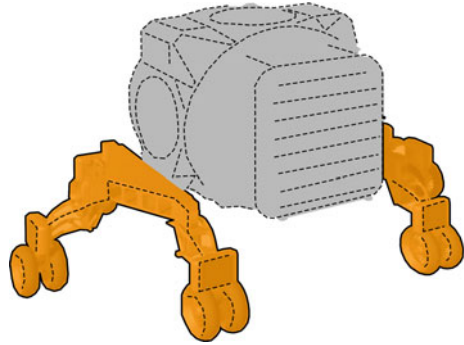


Fig. 18.10 NASA EMC Pressurized Rover with chariot-type mobility system (Root Image: Howe; redrawn by Halıcı)



Regarding the surface mobility component, the study utilizes three concepts: a spring-loaded hopper system for low-gravity environments (i.e., Phobos); the ATHLETE articulated-limb platform for extreme terrain; and a chariot-type wheeled mobility system. The chariot-equipped system resembles an EVA-oriented pressurized rover for cross-country traverse. The rear end of the vehicle features external suitports and a stowable lightweight EVA-stair component. The vehicle has a total pressurized volume of 20.7 cubic meters, of which 14.7 is habitable. The rover has a typical pressurized excursion vehicle-type cabin, in accordance with NASA's recent "common cabin approach with standard interfaces" (Gernhardt 2015) within a frame of commonality (Griffin et al. 2015). The sections of the cabin body, respectively, are the cockpit, aft cabin and the aft enclosure. EMC has a modular set-up with high degrees of compatibility between elements, and provides condition-specific mobility capabilities for near-term operations on Mars' surface.

RAMA

The ESA's Surface Architecture Study includes the RAMA pressurized rover concept by Liquifier Systems Group (LSG). RAMA resembles a mobile laboratory for surface exploration of the Moon and Mars. It is designed for surface missions lasting up to approximately 40 days with a crew of two or three, its source of energy, a liquid hydrogen/liquid oxygen fuel cell, allowing it to be driven and operated during the day as well as at night (Imhof et al. 2010). The rover allows extra-vehicular activity, and a remote manipulator is provided to recover surface samples, to deploy surface instruments and equipment and, in general, to assist the astronauts' field activities. In all cases the rover is thought to be refueled using the products supplied by an in-situ resources facility (Fig. 18.11).

Four wheels are independently propelled and steered. The wheels and their suspension system are mounted on a circular chassis. An ellipsoidal body forms the pressurized cabin. Reactant tanks are fixed to the chassis and are located on all four sides. RAMA has a launch mass of approximately 7,000 kg, a dry mass of about 6,200 kg and surface mission masses of between 7,800 and 8,300 kg (Fig. 18.12).

The RAMA concept is notable for its high level of spatial design quality and efficiency. The total integrated space concept of the mobile habitat was studied by a

Fig. 18.11 ESA RAMA pressurized rover (Root Image: LSG; redrawn by Halıcı)

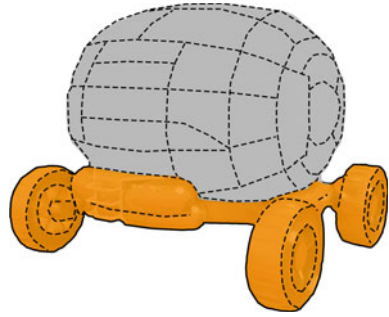
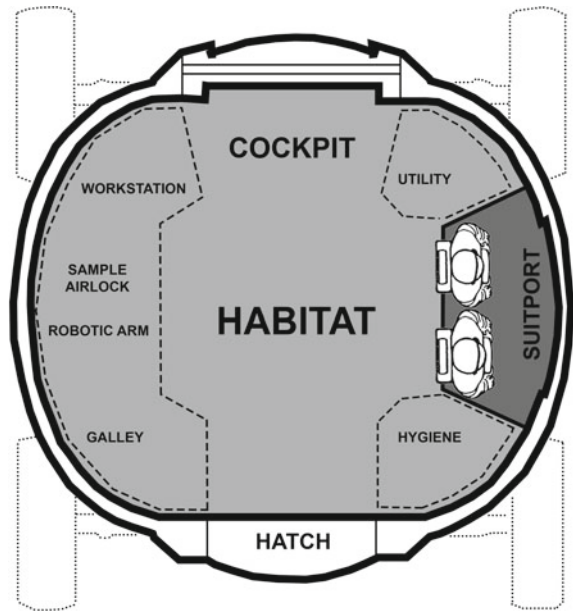


Fig. 18.12 ESA RAMA pressurized rover, interior layout. (Root Image LSG; redrawn by Halıcı)



group of space architects. Each side of the interior space has a line-up of different uses. The rover's pressurized shell is designed to be constructed from a carbon-Kevlar-glass fiber-reinforced phenolic honeycomb. The honeycomb structure is filled with polyethylene to provide additional radiation protection.

MARSCRUISERONE

As part of ESA's European Mars Mission Architecture Study, Architecture and Vision together with SRConsultancy designed the MarCruiserOne pressurized rover, seeking to optimize launch envelope vs habitable volume (Vogler et al. 2007). The design was based on Astrium's mobile pressurized laboratory (MPL) concept (Fig. 18.13).

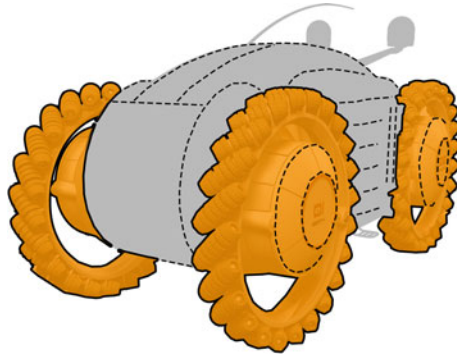


Fig. 18.13 MarsCruiserOne Pressurized Rover (Root Image: Architecture + Vision; redrawn by Halıcı)

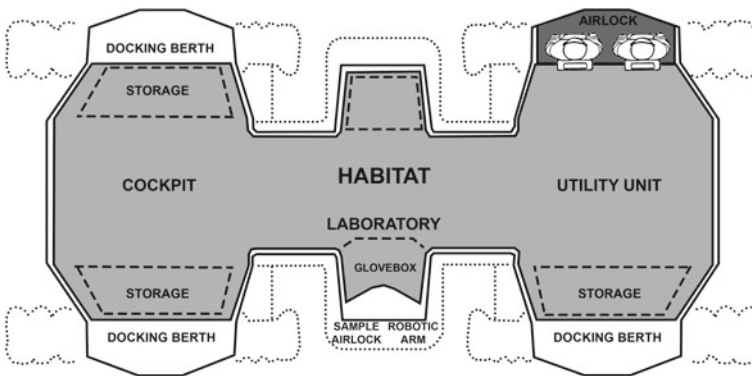


Fig. 18.14 MCO pressurized rover, interior layout (Root Image: Vogler et al.; redrawn by Halıcı)

The use of large-diameter omni-directional wheels and the elimination of the traditional chassis results in an integrated habitable body and enables decreased levels of pressure on unknown ground during exploration sorties. The rover is equipped with suitports and a flexible interior for the crew. The overall configuration is compact enough to fit into a 5 m-diameter launch and landing envelope (Figs. 18.14 and 18.15).

18.3.2 Wheels on Limbs Featuring EMC-ATHLETE, MOBITAT2, MSTs

Adjustable limbs provide motion better suited to changing terrain. The most advanced platform for the wheels-on-limbs type of mobility is the ATHLETE project. The All-Terrain, Hex-Limbed, Extra-Terrestrial Explorer (ATHLETE) system is being

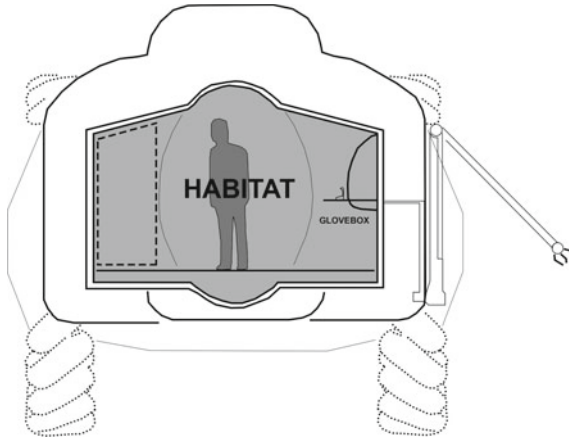


Fig. 18.15 MarsCruiserOne pressurized rover, interior layout in section (Root Image: Vogler et al.; redrawn by Halıcı)

developed by JPL as part of the Human–Robot Systems (HRS) at Johnson Space Center (Wilcox et al. 2007). The limbs of the ATHLETE have six degrees of freedom, so that the wheels can be used as feet, as well as general-purpose manipulators interacting with their surroundings (NASA 2013; NASA JPL n.d.) (Fig. 18.16).

In the early stages of research, the ATHLETE design was envisaged to support human return to the Moon, specifically to transport large masses (cargo and habitats) in the lunar environment (Wilcox et al. 2007). Subsequently, the team has focused on new developments for ATHLETE, aiming for a fully equipped re-usable mobile

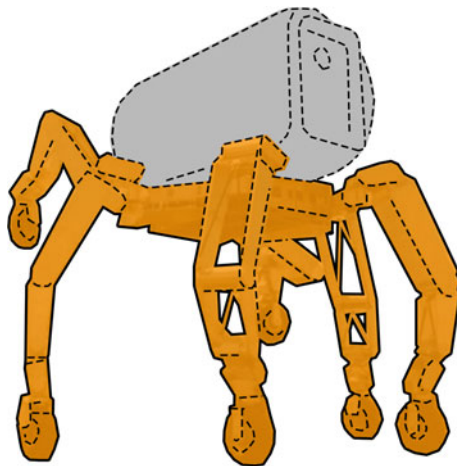


Fig. 18.16 Second-generation ATHLETE unloading cargo (Root Image: NASA; redrawn by Halıcı)

habitat on planetary surface. In consultation with the habitat team led by Larry Toups of JSC, the ATHLETE team created two “micro-habitat” mock-up shells made of graphite composite, each 2.34 m in diameter and 3.66 m long. Each has an aluminum honeycomb floor, “ring frames” that allow other outfitting (especially soft goods such as hammock-style bunks or bags of provisions) to be suspended, and solid-state programmable-color lighting (Wilcox 2009).

EMC Rover (ATHLETE-type)

An adapted version of ATHLETE can be seen in one of the EMC mobile modular system proposals. The vehicle design, intended for mobility on rough terrain, consists of a pressurized excursion vehicle-type cabin with EVA components (i.e., suitports and stowable stairs or a deployable EVA platform) and an ATHLETE-type mobility component (Fig. 18.17).

MOBITAT 2

The Mobitat2 differs from the other examples in that its design focus is on the cabin structure. Howe and Gibson’s research on a self-assembling mobile modular habitat has centered around a set of modular structural elements, the Trigon panels, that are able to self-assemble and carry specialized payload panels. It is important to note that the Trigon panels can also be configured as a pressurized rover with either four or six wheels, or other vehicles (Fig. 18.18).

The Mobitat2 has an eight-legged mobility system, with pressure ports centered on each of the four sides. The maneuverability of the legs in both folded/park mode and walking/rolling mode allows for very fine adjustment in a full six degrees of freedom for horizontal, vertical, yaw, pitch, and roll to ensure neighboring pressure ports line up for docking. A legged habitat is also able to adapt to extreme slopes or variations in the surface of the terrain. The primary mobility system uses ATHLETE-type folding legs (Fig. 18.19).

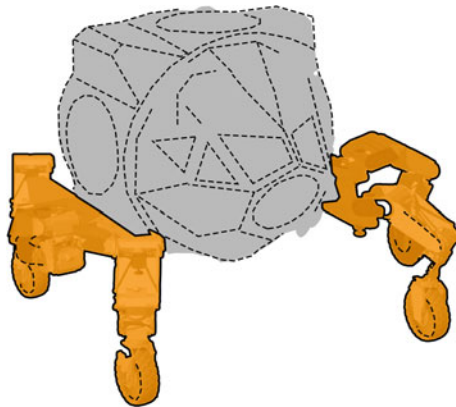


Fig. 18.17 NASA EMC pressurized rover with ATHLETE-type mobility system (Root Image: Howe; redrawn by Halıcı)

Fig. 18.18 Mobitat2 Mobile Habitat with 8-legged ATHLETE-type mobility system (Root Image: Howe & Gibson; redrawn by Halici)

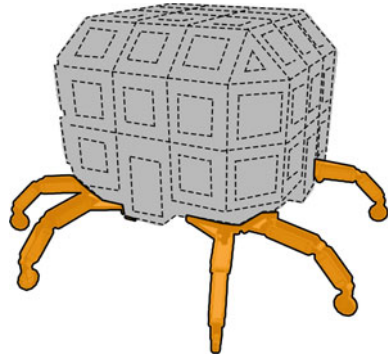
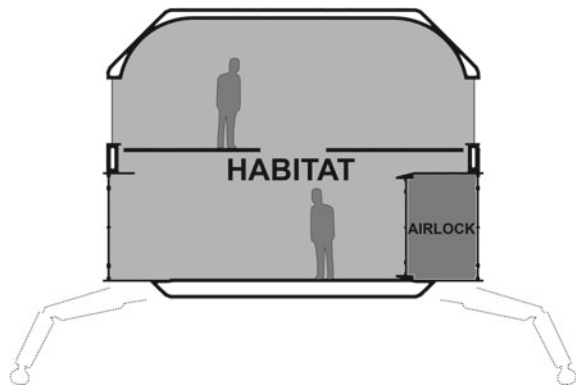


Fig. 18.19 Mobitat2 Mobile Habitat, interior layout in section (Root Image: Howe & Gibson; redrawn by Halici)



MSTS

The Mars Surface Transportation System (MSTS) is an inflatable habitat/laboratory module with a supporting space truss in the form of a crane chassis (Collins et al. 1999). The system has four sets of electrically powered trucks. The proposed inflatable habitat module features an inflatable shell of Kevlar-reinforced material. While the lightweight habitat module suggests a low-mass cabin approach and needs less storage volume for transportation, a set of possible internal configurations with rack-mounted interchangeable equipment modules offers flexibility (Fig. 18.20).

The spatial arrangement of the inflatable habitat module is linear and stacked, specifically in the crew berthing segment. The foredeck contains the cockpit. In the central section there is a habitat/laboratory area. This space contains crew berthing, placed above hardware units. The aft deck houses an EVA airlock and a mating hatch (Fig. 18.21).

Fig. 18.20 MSTS Mobile Habitat (Root Image: Collins et al.; redrawn by Halıcı)

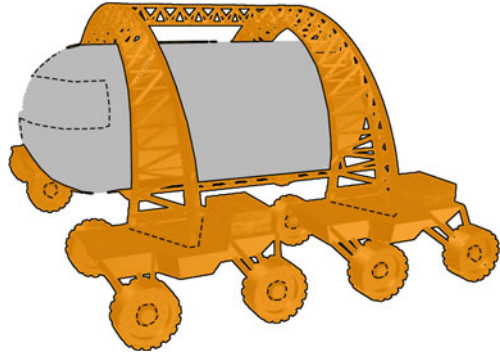
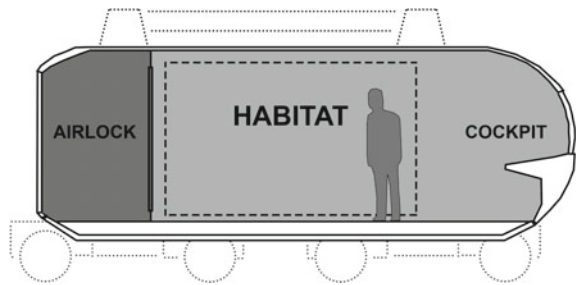


Fig. 18.21 MSTS Mobile Habitat, interior layout in section (Root Image: Collins et al.; redrawn by Halıcı)



18.3.3 *Eccentric Concepts, Featuring SEED and HOPPER EMC*

SEED

SEED (Spherical Environment Exploration Device) is a spherical-rolling habitat concept for Mars surface (Özdemir and Halıcı 2016). The concept, initially proposed at the MarsMobil Workshop in 2012, was to be developed into a preliminary design to be presented at ICES 2016 in Vienna. The unique spatial arrangement of SEED differentiates it from other selected designs. The nuclear arrangement of the SEED habitat places the crew at the core of the concept in both the operational and physical sense (Fig. 18.22).

There are three main layers, the inner sphere, the hollow crust and the outer sphere. The mobility system of the SEED is based on the powered rolling of the cushioned outer sphere. In between the hollow crust and outer sphere, there is a structural traction grid with wheels, driving the outer sphere of inflatable cushions. The inner sphere is effectively free moving on demand, with the option to provide steering data in mobility mode. The hollow crust hosts a suitport, a mating hatch and all the sub-systems of the habitat. The hollow crust serves as a storage space for utensils so the rotatable cabin is kept free for the user (Fig. 18.23).

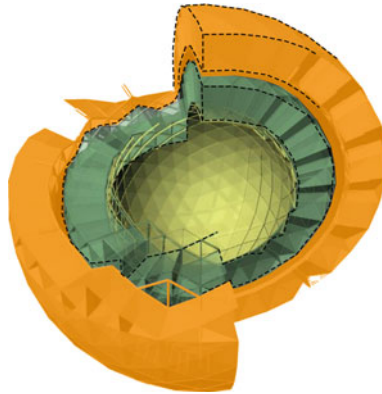


Fig. 18.22 SEED Mobile Habitat (*Credit Halıcı*)

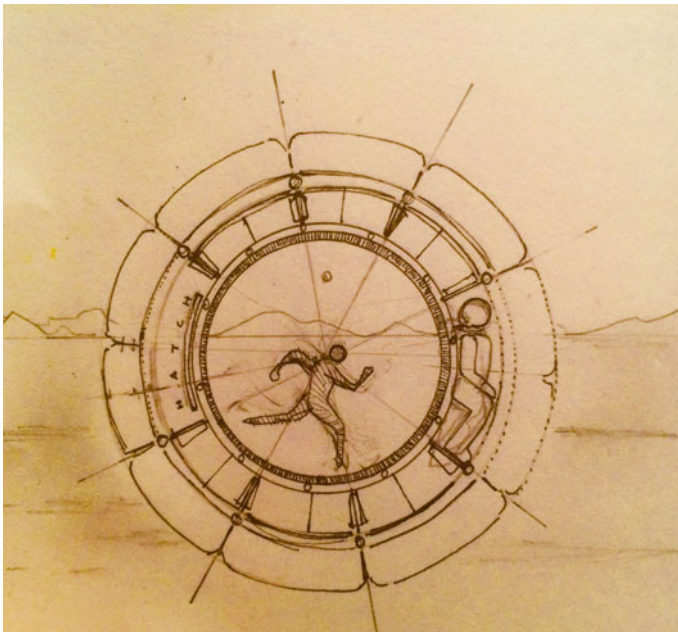


Fig. 18.23 SEED Mobile Habitat, section sketch (*Credit Özdemir*)

EMC Rover (hopper type)

Although not for the actual planetary surface of Mars, an alternative concept for the low-gravity surface of Mars' moon Phobos is included in NASA's EMC study. The vehicle features an adapted version of ATHLETE components on a hopper platform. The ATHLETE system, which includes spring actuators in its tool suite, is thought

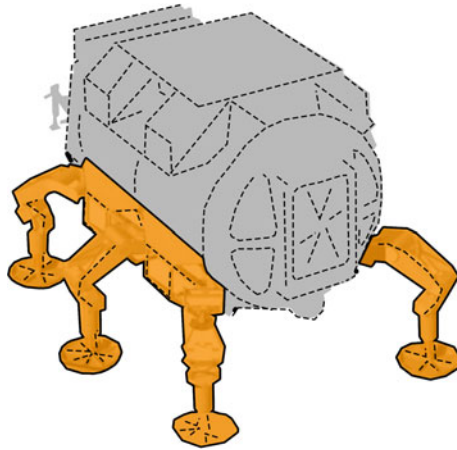


Fig. 18.24 NASA EMC Pressurized Rover with adapted ATHLETE-type mobility system for hopper mode (Root Image: Howe; redrawn by Halıcı)

to be particularly suitable as a Phobos hopper (Howe 2015). During a series of operations on Mars and Phobos, it would be possible to take the same ATHLETE fitted out as a hopper mobility system down to the Mars surface, swap the spring actuators for wheels in a matter of minutes, and proceed as a wheeled mobility platform for carrying habitat modules and other payloads (Fig. 18.24).

18.4 Evaluation

The set of concepts presented here provides a spectrum of alternative approaches to the issue of crew mobility on Mars in shirt-sleeve environments. While clumsy-looking initial designs (e.g., MRDM 1.0) provide a starting point for thinking about self-contained mobile habitats, the latest cluster of NASA EMC modular designs shifts the spotlight to efficiency in the upcoming era of human Mars missions. Though partly free from actual mission constraints, alternative designs (such as SEED and MarsCruiserOne) signal the widening of horizons in human-centered design in space (Table 18.5).

Table 18.5 Selected design concepts (*Data sources see Sect. 18.3.*)

	Mission		Mobility			Habitat		
	Crew	Duration days	Drive system	Mass and size range	Power	Cabin	Spatial arrangement	Unique features
MDRM 1.0	N/A	N/A	4-wheel chassis	16,500 kg	Dynamic isotope	Cylindrical	Linear	Conestoga wagon style, all-in-one big rover
SEV	2	14	12-wheel chassis	4000 kg	Fuel cells	SEV 1.gen cabin	Linear	Low mass/low volume
EMC- Chariot	2	14	12-wheel chassis	5442 kg	Fuel cells	SEV 2.gen cabin	Linear	Low mass/low volume
RAMA	2	40	4-wheel chassis	7800 kg	Fuel cells	ellipsoidal composite shell	Central	Compact, high-quality space
MCO	4	10	4-wheel chassis	4500 kg 4,5 × 3,8 × 9,5	Fuel cells	Monocoque shell	Linear, varying	Compact and optimized hab volume
MSTS	N/A	N/A	4 × 4-wheel chassis + crane	N/A	Fuel cells	Kevlar-reinforced inflatable membrane	Linear, stacked	Multi-use crane chassis

(continued)

Table 18.5 (continued)

	Mission		Mobility		Habitat			Unique features
	Crew	Duration days	Drive system	Mass and size range	Power	Cabin	Spatial arrangement	
EMC-ATHLETE	2	14	ATHLETE	6492 kg	Fuel cells	SEV 2.gen cabin	Linear	Low mass/low volume
MOBITAT2	N/A	N/A	Adapted ATHLETE	6300 kg	Fuel cells	HABOT-type, modular skin	Central, stacked	Modular structure
SEED	1	N/A	Whole body mobility	<2000 kg	Fuel cells	all-surface inflatable cushions, uniform interior	Nuclear	Small size, one-person excursion hab
EMC-Hopper	2	Temporary access	Adapted ATHLETE	7422 kg	Fuel cells	SEV 2.gen cabin	Linear	Low mass/low volume

Acknowledgements The authors would like to thank our reviewers, Tim Lexen and Stephen Ransom, for their kind support in improving the quality of this chapter.

References

- Cohen, M.M. 2015. First mars habitat architecture. In *AIAA SPACE 2015 conference and exposition*, p 4517.
- Collins, C. et al. 1999. Conceptual design of a mars surface Transportation system (MSTS). In *2nd Annual HEDS-UP forum proceedings*.
- Connors, M. et al. 1994. Interviews with the Apollo lunar surface astronauts in support of planning for EVA systems design. Ames Research Center. NASA-TM-108846, 1 Sept 1994.
- De Haas, H. 2009. Mobility and human development. United Nations Development Programme Human Development Reports Research Paper. http://hdr.undp.org/sites/default/files/hdrp_2009_01_rev.pdf. Accessed 15 Oct 2020.
- ESA. 2020. Why go to mars? https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Exploration/Why_go_to_Mars. Accessed 15 Oct 2020.
- Gasser, A. 2003. World's oldest wheel found in Slovenia. Government Communication Office of the Republic of Slovenia.
- Genta, G. 2017. Mobility on mars. In *Next stop mars*. Springer Praxis Books. Springer, Cham. https://doi.org/10.1007/978-3-319-44311-9_9.
- Gernhardt, M. 2015. Human exploration of phobos. https://aeronautical.org/sites/default/files/astonauts-robots/2015/astronauts-robots_2015-05-12-1515_gernhardt.pdf. Accessed 15 Oct 2020.
- Griffin, B. et al. 2015. Small habitat commonality reduces cost for human mars missions. In *AIAA space 2015*, Pasadena, CA.
- Hoffman, S.J., D.I. Kaplan. 1997. Human exploration of mars: the reference mission of the NASA Mars exploration study team. National Aeronautics and Space Administration, Lyndon B. Johnson Space Center.
- Hong, S.B., and J.A. Hoffman. 2008. Design of power systems for extensible surface mobility systems on the Moon and Mars. In: *AIAA SPACE 2008 conference & exposition*, San Diego, California, 9–11 Sept 2008.
- Howe, A.S. 2015. A modular habitation system for human planetary and space exploration. In: *45th international conference on environmental systems*.
- Howe, A., I. Gibson. 2006. MOBITAT2: a mobile habitat based on the trigon construction system. In *Collection of technical papers–space 2006 conference*.
- IAWG, International Agency Working Group. 2016. Dust mitigation gap assessment report. <https://www.globalspaceexploration.org/wordpress/docs/Dust%20Mitigation%20Gap%20Assessment%20Report.pdf>.
- Imhof, B., et al. 2010. Pressurized rover for moon and mars surface missions. In *38th COSPAR scientific assembly 2010*. <https://www.cospar-assembly.org/abstractcd/OLD/COSPAR-10/abstracts/data/pdf/abstracts/B01-0027-10.pdf>. Accessed 15 Oct 2020.
- Liquifier Systems Group. 2008. *RAMA*. Final Report, ESA: Concept Study for a Pressurized Rover.
- Moore, C. 2016. The evolvable mars campaign. https://www.nasa.gov/sites/default/files/files/3-EMC_for_NAC_Research_SubCom_Moore.pdf. Accessed 15 Oct 2020.
- NASA. 2007. *space flight human system standard*, vol. 1. Crew Health.
- NASA. 2013. ATHLETE rover busts a move: a dancing robot. <https://www.youtube.com/watch?v=BUrKVjFvz2o>. Accessed 15 Oct 2020.
- NASA. 2019. Apollo program overview. https://www.nasa.gov/mission_pages/apollo/index.html.
- NASA. 2020. Wheels and legs. <https://mars.nasa.gov/mars2020/spacecraft/rover/wheels/>. Accessed 15 Oct 2020.

- NASA JPL. 2012. Balloon development challenges for mars. https://www2.jpl.nasa.gov/adv_tech/balloons/mars_overview.htm. Accessed 15 Oct 2020.
- NASA JPL. n.d. All-terrain hex-limbed extra-terrestrial explorer. <https://athlete.jpl.nasa.gov/>. Accessed 15 Oct 2020.
- Ozdemir, K., and S.M. Halıcı. 2016. Roll SEED Roll: an architectural assessment of a spherical mobile habitat for mars (SEED_Spherical Environment Exploration Device). In *46th international conference on environmental systems*, Vienna, Austria, 10–14 July 2016.
- Perminov, V.G. 1999. *The difficult road to Mars: a brief history of Mars exploration in the Soviet Union*. National Aeronautics and Space Administration Headquarters. <https://archive.org/details/difficultroadtom00perm/page/70/mode/2up?q=rover>. Accessed 15 Oct 2020
- Rapp, D. 2006. Radiation effects and shielding requirements in human missions to the Moon and Mars. *Mars 2*: 46–71. <https://doi.org/10.1555/mars.2006.0004>.
- Schreiner, S., et al. 2015. An overnight habitat for expanding lunar surface exploration. *Acta Astronautica* 112: 158–170. <https://doi.org/10.1016/j.actaastro.2015.03.012>.
- Stoff, J. 2004. *Building Moonships: the Grumman lunar module*. Arcadia Publishing
- Tillman, J.E. 1997. Mars. http://www-k12.atmos.washington.edu/k12/resources/mars_data-information/mars_overview.html#Meteorological%20highlights. Accessed 2 June 2020
- Vogler, A. et al. 2007. MarsCruiserOne. SAE Technical Paper.
- von Braun, W. 1969. Manned mars landing presentation to the Space Task Group. https://www.nasa.gov/sites/default/files/atoms/files/19690804_manned_mars_landing_presentation_to_the_space_task_group_by_dr_wernher_von_braun.pdf. Accessed 15 Oct 2020
- Wilcox, B.H. 2009. ATHLETE: a cargo and habitat transporter for the moon. In *IEEE aerospace conference proceedings*.
- Wilcox, B.H., et al. 2007. ATHLETE: A cargo handling and manipulation robot for the moon. *Journal of Field Robotics* 24 (5): 421–434.
- Zakrajsek et al. 2005. Exploration rover concepts and development challenges. In: *First AIAA Space Exploration Conference: Continuing the Voyage of Discovery*, Orlando, Florida, 30 January–1 February 2005
- Zubrin, R. 1992. Methods for achieving long range mobility on Mars. In: *28th Joint Propulsion Conference and Exhibit*

Chapter 19

Local Resource Creation on Mars



Robert Zubrin

Abstract This chapter presents methods by which materials found on Mars can be turned into useful resources. The development of Mars can be viewed as consisting of three phases, exploration, base building, and settlement, each with its own resource development priorities. In the exploration phase, propellants and consumables are of primary interest. In the base-building phase, the production of structural materials, including steel, plastics, ceramics, bricks, fabrics, and glass becomes important. In the settlement phase the full range of industries are of interest, including especially energy production. We discuss each of these phases and the relevant chemical engineering processes to meet their needs. The material inventory of Mars readily lends itself to technologies supporting the development of each phase. In the exploration phase, CO₂ and water are the primary local materials of interest, providing the basis for the production of fuel, oxygen and food. In the base-building phase, interest expands to include making use of plentiful Martian crustal materials, especially iron oxide and silicon dioxide, enabling production of structures, including habitations and greenhouses. In the settlement phase, all materials are of interest, notably including deuterium, which is five times as plentiful on Mars as it is on Earth, and which could provide an abundant energy resource to power a growing Martian civilization.

19.1 Introduction

There is no such thing as a natural resource. There are only natural raw materials. It is human ingenuity, manifested as technology, that transforms raw materials into resources. Land was not a resource on Earth until people invented agriculture. Oil was not a resource until people developed oil drilling and refining, and devices that could operate using the resulting product. Aluminum was not a resource until people developed the technology to extract it from its oxide. Before that it was simply rock or dirt. Uranium was not a resource until people developed nuclear fission. Deuterium

R. Zubrin (✉)

Pioneer Astronautics, 11111 W. 8th Ave. unit A, Lakewood, CO 80215, USA

e-mail: zubrin@aol.com

is not a resource of any great consequence today, but it will be an enormous one once people have developed fusion energy.

Mars appears barren to many people today, just as Ice-Age Europe and Asia must have appeared to early humans migrating out of our original tropical African natural habitat. Yet, by developing new technologies, our ancestors were able to create the resources to not only sustain themselves, but flourish with ever-increasing prosperity across the entire planet. In doing so, they transformed humanity from a local biological curiosity of the Kenyan Rift Valley to a global family, hundreds of nations strong, sporting innumerable contributions to thought, literature, art, science, and technology. Having achieved such success, the task before us is to take it further, by developing the technologies to create the resources that will enable the birth and sustain the growth of new vibrant branches of human civilization on the many worlds that surround us.

Of the worlds currently within our reach, Mars possesses by far the richest assortment of raw materials for transformation into resources. It thus presents the best prospect for settlement. Furthermore, Mars offers not just one new world, but an open frontier where we will learn what we need to know to take on the rest. It is there that we will develop not only many of the key technologies, but more importantly the attitudes needed to go, settle, and create life beyond.

Taking on the challenge of Mars is thus the central task of our time.

The settlement of Mars will occur in three phases, each enabled by its own capabilities for local resource creation. These phases are:

1. *The Exploration Phase.* The activities of this phase center on field exploration to resolve central scientific questions, such as those relating to the possible origin, extent, and diversity of life on Mars, and to assess the amounts and locations of potentially useful materials to support further development. This phase will involve tens of people on Mars. Key resources that need to be created include propellants and life-support consumables.
2. *The Base-Building Phase.* The activities of this phase center on creating a base that will not only support greatly expanded exploration, but the development of an ever-expanding technological, industrial, and agricultural capability to enable human settlement. This phase will involve hundreds to thousands of people on Mars. Key resources that need to be created include structural materials such as brick, concrete, ceramics, iron, aluminum, glass, fabrics, and plastics; small-scale manufacturing and agricultural capabilities; repair capabilities; and habitable and farmable space.
3. *The Settlement Phase.* The activities of this phase center on founding and growing cities on Mars that will represent foundations of new branches of human civilization on the Red Planet. This phase will involve hundreds of thousands to millions of people on Mars. Key resources that need to be developed include an increasingly full spectrum of large-scale manufacturing and agricultural capabilities, energy production, and development of commercial products for export.

We discuss each of these in turn.

19.2 The Exploration Phase

The essential resource creation capability required to support a robust and effective exploration program is the ability to produce propellants. The critical materials necessary for this activity are CO₂ and water, both of which are abundant on Mars.

Carbon dioxide comprises 95% of the Martian atmosphere, and can be obtained anywhere by running a roughing pump. Alternative methods of CO₂ acquisition that could potentially be implemented with lower power requirements than the straightforward roughing pump approach include freezing, solid sorption pumps, and liquid sorption pumps.

Acquisition of CO₂ via freezing was first proposed by Robert Ash and his Jet Propulsion Lab collaborators in their seminal 1976 paper (Ash et al. 1978). At 8 mb pressure, CO₂ can be directly frozen out of the atmosphere via contact with a surface refrigerated to 150 K. The problem with this approach is that as CO₂ freezes on a surface, its frost insulates the surface above it from the refrigerator below. In his original paper, Ash proposed that this could be dealt with by employing a windshield wiper-type arrangement to continually remove dry ice frost from the refrigerated surface, scraping the product into a bin, which could then be sealed and heated to produce CO₂ gas at pressure. This approach has yet to be demonstrated experimentally. A simpler system, in which CO₂ was simply frozen in a refrigerated can, was demonstrated by Frankie at Pioneer Astronautics (Frankie and Zubrin 1999) in 1999, but performance was limited to about 20% of the can's nominal volumetric capacity by the self-insulating nature of CO₂ frost. To address these problems, NASA has pursued CO₂ freezers along with rapid cycle sorption/desorption for full-scale production of O₂ or CH₄/O₂ on Mars. A full-scale Mars CO₂ freezer based on earlier pilot-scale freezers is described in Reference (Meier et al. 2018). That design was based on prior work at NASA KSC (Muscatello et al. 2017), in which a copper cold head with a large surface area was utilized to minimize the insulating effects of dry ice. A short freeze/vaporize cycle was also employed to keep the layer thin along with using warm water from the active cryocooler to assist in vaporizing the dry ice.

Solid sorption beds made of zeolite or activated carbon can absorb 10 to 20%, respectively, of their weight as CO₂. Because zeolite is denser than activated carbon, at 10% take-up it will actually absorb about 50% more CO₂ by volume than graphite will at 20% take-up. In experiments done at Martin Marietta (later Lockheed Martin) by Zubrin, Price, and Clark in 1994–95 (Zubrin et al. 1995) the effectiveness of both zeolite and graphite solid sorption pumps was demonstrated. In these experiments, sorption beds were cooled to –90 °C and exposed to an inlet of simulated Mars atmosphere at 8 mb. When the simulant gas was pure CO₂, the beds rapidly sucked in gas, saturating themselves. This gas was then released from the beds at 1 bar by heating them up to 300 °C, to be used as feed for a Sabatier-electrolysis propellant production unit. However, when Mars simulant gas composed of 95% CO₂ and 5% N₂ was used as feed, the sorption pumps slowed down and eventually stopped taking up CO₂ well short of saturation. It was determined that the operation of the sorption bed was being impaired by the accumulation of unsortable nitrogen in the bed,

blocking access to additional CO₂. This was remedied by installing a small blower into the system to move unsorbed N₂ out of the way. Since then, such enhanced sorption beds have been used on a number of projects (Sanders et al. 2015).

An advantage of the sorption-bed approach is that the day/night cycle itself can be used to drive CO₂ capture. Nighttime temperatures on Mars typically fall to -90°C . This is not cold enough to freeze CO₂, but it can very effectively drive CO₂ take-up by a sorption bed. Thus the only power needed to drive a sorption-bed system is heating the bed, which could be done using either day-time photovoltaic power, concentrated sunlight, or waste heat from a radioisotope or nuclear reactor system. Such an approach, however, would require a massive sorption-bed system, needing 1000 kg of zeolite, for example, plus an accompanying large structural mass, if it were desired to obtain 100 kg of CO₂ per day. To get around this limitation, the Martin Marietta team employed a battery of two small alternating sorption pumps, with one cold one sorbing while the other was heated to produce gas for the system. Such an approach could enable effective sorption-bed systems of modest mass. However, power for active refrigeration would be required to support operation.

In order to avoid these problems, the liquid sorption pump was proposed by Zubrin in 2017, and demonstrated by Zubrin, Fatur, Rose, and Harrison (Zubrin et al. 2019) at Pioneer Astronautics in 2018. In this system, cold ethanol at temperatures from -90° to -115°C is trickled down an absorption column. A blower is used to drive Martian air into the base of and then up the column. The CO₂ is highly soluble in cold ethanol and can go into solution with as much as 10% concentration, while the nitrogen and argon minority components of Mars air, being much less soluble, are vented out of the top of the absorption column. The CO₂-saturated ethanol is then piped to a desorption column, where at 60°C the CO₂ can be forced out of the ethanol at pressure. A condenser is used to stop loss of ethanol with the CO₂ product. Heat exchangers are employed between the absorption and desorption columns to minimize required heating and cooling power. The CO₂ product so obtained is nearly 100% pure. The system can operate using nighttime temperature of -90°C , but pressurizing the input gas to ~ 100 mb is required to obtain effective operation at that temperature. To avoid this, the optimal strategy appears to be to operate the system at night, but employ a one-stage active refrigerator to cool the ethanol to -110° to -115°C (ethanol freezes at -116°C) to maximize effective CO₂ take-up without the need to pressurize the gas above ambient conditions. At 10% take-up a system pumping ethanol at a rate of 1 kg/minute can produce 100 g/min of high-purity pressurized CO₂ or 60 kg in the course of a 10-h Martian night. If raised to 5 bar, the CO₂ product could then be stored as a liquid for use during the day as well. Calculations indicate that such a production rate, yielding 30 tons of CO₂ in the course of a 500-day Mars surface stay, could be supported with about 2 kWe of power. This is about a factor of 5 less than that needed by a roughing pump delivering equivalent product, a significant gain, albeit at the cost of added complexity.

Water can be obtained on Mars from the atmosphere, soil, surface ice, or subsurface brines.

Obtaining water from the Martian atmosphere was investigated by Adam Bruckner et al. at the University of Washington in the early 1990s (Williams et al. 1995). In their

system, known as a WAVAR, powerful fans drove very large volumes of Martian air through a zeolite 3A bed, which has a very high affinity for water. Studies showed that the idea was feasible, but as a result of Mars air containing only about 0.001 g of water per cubic meter, the system needed to be very large, with a 5 square-meter intake duct and a fan driving Mars air at a rate of 100 m/s needed to produce about 45 kg of water per day. Power requirements would also be high; the system described above would need to accelerate about 8 kg/s of Martian air to 100 m/s. At 100% efficiency this would require about 40 kWe for 24 h, or 960 kWh, plus 16 kWh to drive the water out of the zeolite.

In 2002 NASA's Mars Odyssey spacecraft reached Mars and mapped the water content of Martian soil from orbit. Odyssey found continent-sized regions of Mars at high latitudes that are up to 60% weight water in the soil, with soil water concentrations of 5% being common at the equator. These results suggested that baking water out of the soil on Mars would be a much more efficient method of obtaining water on Mars than trying to take it from the atmosphere. Such an approach was demonstrated by Mark Berggren, Heather Rose, and their collaborators at Pioneer Astronautics in 2017 (Berggren et al. 2017). In the Mars Water Acquisition System (MWAS) demonstrated by Berggren, Mars regolith simulant containing 5% water by weight was placed in a sealed vessel through which hot CO₂ was driven by a compressor. Water vaporized in the vessel was carried out with the CO₂ to be collected in a condenser vessel. A heat exchanger was placed between the input and output to the condenser to improve efficiency. By taking the water out of the soil as vapor, rather than melting it out, pure water was obtained without saline contamination, a result which Rose dramatized by using the water product both to grow plants and as feedstock in a solid polymer electrolyzer.

Berggren's MWAS worked well enough, but while power requirements were an order of magnitude lower than those needed for a WAVAR-type system, they were still much higher than the 0.75 kWh/kg needed to vaporize water, let alone the 0.14 kWh/kg necessary to melt -50°C ice. In the 1600 s, Huygens observed white caps covering Mars' poles and claimed that they represented frozen water ice. While endlessly disputed, the evidence supporting this hypothesis increased over the following centuries, particularly after 1971, when NASA's Mariner 9 orbiter imaged networks of water erosion features across the Red Planet. All doubt was finally put to rest when Mars' North Pole was definitively proved to be water ice by NASA's Phoenix probe, which landed there in 2007.

In 2018, the team operating the SHARAD ground-penetrating radar on NASA's Mars Reconnaissance Orbiter (MRO) reported the detection of massive formations of glaciers composed of pure water ice, covered by only a few meters of dust, extending from the North Pole southward to 38°N . This is the same latitude as San Francisco or Athens on Earth. The total amount of water in these glaciers is estimated to exceed that present in America's Great Lakes.

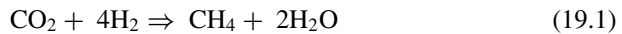
If the landing site is chosen in a location near such a glacier, water can be accessed at very low energy cost, using a proven technology known as Rodriguez wells, or Rodwells. Rodwells, which have been used for years to provide water to scientific bases in Antarctica, work by pumping hot water down into ice, melting some of

it, and returning a larger volume of cold water to the surface. Some of this water can be reheated and sent back down again, repeating the cycle. The pumping power required by Rodwells is insignificant, enabling the production of water at little more than the tangible heat required to raise it to 0 °C, plus the heat of fusion. On Mars this is 0.14 kWh/kg, so using 1 kW of waste heat, over 150 kg of liquid water could be acquired each day. The vapor pressure of water at 0 °C is 6 mb, which is less than average ambient low-altitude pressure on Mars. If the water were acquired from a high-altitude location, where ambient pressure is less than 6 mb, some kind of choke would need to be applied to the well head to keep the melted water in the liquid phase.

Finally, in 2018 the ESA Mars Express orbiter MARSIS ground-penetrating radar team discovered subsurface liquid saline water on Mars. While of great interest to exobiology research, these reservoirs are located more than a kilometer underground, and so represent a less accessible potential source of water for a Mars base than the glaciers discussed above.

19.2.1 The Chemistry of Propellant Manufacture on Mars

The simplest technique to make rocket fuel and oxygen on Mars is to electrolyze water to produce hydrogen and oxygen and then react the hydrogen with the CO₂ that comprises 95% of the Martian air as follows.



Reaction 19.1 is known as the Sabatier reaction and has been widely performed by the chemical industry on Earth in large-scale one-pass units since the 1890s. It is exothermic, occurs rapidly, and goes to completion when catalyzed by ruthenium on alumina pellets at 400 °C. This was proposed by Ash et al in their cited 1976 paper. I first demonstrated a compact system appropriate for Mars application uniting this reaction with a water electrolysis and recycle loop while working at Martin Marietta in Denver in 1993. The methane/oxygen combination produced is an excellent rocket propellant, offering a specific impulse as high as 380 s. This is the second-highest specific impulse available from any plausible rocket propellant combination, being exceeded only by hydrogen/oxygen, which can deliver 450 s. Methane/oxygen, however, is three times as dense as hydrogen/oxygen, and can be stored in compact isothermal tanks, as liquid methane and oxygen can store at the same temperature. Moreover, it takes much less power to liquefy and store liquid methane at 95 K than liquid hydrogen at 20 K. It was for this reason that I chose methane/oxygen as the propellant to use in the Mars Direct mission plan in 1990 (Zubrin and Baker 1999), and Elon Musk chose it for use in his Starship Mars mission architecture in 2016 (Musk 2016). Methane/oxygen was first demonstrated for use in 20,000 lbf Pratt and Whitney RL-10 engines in the 1980s. Since that time,

higher-thrust methane/oxygen engines have been demonstrated, notably including the 500,000 lbf SpaceX Raptor and the 550,000 lbf Blue Origin BE-4.

The water produced by the Sabatier reactor can either be consumed as such or electrolyzed to make oxygen for propellant or consumable purposes, and hydrogen, which is recycled. In the 1990s, when I proposed Mars Direct, the availability of Mars water was unclear. I therefore proposed to bring the required hydrogen for methane/oxygen manufacture from Earth. In this case, 95% of the mass of the resulting propellant would come from Mars, with only 5% needing to be imported from Earth. Such a ratio is termed a “leverage” of 20:1. This might be reduced to 18:1 if it were desired to run the rocket engine with a slightly fuel-rich (O/F) mixture ratio of 3.5:1 rather than the stoichiometric O/F ratio of 4:1. Despite the availability of water ice on Mars, such a strategy to produce propellant at very high leverage might still be considered for initial missions, as it reduces local material acquisition requirements for propellant production to simply gathering CO₂, which can be done anywhere. The required hydrogen can be transported to Mars inside the ascent vehicle lower-stage tanks that will later contain methane. For example, assuming two methane and two hydrogen tanks per stage, we could first empty one of the lower-stage methane tanks containing hydrogen to produce methane and water, with the methane product going into the upper stage, and the water going into a compact bladder. Once that is done, the first lower-stage tank could receive the methane produced from the hydrogen stored in the second lower-stage tank. Then, with the transported hydrogen all gone, the system could proceed to make the rest of the propellant using hydrogen produced on an ongoing basis from the stored water.

A methane/oxygen engine operating with a specific impulse of 380 s, using propellant produced at a leverage of 18:1, may be said to have an “effective specific impulse” of $380 \times 18 = 6840$ s. *This is as high as that offered by electric propulsion, but at high thrust.* Moreover, the thrust is generated by a lightweight chemical engine, without the need to lug around a massive power plant. Effectively what local propellant production allows us to do is to take the energy produced by a large nuclear or solar power system on a planetary surface and integrate it over time, storing the energy in portable chemical form. This is by far the most effective way to use electric power for propulsion in space.

Another system that has been demonstrated for Mars resource utilization is direct dissociation of CO₂ using zirconia electrolysis cells. The reaction is:



Reaction 19.2 is very endothermic and requires the use of a ceramic membrane system with high-temperature seals operating above 1,000 °C. Its feasibility was first demonstrated by Robert Ash at the Jet Propulsion Lab in the late 1970s, and the performance of such systems was later significantly improved by Kumar Ramohali and K.R. Sridhar at the University of Arizona (Rapp 2013). Its great advantage is that no cycling reagents are needed. Its disadvantage is that it requires a lot of power—about twice that of the Sabatier process to produce the same amount of propellant. A small-scale (20 gm oxygen/hour production rate) version of such a system, called

MOXIE, has been placed on the Perseverance rover, and was shown to work on Mars (Hecht and Hoffman 2020).

Because it involves the use of a multitude of small ceramic tubes, each with its own high-temperature seal, questions remain as to the potential of this technology to scale up to meet the production requirements relevant to human Mars expeditions.

Still another method of Mars propellant production is the Reverse Water Gas Shift (RWGS).



This reaction is mildly endothermic and has been known to chemistry since the nineteenth century. Its advantage over the Sabatier reaction is that all the hydrogen reacted goes into the water, from where it can be electrolyzed and used again, allowing a nearly infinite amount of oxygen to be produced from a small recycling hydrogen supply. It occurs rapidly at 400 °C. However, its equilibrium constant is low, which means that it does not ordinarily go to completion, and it is in competition with the Sabatier reaction (19.1), which does. Working at Pioneer Astronautics in 1997, Brian Frankie, Tomiko Kito, and I demonstrated that copper on alumina catalyst was 100% specific for this reaction, however, and that by using a water condenser and air separation membrane in a recycle loop with a RWGS reactor, conversions approaching 100% could be readily achieved (Zubrin et al. 1997).

Our initial RWGS unit produced water at a rate of about 1 kg per day, which would be appropriate to make the oxygen propellant needed for the ascent vehicle of a robotic Mars sample return mission. Building on this work, in 2017, at Pioneer Energy, a commercial spin-off company of Pioneer Astronautics, my R&D team demonstrated an RWGS system operating at a rate of 80 kg of water production per day, sufficient to make all the oxygen propellant needed for a Mars Direct-scale human exploration mission.

19.3 The Base-Building Phase

The activities of this phase center on creating a base that will not only support greatly expanded exploration, but the development of an ever-expanding technological, industrial, and agricultural capability to enable human settlement. This requires expanding the material base of the life-support system to enable supporting hundreds to thousands of people on Mars. To reduce its logistic requirements, the base will need to be able to produce the bulk materials needed such as brick, concrete, ceramics, iron, aluminum, glass, fabrics, and plastics, engage in small-scale manufacturing and agriculture, and be able to create habitable and farmable space.

19.3.1 Local Resource Creation for Life Support

Using either reaction (19.2), or reaction (19.3) together with water electrolysis, unlimited amounts of oxygen may be produced for life-support purposes, with the only local material required being CO₂. Water acquisition systems meeting propellant production requirements can also easily meet life-support system needs, as these are typically one to two orders of magnitude less. Using water and CO₂, plants may be grown, supplying food and fabric. On Mars, buffer gas for breathing systems, consisting of nitrogen and argon, can be extracted directly from the atmosphere using pumps, as these gases comprise 2.7% and 1.6% of the air there, respectively. While comprising less than 5% of Martian air, the disparity between atmospheric acquisition on a scale to meet propellant production requirements and those of life support means that any system designed to acquire enough Mars air to produce propellant will also process more than enough gas to meet crew buffer gas requirements.

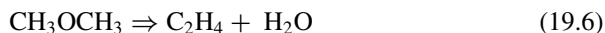
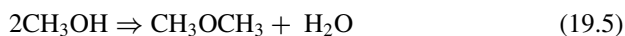
19.3.2 Building Greenhouses

Food can be grown on Mars for experimental purposes in greenhouses transported from Earth. However, if an agricultural base is to be established of sufficient size to support human settlement, the bulk of the materials needed to construct greenhouses need to be producible on Mars. Growing plants requires a great deal of energy in the form of light. For example, a prime hectare of Iowa land produces about 9 metric tons of corn per year. If transformed into an equivalent mass of diverse foods, this could support about 20 people. At high noon, such land receives about 10 MW of sunlight, which works out to 500 kW per person. This is two orders of magnitude higher than the per-person life-support electric power requirement on Mars (and three orders of magnitude higher than average per-person electric power consumption on Earth). If this light were supplied using solar energy, 12 hectares of land would need to be covered with 20%-efficient photovoltaic panels to illuminate each hectare of cropland. While conceivably a highly efficient greenhouse system might improve on these results significantly, it is apparent that Mars settlers will need to take advantage of natural sunlight to support plant growth. Therefore, Mars settlers will need the capability of producing transparent structural materials in bulk. Such materials could include plastics, glass, and ice.

19.3.3 Plastics Production

Running the RWGS with extra hydrogen, a waste gas stream consisting of CO and H₂ can be produced. This is known as “synthesis gas” and can be reacted exothermically in a second catalytic bed to produce methanol (reaction 19.4.) The methanol in turn

can be used to produce dimethyl ether (DME) by running it over a gamma alumina catalyst at 400 °C (reaction 19.5). Such use of RWGS “waste” gas to make methanol was first demonstrated during the 1997 Pioneer Astronautics program, and then on a much larger scale (5 kg methanol/hour) during a 2017 Pioneer Energy project, which also demonstrated the ability to transform the methanol product into DME. Methanol and DME are both important chemical commodities on Earth. Methanol can be used in flex-fuel internal combustion automobile engines and in fuel cells. DME can be used as diesel fuel, and in fact offers superior performance to petroleum-derived diesel fuel, possessing a cetane rating of 65, compared to petroleum diesel’s 48. DME can also be used to produce ethylene or propylene via reactions (19.6) and (19.7). In addition to being excellent rocket fuels, these substances are also the basis of the plastics industry, as they are the raw materials for production of polyethylene and polypropylene, the two most important plastics.



It may be noted that in addition to supplying a source of hard plastic for the manufacture of spare parts by casting or 3-D printing, polypropylene is an attractive material for the manufacture of fabrics for clothing, and is currently widely used for production of superior outdoor attire.

Ethylene and propylene can also be used to make methyl methacrylate (MMA), $\text{CH}_2 = \text{C}(\text{CH}_3)\text{COOCH}_3$. MMA can then be polymerized to create poly(methyl methacrylate) or PMMA, an extremely useful greenhouse construction material, generally known by its trade name, Plexiglas.

An endless variety of other plastics, including strong structural ones such as nylon and PVC, can be made by combining CHO compounds with nitrogen or chlorine, both of which are available on Mars.

19.3.4 Glass Production

The manufacture of clear glass on Earth requires silica sand (SiO_2), sodium oxide (Na_2O) from soda ash, calcium oxide (CaO) from limestone/dolomite, as well as well as feldspar (containing Al_2O_3). All of the elements in these minerals are available on Mars, but not necessarily the minerals themselves. NASA’s Spirit and Curiosity rovers have both detected large silica deposits in Gusev and Gale Craters respectively.

In 2012, scientists using the CRISM instrument on MRO announced that they had detected quartz, feldspar, and amorphous silica deposits near Antoniadi Crater on Mars. NASA's Curiosity rover has found calcium-sulfate veins in rocks, apparently left behind by water action in Yellowknife Bay. Sodium salts have also been found in various evaporate deposits explored by the Mars rovers.

It is thus apparent that the raw materials for glass production exist on Mars, but that novel approaches may need to be developed to minimize the complexity of glass production using the mineral feedstocks available within reasonable transport range of the Mars base. This would be an excellent topic for Earth-based research using data now to hand on Martian mineralogy from the rover missions.

Glass is an extremely important material for modern civilization. It may be noted that while glass has been used on Earth since circa 3500 BCE, techniques for making transparent glass were not discovered until circa 100 CE in Alexandria, and truly clear glass had to wait for its development till 15th-century Venice. Of particular concern if Mars glass is to be made fully transparent is the elimination of iron-oxide dust impurities from the feedstock.

19.3.5 Ice Architecture

The discovery of copious supplies of water ice in mid-latitudes on Mars opens up the possibility for large-scale construction of greenhouses and habitats on Mars using ice as a primary structural material. Employing Rodwells to extract water from glaciers, water can be obtained at an energy cost of 0.14 kWh/kg. A 1 MWe reactor operating at 25% efficiency would produce 4 MW of waste heat. If stationed near a glacier, this would be sufficient to acquire 685 metric tons of water per day. While weak in tension, ice is strong in compression, and structures can be built made of it using blocks or any other conventional technique involving bricks, concrete or other compression-strong/tensile-weak materials. However, as ice can be readily liquefied, it should be possible to 3-D print ice structures in ways that are not possible with bricks. Furthermore, as a primary objective of a Mars structure is to contain pressure, the tensile weakness of ice can be productively remedied by supporting it from below by a pressurized structure, whose pressurization itself is assisted by the weight of the ice above it. It takes about 30 m of ice to weigh down with a pressure of 1 bar on Mars. So a polyethylene dome pressurized to 50 mb—sufficient for a greenhouse—could support and have its pressure contained by a 1.5 m-thick layer of ice. A still more interesting possibility might be to cover a polyethylene dome habitat pressurized to 300 mb—Skylab space station pressure—with 8 m of liquid water with a thin ice layer on top. The inhabitants of such a dome would enjoy ample shielding, protected from cosmic radiation by a transparent aquaculture greenhouse, growing abundant kelp, fish, and other seafood above their heads.

Ice structures could also be built by melting tunnels into glaciers. Alternatively, using waste heat from nuclear reactors, the contents of ice-filled craters could be melted from beneath their icy surfaces, to create large habitable aquatic environments.

For example, a 500 m-diameter crater 20 m deep could be entirely melted out below its surface by the 1 MWe reactor described above in about 16 years, creating a large ice-covered lake which could be farmed for fish and kelp, and also provide a convenient location for well-shielded submarine human habitats.

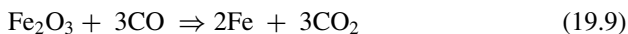
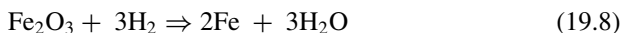
The unique advantages and potential opportunities offered by ice architecture on Mars are profound, and call for further research and aggressive development.

19.3.6 Bricks and Concrete

Mars is rich in clay-like materials suitable for brick manufacture, and possesses gypsum, a key ingredient for the manufacture of Portland cement. So it should be possible to manufacture these traditional building materials on Mars as well.

19.3.7 Metals Production

Iron can also be produced on Mars using either reactions (19.8) or (19.9). Indeed, the solid feedstock, Fe_2O_3 , is so omnipresent on Mars that it gives the planet its red color, and thus indirectly, its name.



Reaction (19.8) is mildly endothermic and can be used with a water electrolysis recycling system to produce oxygen as well. Reaction (19.9) is mildly exothermic and can be used in tandem with an electrolyzer and an RWGS unit to also produce oxygen. The iron can be used as such or turned into steel, as carbon, manganese, phosphorus, silicon, nickel, chromium, and vanadium, the key elements used in producing the principal carbon and stainless steel alloys, are all relatively common on Mars. To show this, in 2017, Pioneer Astronautics demonstrated the use of reaction (19.9) to make carbon steel out of Mars soil simulant samples.

While the above reactions readily produce metallic iron and oxygen from iron oxide, unless nearly pure iron oxide feedstock is obtained, there will be a significant additional task required to separate the metallic iron product from surrounding unreduced oxides of silicon, aluminum, calcium, etc. In the 2017 work at Pioneer Astronautics, this was accomplished using the traditional, but energy-intensive, technique of heating the entire reduced iron/unreduced oxide mixture to 1500 °C, after which the iron melted out and was extracted as a slug.

An alternative approach to separate the iron is to produce iron carbonyl, as long advocated by the University of Arizona's Professor John Lewis (Lewis and Lewis 1987).

For example, carbon monoxide produced by the RWGS can be combined with iron at 110 °C to produce iron carbonyl ($\text{Fe}(\text{CO})_5$), which is a liquid at room temperature. Then iron carbonyl can be poured into a mold, and then heated to about 200 °C, at which time it will decompose. Pure iron, very strong, will be left in the mold, while the carbon monoxide will be released, allowing it to be used again. Similar carbonyls can be formed between carbon monoxide and nickel, chromium, osmium, iridium, ruthenium, rhenium, cobalt, and tungsten. Each of these carbonyls decomposes under slightly different conditions, allowing a mixture of metal carbonyls to be separated into its pure components by successive decomposition, one metal at a time.

An additional advantage of this technique is the opportunities it offers to enable precision low-temperature metal casting. One can take the iron carbonyl, for example, and deposit the iron in layers by decomposing carbonyl vapor, allowing hollow objects of any complex shape desired to be made. For this reason, carbonyl manufacturing and casting will no doubt also find extensive use on Mars, as well as the asteroid worlds that were Lewis's primary target.

Yet another alternative method to produce pure iron on Mars, the Moon, or the asteroids, would be to employ electrochemistry. Electroplating of iron from ore was first demonstrated on Earth in 1859, but has not been employed much since reduction of iron ore using carbon monoxide produced by the incomplete combustion of coal is so much cheaper. In extraterrestrial environments this could change a lot. In 2020, Alex Roman and Diana Aksenova working at Pioneer Astronautics achieved good results electroplating iron from JSC-1 lunar simulant using a potassium-hydroxide solution as his electrolyte. Further work to characterize the feasibility and merit of utilizing this process to produce iron in extraterrestrial environments is ongoing.

If pure metals can be obtained, 3-D printing technologies can be developed and employed, as the feedstock will have predictable qualities. This will offer enormous advantages, as it will allow Mars base personnel to produce anything they can draw, including notably spare parts for machines, an essential capability to reduce the logistics requirements of a base of any size.

19.3.8 Aluminum Production

Aluminum oxide, or alumina Al_2O_3 , is common on Mars, comprising about 4% of the planet's surface, by weight. However, the aluminum is very tightly bound to oxygen, and so directly reducing aluminum by a reaction analogous to (19.9) would require extremely high temperatures. It is for this reason that, while iron has been in common use for 3,000 years, aluminum metal was unknown to science until the nineteenth century, and did not come into general use until the twentieth century. The process involves dissolving the alumina in cryolite at 1000 °C, using carbon electrodes, which are consumed in the process, while the cryolite is unharmed.

The required carbon can be produced on Mars from RWGS carbon monoxide via:



This reaction is moderately exothermic and occurs spontaneously at high pressure and temperatures of about 600 °C. Carbon could also be produced by pyrolyzing the methane product of reaction (19.1), but this would be less efficient, as it requires electrolyzing four water molecules for every carbon produced, instead of the two required by the RWGS route.

In addition to its complexity, the production of aluminum also comes with a high energy cost, about 20 kWh/kg, ten times the 2 kWh/kg required to produce an equal mass of steel. For this reason, until a Mars settlement becomes truly energy rich, iron and steel will be the primary metals produced for use on the Red Planet.

19.3.9 Graphite Production on Mars

In recent years, graphite products have become important structural materials, increasingly replacing both steel and aluminum. For example, top-quality bicycles, once constructed of steel, and then aluminum, are now being made from graphite-derived carbon-carbon. Other graphite products, such as graphene and carbon nanotubes, hold promise for revolutionary applications ranging from superconductors and photovoltaics to ultra-strong fibers. Indeed, if the development of high-quality steel provided the basis for the industrial revolution of the nineteenth century, and the advent of aluminum enabled the jet-age globalized civilization of the 20th, it has been suggested that graphite may provide the material basis of the twenty-first century and beyond. If so, the raw material to make it is available everywhere on Mars as free as air.

Using reactions (19.3) and (19.10), combined with water electrolysis, graphite may be produced on Mars at an energy cost of about 16 kWh/kg, with 2.7 kg of oxygen also being produced for every kilogram of graphite product. This may not seem attractive, but if oxygen is being produced for propellant purposes by reactions (19.2) or (19.3), then using reaction (19.10), 0.375 kg of graphite can be produced for every kilogram of oxygen produced, *at no extra energy cost whatsoever*. Thus the development of technologies for producing good structural materials out of graphite could be of extraordinary value for a developing Martian civilization.

19.4 The Settlement Phase

While exploring and base building on Mars can be supported by government, corporate, or non-profit largesse, a Martian civilization of millions of people will require an economic foundation. For the same reason that no nation on Earth is truly autarchic

(and those that try to be, such as North Korea, are very poor in consequence), it will never be possible, or desirable, for Mars to be totally self-sufficient. That said, both the purchase and the transport of goods from Earth will need to be paid for, with the latter representing a formidable burden even if interplanetary transport costs can be brought down several orders of magnitude. It is therefore imperative that everything heavy—for example all goods too massive in respect to their value to use intercontinental air freight on Earth today—needed on Mars, be made on Mars. Only high-value lightweight goods requiring Earth's vast division of labor for their manufacture should be imported from Earth. To achieve this objective a vast range of technologies will be necessary, and the capabilities of the limited Martian workforce will need to be greatly multiplied in both quantity and diversity through broad-application labor-saving machinery, automation, robotics, biotechnology, and artificial intelligence. Mars will also need its own large-scale energy sources, for which the only viable candidates are nuclear fission and fusion. While all these technology areas are of interest on Earth as well, they will be far more critical for Mars, a fact that will drive the Martians to be leaders in these fields. If they can be successful at that, they will have a great deal of value to sell on Earth, specifically the patent rights to their innovations. Through the sale of such intellectual property, they will be able to pay for their imports from the mother planet.

However, if Mars is to be the springboard for humanity to become an interplanetary civilization, it will also need to be able to economically export material goods, not only or even primarily to Earth, but to worlds beyond Mars. There is much discussion today of the possibility of mining precious metals in the main asteroid belt for sale on Earth. Small mining outposts scattered throughout the belt will have much more limited production capability and division of labor than will be available on Mars. So, just as during the colonial period the relatively well-developed American colonies paid for their advanced imports from Britain by selling their craft goods to Caribbean plantations, who in turn earned hard cash by selling rum to England, so a Mars capable of exporting its products will be able to profit by participating in a triangle trade with Earth and the asteroids.

The full array of local resource creation, manufacturing, and labor-amplifying technologies that will need to be developed to enable Mars settlement is beyond the scope of this paper. However, as energy and export technology are fundamental to the entire project, I will address them briefly here.

19.4.1 Energy for Mars Settlement

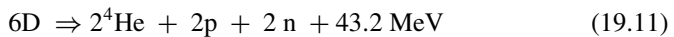
On Earth power can be generated by combustion of biomass or fossil fuels, tapping waterfalls and wind, converting sunlight by photovoltaic or concentrated thermal means, accessing geothermal sources, or nuclear power. On Mars, combustion is not a net energy source, as both the fuel and the oxygen to burn it need to be made. Water and wind power are unavailable, and will remain so until the planet is terraformed. Solar energy can be used on Mars, and has been, but the flux is only 40% as strong as

it is on Earth, and it is subject to months-long cut-off by dust storms, making it even less attractive for large-scale use than it is here. Geothermal power is a possibility, but only in limited amounts in limited locations. So the energy basis for a Martian civilization will need to be nuclear power.

Uranium and thorium have been detected on Mars widely dispersed at about 1 ppm concentration, which is about the same as is typical on Earth. Concentrated uranium ores, typically 2%, but in a few cases up to 20% uranium, have been found on Earth, and there is no reason why they should not also exist on Mars. So in principle, nuclear fission reactors could be fueled from uranium and/or thorium mined on Mars. The problem, however, is that refining high concentrations of these materials out of mined ores is a highly complex and involved industrial process, made much harder if it is necessary to enrich fissile material content from 0.7% U235 to 3% U235 via isotope separation to create reactor-grade fuel.

Deuterium, however, is about five times as common in Martian water as it is on Earth (833 ppm on Mars vs 166 ppm on Earth.) Because the atomic weight of deuterium is twice that of ordinary hydrogen, separating them is much easier than separating U-235 from U-238, which only differ by weight by 1.3%. Moreover, no mining is required, as a Mars colony will constantly be using water, and the physical-chemical life-support system will need to electrolyze at least 1 kg of water per day for every colonist. Thus a Mars colony of 100,000 people will need to electrolyze at least 100 metric tons of water per day, within which there will be 11.1 tons of hydrogen, from which 18.5 kg of deuterium can be extracted. This could be used as fuel in a D-D fusion reactor, whose net reaction, including highly reactive side-products.

T and ^3He which will be consumed in situ, is:



The primary energy cost of hydrogen distillation methods of deuterium production is electrolysis. In this case, the electrolysis of water for life support will require about 480 MWe-hrs, or 20 MWe all day long. The amount of deuterium, however, will be enough to produce 1850 GW-hours, roughly 3,850 times as much energy as it took to extract, or 1,000 times as much power if we assume the fusion energy is converted into electricity at 26% efficiency.

If uranium can be mined and refined Mars, the presence of plentiful deuterium would facilitate the building of reactors of the CANDU type, which use heavy water to moderate their neutron flux, and thus can employ natural uranium, without isotope enrichment, as their fuel. This would at least eliminate the need for the very difficult U-235/U-238 isotope separation from the fuel cycle. Alternatively, natural uranium or thorium could be put in a blanket surrounding a D-D fusion reactor, and making use of the reaction's neutron product, be bred into fissile Pu-239 and U-233, which could then enable uranium and thorium breeder reactors.

But the simplest cycle is simply D-D fusion, made very attractive by the abundant deuterium that will be produced as a byproduct of the operation of the Mars' settlement life-support system. It is for this reason that if fusion is not developed on Earth, it certainly will be on Mars.

It may be noted that deuterium currently sells for about \$4,000/kg, so that the above 18.5 kg of deuterium would be worth about \$74,000 per day, or \$27 million/year if exported to Earth—which is not very much. It could be worth a lot more if exported to the asteroid belt.

Buy cheap, sell dear.

19.4.2 *Enabling Martian Exports*

A Starship-type reusable launch vehicle, running on locally produced methane/oxygen propellant, could operate between the Martian surface and orbit, transporting cargos to an orbital node. From there they could be transferred to an interplanetary craft, propelled either by chemical methane/oxygen engines fueled from Mars, or electric engines fueled with argon from Mars, which could then ship them off to the Earth, the Moon, or the asteroid belt. The ΔV between the Martian surface and low orbit is about 3.8 km/s, from the surface to a 1-sol period highly elliptical near-escape orbit is 4.7 km/s, and from the surface to trans-Earth injection is 6.5 km/s. All of these maneuvers are, in principle, within the capability of the Starship. However, the lower the ΔV , the less propellant it needs and the more cargo it could lift. This suggests a promising application for space elevator, or skyhook, technology.

As originally proposed by Soviet engineer Y. N. Artsutanov in 1960 and revived by Jerome Pearson in the 1970s (Pearson 1975), a skyhook tether could be placed with its center of mass at geostationary orbit, and hung down 36,000 km to Earth, allowing ascents to orbit to be achieved by cable car. While mathematically sound, the engineering of such a system is impractical with current or foreseeable materials, because while the bottom of the cable would only need to lift the payload, the next bit above it would need to lift the payload plus the cable below, and so it would have to be thicker to bear the load. Proceeding upwards in this way, the cable would need to get thicker and thicker, with its mass growing exponentially as it went, causing the system to ultimately require billions of times the payload mass for its construction.

On Mars, however, lower gravity would allow the cable to grow thicker at a much slower rate. Furthermore, if the cable were not in a geostationary orbit, but a lower one, the cable would be shorter and centrifugal force would lower the apparent gravity felt by the cable still more. A cable could be hung 5,800 km down from Phobos, which orbits at a speed of 2.14 km/s 6,000 km above the Martian surface, which itself is 3,400 km from the planet's center. This being the case, the bottom of the tether would be traveling at a speed of just 0.82 km/s above the Martian equator, which itself is turning in the same direction at a speed of 0.24 km/s. As a result, a rocket vehicle taking off from the equator would be able to reach the tether with a ΔV of just 0.58 km/s, catch it, and then be hauled up to Phobos by a cable car. If there were another cable extending outward from Phobos, the vehicle could be *lowered outward* (sic—along this cable, effective gravity would point away from

Mars, since centrifugal force would be greater than gravity) to reach escape velocity at 3,726 km beyond Phobos and still greater speeds at distances beyond. Assuming a tether tensile strength equal to Kevlar (2,800 megapascals), the tether would have a mass less than ten times the payload it could lift. By using such a system, cargo could be sent back to Earth or whipped out to the asteroid belt or beyond with a tenth of the rocket ΔV required to do so directly. This would make it feasible for Mars colonists to transport goods cheaply to Earth, the Moon, the asteroid belt, and the moons of Jupiter, Saturn, Uranus, and Neptune. High-technology goods needed to support asteroid mining may have to come from Earth for some time. But since food, clothing, deuterium, 3-D printed steel and plastic parts, and other necessities can be produced on Mars with much greater ease than would be possible anywhere further out, Mars could become the central base and port of call for exploration and commerce heading out to the asteroid belt, the outer solar system, and beyond.

19.5 Conclusion

An environment becomes habitable once people develop the technology to allow them to create the resources necessary to support human habitation there out of local materials. Thus the NASA acronym ISRU, which stands for In-Situ Resource Utilization, is based on a fundamental misunderstanding of the nature of reality and the human condition. As humans have expanded our physical and technological reach, we have not “utilized” and thereby diminished a stockpile of “natural resources” drawn from some fixed inventory existing before us. Rather we have created the resources available to support us on Earth, and continue to do so, with radically increasing effectiveness, today. There are no natural resources on Mars—or anywhere else in the universe—today. But there will be once there are resourceful people there.

References

- Ash, R.L., W.L. Dowler, and G. Varsi. 1978. Feasibility of rocket propellant production on Mars. *Acta Astronautica*, 5: 705–724. (July–Aug 1978).
- Berggren, Mark, Robert Zubrin, Heather Rose, Anterra Kennedy, and Maxim Shub. 2017. Advanced mars water acquisition system. Phase I Final Report, NASA Contract NASA SBIR Phase I Contract NNX17CC70P, delivered to NASA GRC December 8, 2017.
- Frankie, B., and R. Zubrin. 1999. Chemical engineering in extraterrestrial environments. *Chemical Engineering Progress* 94 (2): 45–54.
- Hecht, Michael H., and Jeffrey A. Hoffman. 2020. The Mars oxygen ISRU experiment (MOXIE) on the Mars 2020 Rover. In *3rd international workshop on instrumentation for planetary missions*. <https://www.hou.usra.edu/meetings/ipm2016/pdf/4130.pdf>. Accessed 29 Dec 2020
- Lewis, J., and R. Lewis. 1987. *Space resources: breaking the bonds of earth*. NY: Columbia University Press.

- Meier, A.J., M. Grashik, M. Shah, J. Sass, J. Bayliss, P. Hintze, and R. Carro. 2018. Full-Scale CO2 freezer project developments for mars atmospheric acquisition. In *2018 AIAA SPACE and astronautics forum and exposition*, 5172.
- Muscattello, A.C., P.E. Hintze, A.J. Meier, J. Bayliss, R. Formoso, M. Shah, B. Vu, R. Lee, and J. Captain. 2017. Testing and modeling of the mars atmospheric processing module. In *AIAA SPACE and astronautics forum and exposition*, 5149.
- Musk, E. 2016. Presentation to International Astronautical Congress, Guadalajara, Mexico, September, 2016
- Pearson, J. 1975. The orbital tower: a spacecraft launcher using the earth's rotational energy. *Acta Astronautica*, 2, Sept–Oct 1975, p.785.
- Rapp, D. 2013. *Use of extraterrestrial resources for human space missions to moon or Mars, Chapter 2*. Chichester, Heidelberg: Springer.
- Sanders, Gerald B., Aaron Paz, Lara Oryshchyn, Koorosh Araghi, Anthony Muscatello, Diane L. Linne, Julie E. Kleinhenz, and Todd Peters. 2015. Mars ISRU for production of mission critical consumables-options, recent studies, and current state of the art AIAA 2015-4458. Published Online:28 Aug 2015. <https://doi.org/10.2514/6.2015-4458>. Accessed 29 Dec 2020.
- Williams, J., S. Coons, and A. Bruckner. 1995 Design of a water vapor adsorption reactor for Martian in situ resource utilization. *Journal of the British Interplanetary Society*, August 1995b.
- Zubrin, R., and D. Baker. 1999. Mars direct, humans to the red planet by 1999. IAF-90-672. In *41st congress of the international astronautical federation*, Dresden, Germany, October 1990. (Republished in *Acta Astronautica*, 26, no. 12 (1992): pp.899–912. See also R. Zubrin, *The Case for Mars: The Plan to Settle the Red Planet and Why We Must*, Simon and Schuster, NY, 1996, 2011).
- Zubrin, R., S. Price, L. Mason, and L. Clark. 1995. An end to end demonstration of mars in-situ propellant production. AIAA-95-2798. In *31st AIAA/ASME joint propulsion conference*, San Diego, CA, July 10–12.
- Zubrin, R., B. Frankie, and T. Kito. 1997. Mars in-situ resource utilization based on the reverse water gas shift, AIAA-97-2767. In *33rd AIAA/ASME joint propulsion conference*, Seattle, WA, July 6–9, 1997.
- Zubrin, Robert, Steven Fatur, Heather Rose, and Maxim Shub. 2019. Liquid sorption pump. Final Report on NASA SBIR Phase I Contract #: 80NSSC18P1936, Delivered to NASA GRC January 25, 2019.

Chapter 20

Planetary Exploration of Mars



**Robert C. Anderson, James M. Dohm, Debra Buczkowski,
Danielle Y. Wyrick, and Timothy J. Parker**

Abstract During the last six decades of planetary exploration, rapid developments in technology have helped enable the exploration of Mars. To demonstrate the technical feasibility of off-Earth mining, new technologies are being developed to identify, remotely extract, and collect commodities under different environmental conditions. Impressive machines, tools, and operational procedures have been developed that allow scientists to study in detail the atmosphere, surface, and subsurface of Mars, progressing from flybys to orbiters to stationary landers to rovers. Key technologies have been developed to provide data for understanding and characterizing the geology, mineralogy, internal structure, atmospheric composition, and dynamics, as well as to identify the possibility of life on Mars. This makes Mars a major target for human colonization, exploration, and mining.

R. C. Anderson (✉) · T. J. Parker
Jet Propulsion Laboratory/California Institute of Technology, 4800 Oak, Grove Dr. MS 183-600,
Pasadena, CA 91381, USA
e-mail: robert.c.anderson@jpl.caltech.edu

T. J. Parker
e-mail: timothy.j.parker@jpl.nasa.gov

J. M. Dohm
Exploration Institute, LLC, 1603 Capitol Ave. #511B, Cheyenne, WY 82001, USA
e-mail: james.m.dohm@explorationinstitute.com

D. Buczkowski
Planetary Exploration Group, Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins
Rd., Mailstop 200-W230, Laurel, MD 20723-6099, USA
e-mail: debra.buczkowski@jhuapl.edu

D. Y. Wyrick
Space Science and Engineering Division, Southwest Research Institute, Culebra Road, San
Antonio, TX 622078238, USA
e-mail: Danielle.wyrick@swri.org

20.1 Introduction

Off-Earth mining is becoming one of today's hottest planetary topics. Within the last decade, numerous start-up corporations have announced plans to extract resources from a variety of different planetary objects, including asteroids, various moons (including our own), and the planet Mars. To demonstrate the technical feasibility of off-Earth mining, these companies are developing new technologies needed to identify and remotely extract and collect commodities under different environmental conditions. For example, worldwide mining companies such as CSIRO have been leading the way in developing fully autonomous vehicles using unmanned, semi- and fully autonomous mineral extraction equipment in their larger-scale mining operations on Earth for eventual use in space. Other companies such as Caterpillar, Komatsu, and Volvo have been developing and deploying autonomous vehicle technologies for use in deep mines for several years; robotic mining vehicles include excavators, loaders, and trucks. Developing these new technologies is the first step toward extraterrestrial mining (Keck Institute for Space Studies 2020).

During the last six decades of planetary exploration, there has been a rapid development in technology to help enable the exploration of distant worlds. Impressive machines, tools, and operational procedures have been developed that allow scientists to study in detail the atmosphere, surface, and subsurface of many planetary bodies, progressing from flybys to orbiters to stationary landers to rovers. Key science aims have been to understand and characterize the geology, mineralogy, internal structure, atmospheric composition, and dynamics of these worlds, as well as the possibility of life.

Getting to Mars has not been easy. There have been numerous devastating failures of both orbiters and landers, followed by incredible successes. Table 20.1 lists most of the successful missions to Mars and the valuable data they returned.

Over time, these missions have broadened our perspectives on Mars, including insights into its internal, landscape/subaerial, mineralogy, and atmospheric structures, and finally its interpreted evolutionary history. Understanding and exploring the many differences and similarities between Mars and Earth will help us to assess the mining potential of the Red Planet.

20.2 Remote Sensing of Mars

Orbiting, landing and roving spacecraft (Table 20.1) have revealed present-day Mars to be cold, dusty, and desert-like, with a thin atmosphere. However, Mars' more dynamic past is revealed by towering volcanoes, mountain ranges, vast canyon systems, igneous plateaus, tectonic basins and faults, and ancient rocks with distinct magnetic signatures. A hydrosphere and a cryosphere have figured prominently in the evolutionary history of Mars, as is evident from present-day glacial and periglacial landscapes, as well as a landscape that has been etched by ancient lakes, giant floods,

Table 20.1 Mars missions, 1964–2018

Mission	Type	Launch date	Mars arrival	Primary instruments
Mariner 4	Flyby	Nov. 28, 1964	July 15, 1965	Camera
Mariner 6	Flyby	Feb. 24, 1969	July 31, 1969	Camera
Mariner 7	Flyby	March 27, 1969	Aug. 5, 1969	Camera
Mariner 9	Orbiter	May 30, 1971	Nov. 13, 1971	Camera
Viking 1 Viking 2	Orbiter and Lander	Aug. 20, 1975 Sept. 9, 1975	July 20, 1976 Sept. 3, 1976	Camera, X-ray fluorescence spectrometer, biological laboratory, weather package
Mars Global Surveyor	Orbiter	Nov. 7, 1996	Sept. 12, 1997	Camera, spectrometer, laser altimeter
Mars Pathfinder	Lander	Dec. 4, 1996	July 4, 1997	Rover camera, lander camera, alpha proton X-ray spectrometer
Mars Odyssey	Orbiter	March 7, 2001	Oct. 24, 2001	Camera, gamma-ray spectrometer
Mars Express	Orbiter	June 2, 2003	Dec. 25, 2003	Still in operations
Mars Exploration Rover: Spirit Mars Exploration Rover: Opportunity	Lander	June 10, 2003 July 7, 2003	Jan. 4, 2004 Jan. 25, 2004	Camera, microscopic imager, miniature thermal emission spectrometer, Mossbauer spectrometer, alpha particle X-Ray spectrometer, rock abrasion tool, magnet array
Mars Reconnaissance Orbiter	Orbiter	Aug. 12, 2005	March 12, 2006	Cameras, spectrometer, radar
Phoenix Lander	Lander	August 4, 2007	May 25, 2008	Cameras, wet chemistry, atomic force microscopes, weather package

(continued)

Table 20.1 (continued)

Mission	Type	Launch date	Mars arrival	Primary instruments
Mars Science Laboratory	Lander	Nov. 26, 2011	Aug. 5, 2014	Cameras, microscope, spectrometers, radiation detectors, weather package
Mars Atmosphere and Volatile Evolution	Orbiter	Nov. 18, 2013	Sept. 21, 2014	Spectrometers, magnetometer
Insight Lander	Landers	May 5, 2018	Still Operating	Cameras, seismic, heat probe, weather package

river valleys, and maybe even oceans. This geomorphic evidence for the persistent presence of water is consistent with the presence of secondary minerals known to form through the hydrothermal alteration of primary volcanic minerals; these secondary minerals may be of interest to future mining endeavors. The complex igneous/tectonic evolution of Mars contributes to the transport and concentration of minerals (e.g., veins).

Methane has also been detected in the Martian atmosphere, albeit controversially, by the Planetary Fourier Spectrometer instrument (PFS) onboard the Mars Express orbiter (MEX) (Formisano et al. 2004), the thermal emission spectrometer (TES) onboard the Mars Global Surveyor (MGS), and Earth-based observations, as well as by the NASA Mars Science Laboratory (MSL). Magmatism, hydrothermal activity, and biological release are a few of the possible explanations for the methane (Formisano et al. 2004; Prieto-Ballesteros et al. 2006; Onstott et al. 2006), and subsurface traps (structural, sedimentary, ground ice, gas hydrates) may have played a significant role as methane reservoirs. These findings may indicate a significant energy source for future human landed missions. This section reviews the state of knowledge about Mars, especially concerning its geology and the potential for off-Earth mining.

20.2.1 *Internal Structure and Atmosphere*

Although Mars is smaller than the Earth (Fig. 20.1), its interior structure is interpreted to be similarly comprised of a core, mantle, and outer crust. Based on limited subsurface data, the Martian core has been estimated to be ~1800 km in radius and composed of iron, nickel, and sulfur surrounded by a silicate mantle (Rivoldini et al. 2011). Calculated as having twice the amount of iron as within the Earth's mantle (Halliday et al. 2001), the Martian mantle is also richer in phosphorus and potassium

(Brückner et al. 2001). The silicate crust is on average 50 km thick and composed of elements consistent with igneous rocks (Nimmo and Tanaka 2005). Observations suggest that parts of the crust have been magnetized, despite the lack of a present-day, core-driven dynamic magnetic field. The Mars Global Surveyor (MGS) mission identified localized crustal magnetic fields associated with the ancient crustal rocks (Acuña et al. 1998, 1999, 2001; Connerney et al. 1999a, b, 2001, 2005; Langlais et al. 2004; Arkani-Hamed 2004; Whaler and Purucker 2005; Lillis et al. 2008). Understanding dynamo timing is a big part of determining Mars' evolution, as it directs the planet's thermal history. Mars' magnetic history should be recorded within the magnetized rocks on its surface.

The oldest stratigraphic units identified on Mars display magnetic signatures that can be differentiated from each other, thereby providing evidence on the relative age of emplacement for each stratum. Estimates of the age of the Martian dynamo range from as early as 4.5 billion years ago to possibly as late as 3.7 billion years ago (Mittelholz et al. 2020). Following the shutdown of the dynamo, internal sources of Martian heat release were relocated to the margins of giant impact basins and large volcanic provinces (Fig. 20.2). Hydrothermal transport in magmatic systems on Earth provide regions for mineral concentrations.

Ancient Mars had a thicker atmosphere than today (Jakosky et al. 2017), shielded by the magnetic field associated with the active dynamo, leading to warmer and wetter conditions. When the dynamo shut down, Mars was left unshielded from the solar wind; this led to the stripping and thinning of the atmosphere, resulting in an environmental shift to more present-day conditions (Lundin et al. 2004). Following

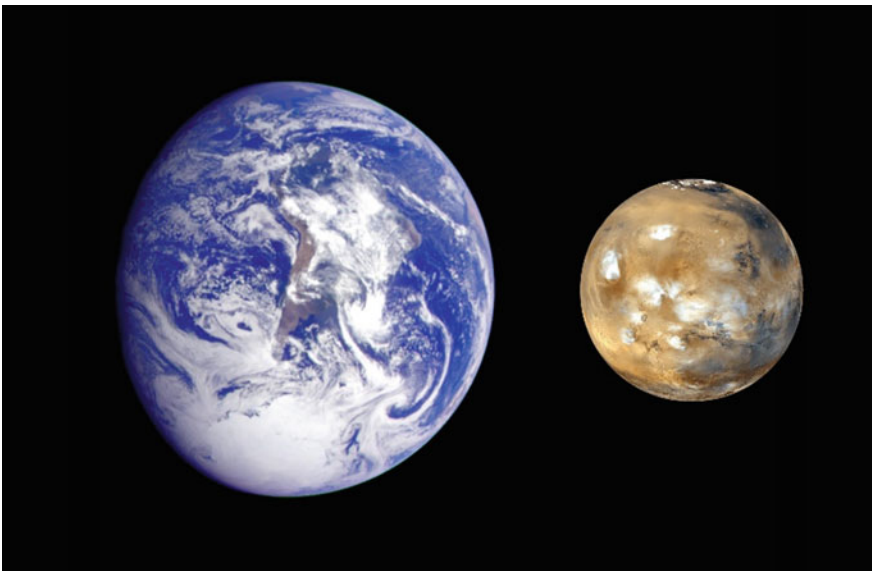


Fig. 20.1 The relative sizes of the Earth and Mars. *Image Credit: NASA*

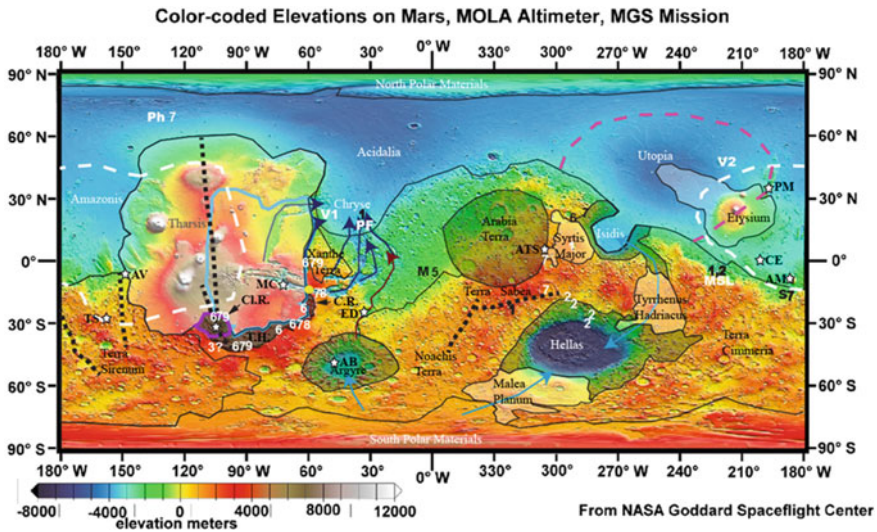


Fig. 20.2 From Dohm (2017). MOLA topographic map showing major features and provinces. Vikings 1 and 2 (V1 and V2), Pathfinder (PF), Spirit (S), Opportunity (M), Phoenix (PH), and Mars Science Laboratory (MSL) landing sites. Circum-Chryse outflow channels (dark blue arrows), Uzboi drainage (dark gray arrow), major glacial routes (light blue arrows), hypothesized Tharsis paleo-basin (solid blue line), inferred plate boundaries (dashed black lines), Tharsis/Elysium corridor (white dashed line). Thaumasia highlands (TH; dark patch), Claritas Rise (violet circle), and Coprates Rise (CR; dark patch). Hellas, Isidis, and Argyre impact basins and hypothesized ancient (>4.0 Ga) impact basin of Arabia Terra (medium-dark patch), hypothesized Utopia impact basin (violet dashed line). Tharsis, Elysium, Tyrrhenus/Hadriacus, Malea Planum, Syrtis Major, and Apollinaris volcanic provinces (light patches). The extent of the Noachian Ocean is approximately marked by the transitional color region of green and yellow. Numbers indicate: (1) possible felsic materials, (2) possible anorthosite, (3) possible ophiolite, (5) hematite, (6) serpentine, (7) carbonate, (8) zeolite, and (9) chlorite. Prime targets for future exploration are shown by black outlined white stars: Abus Vallis (AV), Apollinaris Mons (AM), Argyre basin (AB), Arabia Terra-Syrtis boundary (ATS), Central Elysium (CE), Claritas Rise (Cl.R.), Eberswalde Delta (ED), Melas Chasma (MC), Phlegra Montes (PM), and Terra Sirenum (TS)

the dynamo shutdown, major volcanic outgassing events resulted in transient periods of thicker atmosphere, lasting for thousands of years (Baker et al. 1991, 2007). This outgassing was associated with major surface water flooding events, recorded as outflow channel development and perhaps as many as two ocean phases (Parker et al. 1987, 1993; Dohm et al. 2001a, b, 2007; Fairén et al. 2003).

Present-day Mars has a very thin atmosphere comprised primarily of carbon dioxide, with less than 2% nitrogen and argon, and only trace amounts of oxygen and water (Mahaffy et al. 2013). Methane has also been detected in the current Martian atmosphere; while we do not currently understand the source of methane, it is apparently emanating from discrete regions of the planet's surface and could be a valuable resource for future utilization (Formisano et al. 2004; Komatsu et al. 2011; Dohm et al. 2011).

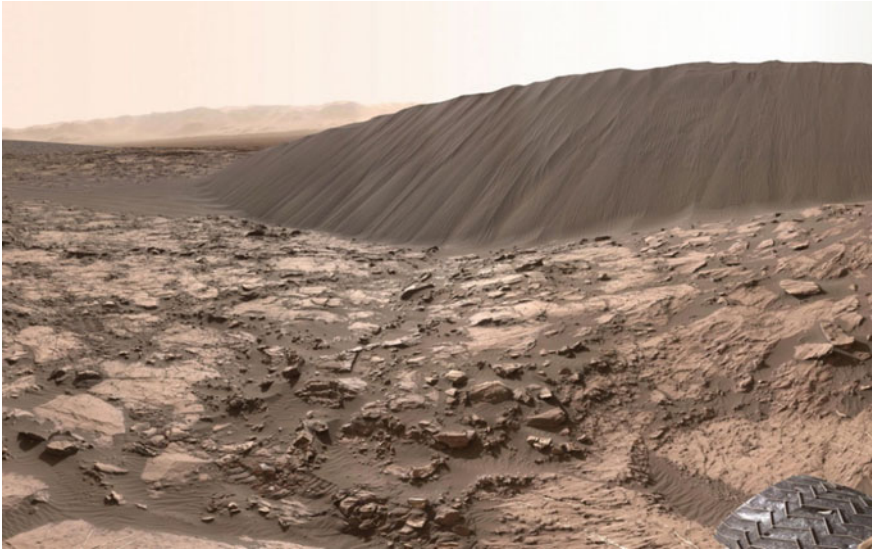


Fig. 20.3 Close-up view of a sand dune called Namib Dune, part of the Bagnold Dunes near Mount Sharp in Gale Crater, as seen by the Curiosity rover on December 18, 2015. Namib is about 16 feet (5 m) tall. *Image Credit: NASA/JPL-Caltech/MSSS*

Despite its low atmospheric pressure (Haberle 2015), the atmosphere of Mars contains a high dust content capable of suspending particles up to 1.5 microns in diameter (Lemmon et al. 2005). Dust storms have been observed on Mars, ranging from small, localized storms to global in extent (Gierasch and Goody 1973; Gierasch 1974). Dune fields (Bourke et al. 2010; Hayward et al. 2007) and dust devils (Metzger and Lancaster 1996; Farrell et al. 2004) have been identified, indicating present-day dynamic atmospheric conditions (Fig. 20.3). These aeolian sediments comprise the uppermost portion of the near-surface strata and may provide valuable resources for future construction.

20.2.2 Topography

The Mars Orbiter Laser Altimeter (MOLA) has returned some of the highest global vertical and spatial resolution (460 m to 115 m/pixel near the poles) digital elevation models for any of the terrestrial planets thus far (Smith et al. 1999, 2001; Zuber et al. 2000; Okubo et al. 2004). MOLA has highlighted the hemispheric dichotomy of Mars in detail (Fig. 20.2), a distinct boundary of which separates the heavily cratered southern highlands from the smoother, less cratered northern lowlands. This crustal dichotomy is a fundamental feature of Mars (Watters et al. 2007), resulting

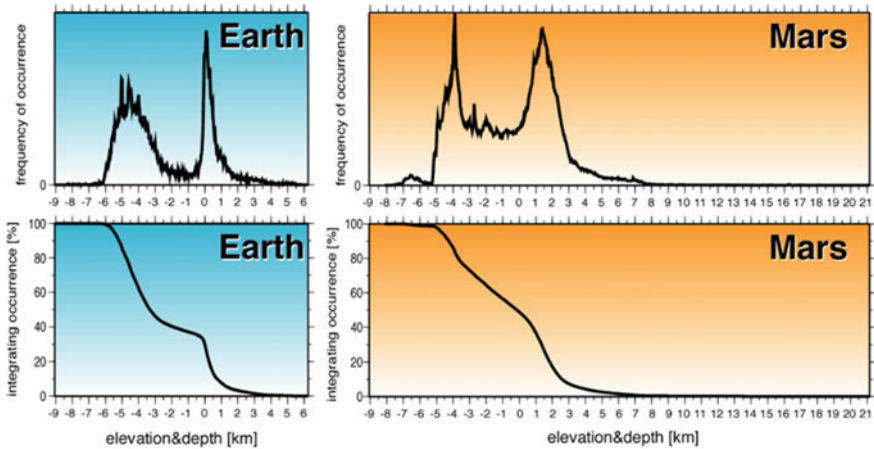


Fig. 20.4 Hypsographic curves and histograms for Earth and Mars, showing that both planets have a bimodal topographic signature. For Earth, 1/3 of the crust by area underlies highlands that are continental in nature, and 2/3 underlies lowlands that are oceanic in nature. For Mars 2/3 of the crust by area underlies highlands, and 1/3 underlies lowlands. *Image Credit: Kido Motoyuki, geophysicist from Tohoku University*

in a bimodal topographic signature, similar to Earth's (Fig. 20.4), but distinct from the other rocky bodies of the solar system, including Mercury, Venus and the Moon.

The bimodal topographic signature of Earth is due to differences between the older and lower continental crust, composed largely of felsic rock materials, and the denser, younger mafic oceanic crust; this crustal differentiation is the result of plate tectonics. Meanwhile, the signature of Mars reflects the crustal dichotomy, the origin of which may have set the course for most of the subsequent geologic evolution of Mars (Fairén and Dohm 2004).

Local and regional variations in topography are pronounced in the southern highlands, which includes the giant Hellas, Argyre, Isidis, and Chryse impact basins. Volcanic provinces, which comprise numerous volcanoes and igneous plateaus, are also found in the southern highlands. In addition, structurally controlled basins and fault systems are often associated with the development of the larger volcanic provinces. The high-resolution MOLA data have been used to study Martian tectonic structures in unprecedented detail (e.g., Golombek et al. 2001; Schultz and Lin 2001; Schultz and Watters 2001; Wilkins and Schultz 2003; Okubo and Schultz 2003, 2004, 2006; Schultz et al. 2006), and have revealed a population of subdued tectonic features in the northern lowlands of Mars partly buried by subsequent emplacement of sedimentary materials (Withers and Neumann 2001; Tanaka et al. 2003). Tectonic systems on Earth provide structurally controlled traps for fluids and gas and should be similar on Mars.

20.2.3 Volcanism

The Tharsis Rise (Fig. 20.2), located in the western hemisphere, is the largest volcanic province on Mars (Scott and Tanaka 1986; Tanaka et al. 2014). The province includes volcanoes, igneous plateaus, fault systems, pit crater chains, and vast canyon systems (Anderson et al. 2001; Dohm et al. 2001a,b, 2007). Some of the prominent volcanoes are Arsia Mons, Pavonis Mons and Ascraeus Mons, collectively known as the Tharsis Montes. The tallest volcano on Mars, Olympus Mons (Fig. 20.5) is also associated with the rise. Syria Mons, a shield complex, has also contributed significantly to the development of Tharsis (Anderson et al. 2001; Dohm et al. 2001a, b). In the east, the Elysium Rise, a smaller volcanic province dominated by Elysium Mons, is found associated with volcanoes (e.g., Hecates Tholus and Albor Tholus) and several flat-lying volcanic paterae. Relatively young, low-relief shield volcanoes occur in Elysium Planitia, to the south and southwest of the rise (Vaucher et al. 2009a, b), while Syrtis Major, another low-relief shield volcano, is found to the west (Greely and Guest 1987). Other volcanic provinces include those along margins of giant impact basins. For example, Syrtis Major is associated with the Isidis impact structure, whereas Malea Planum, Hadriacus Mons, and Tyrrhena Mons are associated with the Hellas impact structure (Crown et al. 1992; Crown and Greeley 2007). One seemingly isolated volcano, Apollinaris Mons (Fig. 20.6), had several stages of development, including possibly interacting with a northern plains ocean during its early formation (Scott et al. 1993; El Maarry et al. 2012).

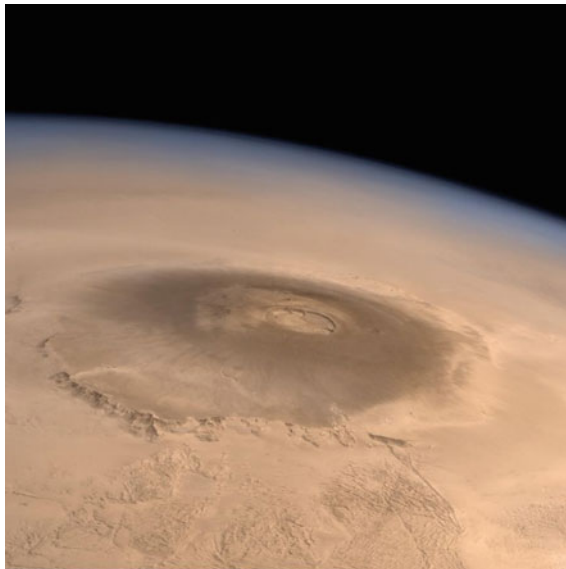


Fig. 20.5 The tallest identified volcano in the solar system, Olympus Mons. *Image Credit: ESA*

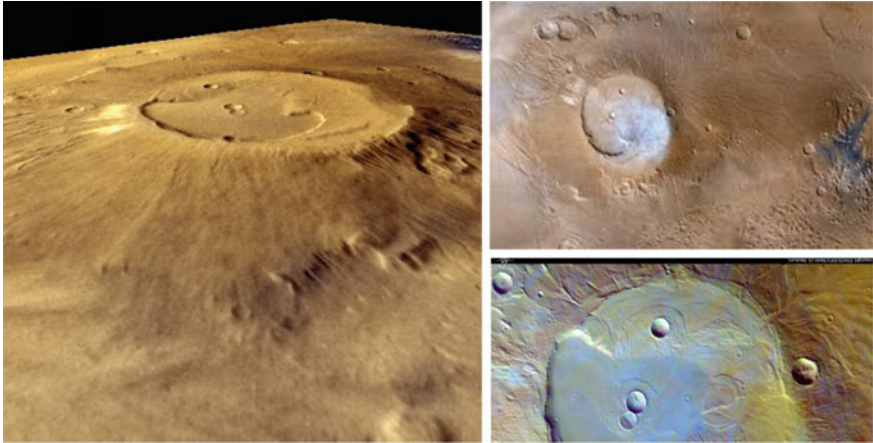


Fig. 20.6 Three images, taken by the High-Resolution Stereo Camera (HRSC) on board ESA's Mars Express spacecraft, show the caldera of Apollinaris Mons, an ancient, 5 km-high volcano north of Gusev Crater. *Image Credit: ESA*

Martian shield volcanoes are similar in shape and slope to terrestrial shield volcanoes that form over mid-plate hot spots, but are significantly larger. This is likely due to the lack of plate tectonics on Mars, where volcanoes sit over a hot spot for extended periods of time, allowing them to keep growing long after a terrestrial volcano would have shifted away due to plate tectonic motion. Basaltic volcanism provides important elements relevant to terrestrial mining operations such as silica, iron, magnesium, sulfur, calcium, and sodium, as part of a typical solid-state solution mineral series. These volcanoes and volcanic plateaus are excellent areas to search for these elements.

20.2.4 Tectonism

Tectonic deformation on Mars reflects its volcanic history. As previously mentioned, there are two major centers of magmatic/volcanic activity on Mars: Tharsis and Elysium. Radial graben are found primarily associated with Tharsis Rise, a region dominated by plume-driven magmatic-tectonic activity (uplift) (Mege and Masson 1996; Baker et al. 2007; Dohm et al. 2007). Radial graben systems like Tharsis result from extensional faulting forming massive fault systems that extend radially out from the center of the dome (Anderson et al. 2001, 2004). Collapse features such as pit crater chains (Wyrick et al. 2004; Ferrill et al. 2004) and chaotic terrain (Rodriguez et al. 2005) have also been recognized.

Concentric graben, on the other hand, have been found to have formed as a result of flexure due to lithospheric loading from a volcano and are found associated to Elysium Mons (Hall et al. 1986). Ring graben have also been found outside the

rim of several of the larger impact basins, such as the Isidis basin (e.g., Wichman and Schultz 1989) and the Argyre basin (Hiesinger and Head 2002; Dohm et al. 2015). For the impact basins, these circumferential graben form as impact-induced ring-fracturing during collapse of the transient basin (McKinnon 1981) or sometimes result from flexure due to subsequent basin fill loads (Wichman and Schultz 1989).

Rift valleys, large elongated depressions with steep walls formed by the downward displacement of surface material between nearly parallel faults or fault systems, have also been identified on Mars. Valles Marineris (Fig. 20.7) has been identified as a rift valley (Anderson and Grimm 1998; Hauber et al. 2010; Andrews-Hanna 2012a, b, c), with its formation related to the development of Tharsis (Dohm et al. 2009b; Hauber et al. 2010) combined with subsidence of the rift floor (Andrews-Hanna 2012a; b, c). This complex region displays a strong interaction between tectono-magmatic processes and water (e.g., Weitz et al. 2003; Komatsu et al. 2004; Quantin et al. 2005) that could have resulted in a diversity of environments conducive to the origin of large mineral deposits. Other trough systems radial to Tharsis, such as Claritas Fossae (Fig. 20.8) or the graben complex in Thaumasia, are also thought to potentially be rift valleys (Hauber et al. 2010; Andrews-Hanna 2012a, b, c). Both areas display features which point to regional extension. There are other areas on Mars that also exhibit major extension such as Basin and Range-like terrain identified in the Terra Sirenum region (Karasozen et al. 2016; Anderson et al. 2019).

Contractional features on Mars are caused by compressional stresses and include the Thaumasia highlands, which record distinct thrust faults as well as other types of tectonic structures (Dohm et al. 2001b, 2013). Lobate scarps appear as linear to arcuate landforms that are generally asymmetric in cross-section, with a steeply sloping face and a gently sloping back limb (e.g., Watters et al. 1998; Watters and Robinson 1999) and are generally recognized to be the expression of surface-breaking

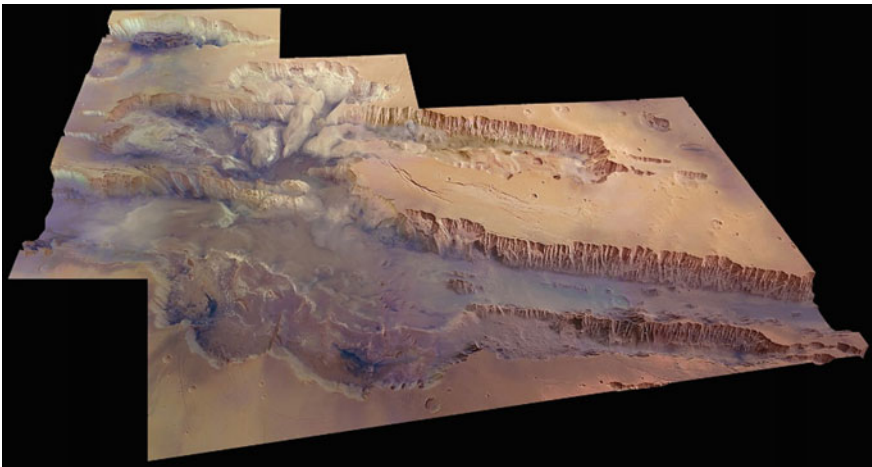


Fig. 20.7 Color image of a portion of Valles Marineris. Called the Grand Canyon of Mars, Valles Marineris is over 4000 km long, 200 km wide and 10 km deep overall. *Image Credit ESA*

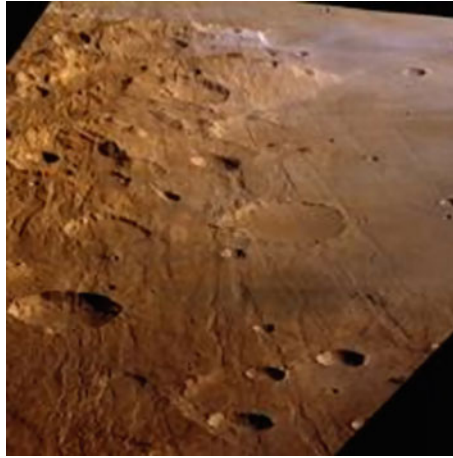


Fig. 20.8 Claritas Fossae is a densely dissected highland terrain on the Tharsis Rise of Mars, located immediately south of the Tharsis Montes. *Image Credit: ESA*

thrust faults (e.g., Melosh and McKinnon 1988). Wrinkle ridges have a distinctive morphology consisting of a broad, low relief arch and a narrow, superimposed ridge (e.g., Strom 1972; Bryan 1973). Mechanical modeling supported by topographic analysis of wrinkle ridges on Mars shows that both the broad arch and the narrow ridge may be formed over a single thrust fault with listric geometry (Watters 2004). Structurally complex tectonic zones provide the means of transport, charge, and trapping of mineralized magmas and groundwater in the Martian crust. Mineralization associated with similar fault, vein, and shear zone systems on Earth provide excellent prospects for in-situ mining.

20.2.5 Paleohydrology

Mars at one time had significant quantities of surface water. Geological and geomorphic investigations on Mars have long pointed to hydrological cycling, including the formation of oceans, lakes, rivers and streams, ice sheets, and glaciers, among other water-influenced landscape architectures (Baker, 1978, 1982, 2001; Baker and Partridge 1986; Parker et al. 1987, 1989, 1993, 2001; Parker and Gorsline 1991; Scott et al. 1991a, b, 1995; Kargel and Strom 1991, 1992; Baker et al. 1991; Cabrol et al. 1998, 2001; Cabrol and Grin 1999, 2001; Clifford and Parker 2001; Grant and Parker 2002; Kargel 2004; Komatsu et al. 2004; Irwin et al. 2004; Howard et al. 2005). These include dendritic valley networks, which are widely distributed on Mars and are similar in scale to those found on Earth. In addition, there are outflow channels that are significantly larger and longer than the valleys. While the valley networks

are thought to form due to the long-lasting flow of water, sourced either by groundwater (Malin and Carr 1999) or precipitation (Craddock and Howard 2002), outflow channels more likely form during large magmatic-driven episodic floods of water (Baker 1978, 1982, 2001; Baker et al. 1991; Dohm et al. 2001a; Fairén et al. 2003). The Martian valleys frequently end at fan-shaped deltas on crater floors, suggesting that the craters once hosted lakes (e.g., Newsom et al. 1996; Cabrol and Grin 1999, 2001). Channels and valley networks that are temporally and spatially associated with tectonic faults and volcanoes, such as observed in the Thaumasia highlands, could indicate groundwater, seismic, and hydrothermal contributions to their formation (Gulick 1993, 1998; Tanaka et al. 1998; Dohm and Tanaka 1999; Dohm et al. 2001b). Similarly, chaotic terrains occur associated with outflow channels, such as Juventae, Capri, and Ganges Chasmata, northeast of Valles Marineris (Fig. 20.2), marking catastrophic releases of floodwaters and other volatiles associated with collapse (Dohm et al. 2001a).

Whether Mars once had an ocean-sized body of water in the northern lowland plains continues to be a question of considerable interest, especially considering the importance of identifying water and/or evaporite deposits for the future exploration of Mars. This question includes uncertainties about the formation, evolution, timing, and ultimate fate of the putative ocean. Significant evidence that an ocean once occupied the northern plains are the identification of numerous deltas at the same elevation along the dichotomy boundary (Achille and Hynes 2010) (Fig. 20.9) and elemental information (Dohm et al. 2009c). Models have been completed to address the plausibility of a high-volume northern ocean on Mars and determine whether it could have existed for an extended period without freezing (Head et al. 1999; Kreslavsky and Head 2002; Carr and Head 2003; Ramirez and Craddock 2018; Turlin and Forget 2019). Other models have been considered to help determine whether wind-driven waves, which would suggest a relatively warm paleoenvironment, would be necessary to produce the observed morphologies (Parker 1994; Kraal 2006), or if ice shoving in a cold climate would have been more likely (Parker and Currey 2001; Barnhart et al. 2005). Recent work identifying paleo-tsunami deposits has added to the evidence for ancient paleo-oceans on Mars (e.g., Rodriguez et al. 2016; Costard et al. 2017, 2019; Di Prieto et al. 2021). Identifying ancient shorelines may provide valuable resources for construction materials of particular grain shapes and sizes.

20.2.6 Present-Day Cryosphere and Aqueous Activity

Water on the surface of Mars is currently in the form of ice. Most of this water is contained in the massive northern and southern polar ice caps, Planum Boreum and Planum Australe, respectively (Fig. 20.10). The ice caps are composed of both water and carbon dioxide ice. Although the southern ice cap is smaller than the northern ice cap, unusual features thought to be geysers can be observed erupting from its surface (Kieffer et al. 2006). Planum Australe may have several large saltwater lakes beneath its surface (Lauro et al. 2021). In addition, smaller deposits of surface ice have been

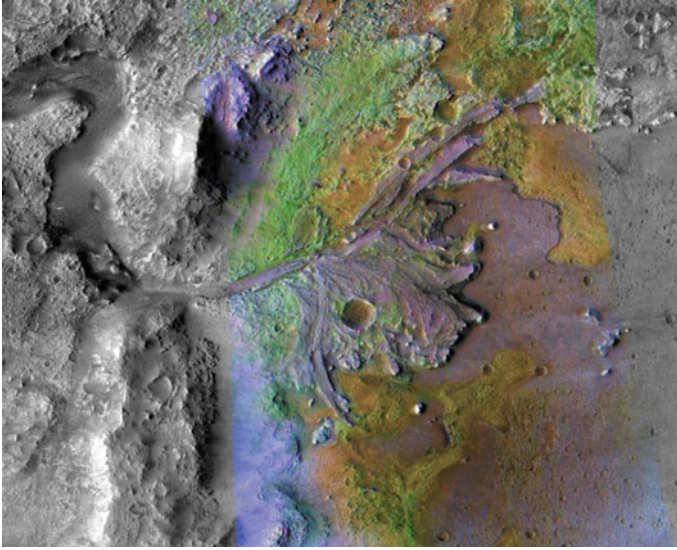


Fig. 20.9 Delta-Like Fan (Jezero) on Mars suggests ancient rivers were present and persistent in the past. Jezero is the landing site for Mars 2020 rover Perseverance. *Image Credit: NASA/JPL-Caltech/MSSS*

detected on the north-facing slopes of impact craters on the northern plains (e.g., Dundas et al. 2013). The Mars Odyssey gamma-ray spectrometer (Boynton et al. 2002, 2004) identified elevated water-equivalent hydrogen at all latitudes on Mars, from the poles to regions at the equator (Feldman et al. 2004), indicating that a combination of ice and water exists beneath much of the Martian surface. Patterned ground, such as polygonal fracturing, provides supportive evidence of the regional presence of permafrost. Lobate debris aprons found on several of the volcanoes are hypothesized to represent ice sheets and glaciers covered by layers of rock and dust (Lucchitta 1981; Scott and Zimbelman 1995; Scott et al. 1998; Plaut et al. 2009). Radar data from the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) onboard the Mars Express spacecraft (Farrell et al. 2005; Mouginot et al. 2010) and the Shallow Subsurface Radar (SHARAD, Fig. 20.11) on the Mars Reconnaissance Orbiter (MRO) (Holt et al. 2008; Phillips et al. 2007) have yielded greater detail as to the quantity of subsurface water ice found on Mars. Soil samples from the Mars Phoenix lander confirmed the presence of shallow subsurface ice (Smith et al. 2009).

Transient flow features, in the form of dark slope streaks and recurring slope lineae (RSL) have been observed on Martian slopes in orbital images acquired from MGS. These features are narrow (0.5–5 m) and exhibit relatively dark markings on steep (25°–40°) slopes (Fig. 20.12). RSL appear during the Martian warm seasons and fade when temperatures get colder (McEwen et al. 2011). Wide-ranging interpretations of their formation include aqueous activity (Ferris et al. 2002) and dust avalanching (Sullivan et al. 2001). One hypothesis suggests that brines from salty groundwater

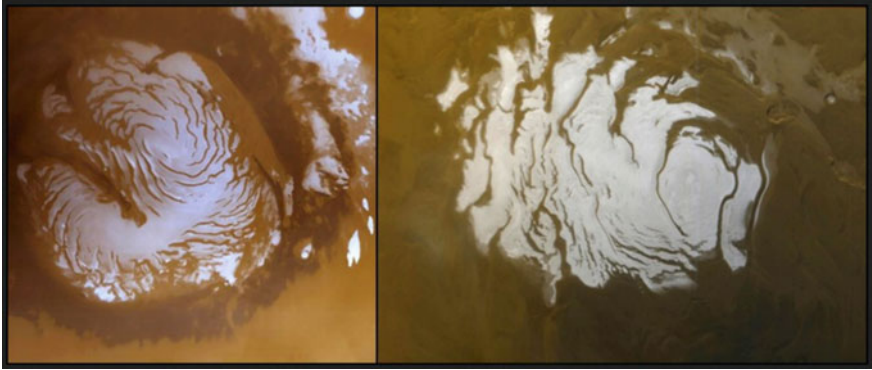


Fig. 20.10 Like Earth, Mars has polar ice caps on each pole. Left: the Martian North Polar Ice Cap (Planum Boreum); Right: the Southern Polar Ice Cap (Planum Australe). These two Martian ice caps are primarily made of frozen water. During the Martian winter, the permanent water ice of the poles is protected by a seasonal layer of frozen carbon dioxide. *Image Credit: NASA*

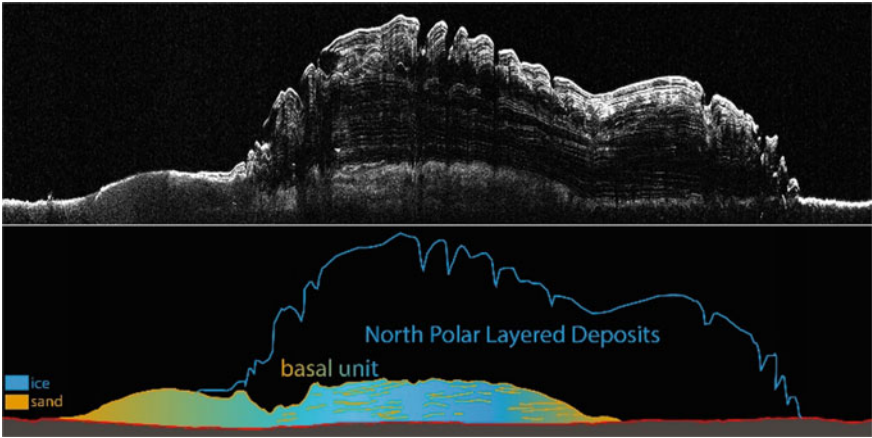


Fig. 20.11 Data from SHARAD, a ground-penetrating radar aboard NASA's Mars Reconnaissance Orbiter, provided detailed radar profiles of alternating layers of ice and sand buried beneath the north polar cap. *Image Credit: NASA/ASI/Nerozzi*

reservoirs of local and regional extent in the cratered southern highlands seep at steep slopes, transiently wetting the Martian surface (Burt and Knauth 2003). In addition to mining directly for water/ice, current groundwater activity influences the emplacement of evaporites and other sites of mineralization that could be valuable for future mining.

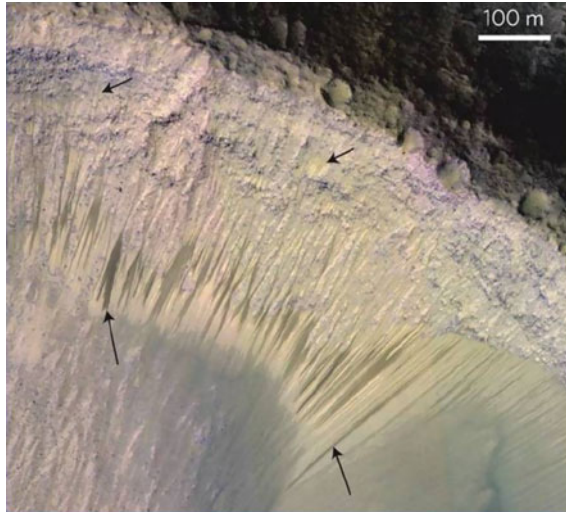


Fig. 20.12 NASA Mars Reconnaissance Orbiter's HiRISE image of recurring slope lineae (RSL) in Melas Chasma, Valles Marineris. Arrows point to several of the identified lineae. *Image Credit:* © NASA/JPL-Caltech/University of Arizona

20.2.7 Mineralogy/Geochemistry

The mineralogy of Mars is indicative of both its volcanic nature and the presence of surface water in the past. Orbital observations by the Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité (OMEGA) (Bibring et al. 2006) and the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) (Murchie et al. 2007) indicated that the minerals detected on the Martian surface are largely volcanic primary minerals such as pyroxene and olivine, consistent with the interpretation that Mars has been a volcanically active planet. Chemical analyses of rocks on the surface confirmed that the Martian crust was built of basalt lava flows, not much different from those on Earth (McSween et al. 2009). However, many alteration minerals have also been detected, including iron oxides, phyllosilicates (Fig. 20.13), sulfates, and opaline silica (e.g., Murchie et al. 2007; Squyres et al. 2008). These alteration minerals tend to form in the presence of water, suggesting that water was once common on the surface of Mars. However, the presence of different types of alteration minerals suggests changes in environmental conditions over time. Smectites, a common phyllosilicate mineral on Mars, form in neutral pH environments, while sulfates form in more acidic water.

The identification of carbonates on Mars also indicates the one-time presence of liquid water on its surface. The Phoenix lander found 3–5 weight % calcium carbonate (calcite) in its sampling of the soil of the northern plains (Boynton et al. 2009). Since then, CRISM has detected carbonates around the southern highlands



Fig. 20.13 The Curiosity rover drilled into this clay-bearing unit on Mars. Drill hole is 1.6 cm in diameter. *Image Credit: NASA/JPL-Caltech*

(Ehlmann et al. 2008), including locations such as Huygens Crater and Nili Fossae, among others.

Magnesium-iron carbonates were also discovered (16–34 wt%) in the Columbia Hills of Gusev Crater (Morris et al. 2010). The presence of opaline silica suggests hydrothermal deposition (Squyres et al. 2008). The MER Spirit rover detected goethite (Morris et al. 2006), an iron oxide that only forms in the presence of water, in the Columbia Hills. Hematite, another iron oxide, was found by the MER Opportunity rover at Meridiani Planum in the form of prevalent spherical concretions, thought to have formed due to the presence of water in the soil (Squyres et al. 2004). The Opportunity rover also identified jarosite (Klingelhofer et al. 2004), a sulfate that only forms under acidic aqueous conditions (Fairén et al. 2004), further establishing the one-time presence of water in the region.

Thermal Emission Spectrometer (TES) data have been interpreted to indicate that Mars' volcanic plains are composed of two surface units (Bandfield et al. 2000). Surface Type 1, common in the southern highlands, was interpreted to be basalt, while Surface Type 2, prevalent in the northern lowlands, was interpreted as basaltic andesite (Bandfield et al. 2000) or weathered basalt (Wyatt and McSween 2002). Andesite would be consistent if the volcanic plains in the northern lowlands were more chemically evolved and volatile-rich than the volcanic materials in the southern highlands (Bandfield et al. 2000). However, weathered basalts, perhaps with thin coatings of secondary minerals, could form through interaction with water-bearing materials (Wyatt et al. 2002).

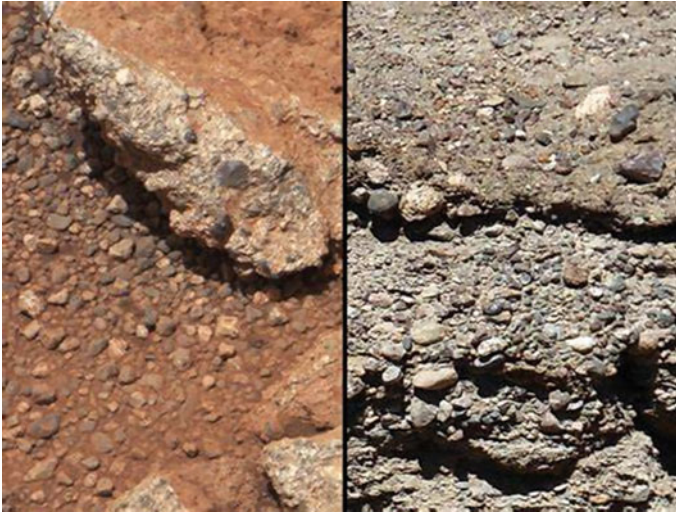


Fig. 20.14 Left: rock outcrop on Mars compared with a terrestrial fluvial conglomerate. Right: conglomerate rocks on Earth indicate the presence of flowing water. *Image Credits: NASA*

Felsic composition rocks have also been identified on Mars, although in sparse quantities and few locations. Alluvial fan deposits (Fig. 20.14) within Gale Crater with a composition similar to terrestrial granites were identified by the Mars Science Laboratory (MSL) Curiosity rover (Baratoux et al. 2014). Utilizing CRISM data, Wray et al. (2013) identified deposits of felsic materials in three different locations on Mars, consistent with the detection of granitic rocks. Eight deposits of anorthosite, containing >90% iron-plagioclase, have also been detected by analysis of CRISM data (Carter and Poulet 2013).

Wray et al. (2013) suggested that the location of the felsic material provided an explanation for how granite could have formed on Mars. On Earth, granite, or its eruptive equivalent, rhyolite, is often found in active subduction zones associated with plate tectonics. But some form of mobile lithosphere (e.g., proto-plate tectonics) has been hypothesized for ancient Mars (Connerney et al. 2005; Baker et al. 2007; Dohm et al. 2007, 2009a, 2013, 2016; Dohm 2017; Yin, 2012a,b). Wray et al. (2013) instead proposed that igneous fractionation occurred, in which the magma slowly cooled in the subsurface, separating low-density melt from higher-density crystals in a process called fractionation. The cycle is repeated over and over for millennia until granite is formed. On Mars, presumably the differentiation enriches the melt in silica, which makes the melt, and eventual rock, lower density and gives it the physical properties of felsic rocks. The presence of large deposits of granitic-like materials would provide a valuable resource for off-Earth mining.

Carter and Poulet (2013) determined that the scarcity of anorthosite detections on Mars suggests this rock formed as localized plutons, similar to how they form on Earth, rather than as a global anorthosite crust, as occurred on the Moon. However,

these authors also concluded that the detections of anorthosite at several locations suggested that magmatic processes that produce highly evolved melts were active on ancient Mars.

20.3 Identifying Geological Features for Mining on Mars

Using terrestrial mining exploration approaches on Mars requires identifying, characterizing, and mapping geologic units to determine the ideal locations to mine materials of interest. Understanding the geologic history, combined with topographic, spectroscopic (both mineralogic and elemental) and geophysical data, can be vital for the selection of prime candidate targets. Through detailed geologic investigation, many anomalous observations have come to light that have a bearing on the interpretation of the geologic history of specific regions (e.g., Baker et al. 2007; Dohm et al. 2013, 2016; Dohm, 2017).

The next steps for future mining exploration and exploitation of Mars will require further analysis into several key areas. For example, areas where anorthosite and felsic crustal materials have been identified may be sources of important incompatible and rare Earth elements. Similarly, regions that expose the evolution and current state of water and ice will require future analyses to determine extractable volumes. Regions that show a high degree of hydrothermal alteration may be sources of secondary minerals important for critical in-situ resources. The locations of relatively low-density materials within the southern highlands offer the possibility of either large expanses of felsic materials or deep sedimentary basins, both of which could potentially host resources. Bands of varying crustal thickness, bimodal topography, and complex structures that have been identified on Mars are similar to structures on Earth that are known to provide areas of concentration for a variety of mineable minerals. All of these areas need further examination to determine if they could represent viable economic resources.

20.4 Can Mars Be Mined like Earth?

Almost all the materials necessary for everyday life on Earth are mined from the crust. The mineralogically diverse crust of Mars points to an elevated mining potential. This diversity includes basalt, basaltic andesite, andesite, granite-like material, anorthosite, sulfates, hematite, opaline quartz, serpentine, clay, carbonate, zeolite, and chlorite (e.g., Borg et al. 1999; Banfield et al. 2000; Bruckner et al. 2001; Chan et al. 2004; Murchie et al. 2007; Boynton et al. 2009; Bishop et al. 2009; Michalski and Niles 2010; Morris et al. 2010; Ehlmann et al. 2010; Carter and Poulet 2013; Halevy et al. 2011; Wray et al. 2013; Viviano-Beck et al. 2014, 2016; Dohm et al. 2018). All of these mineral compositions point to a variety of possible future resources for mining.

Terrestrial mining occurs in areas where large mineral deposits are concentrated, either above or below the surface (e.g., faults, igneous plumes, etc.). Mining on Earth involves a complex and lengthy series of steps using high-technology operations, remote-control equipment, and complex machinery in sensitive or challenging environmental conditions for successful extraction. Recent technological advances in mining applications on Earth will likely be useful for the extraction of resources on Mars. A key example is the achievement of remote mining extraction of the deep ocean basins on Earth (Keck Institute for Space Studies 2020). Such advances in technology will afford astronauts and smart robotic agents access to features and areas on Mars thought to have high mineral potential, but that are difficult or impossible to reach with current engineering and safety constraints. For example, candidate sites of plume-driven manifestations at the Martian surface occur in rugged mountainous terrain at relatively high elevations. Many of these features are likely accompanied by hydrothermal activity (Gulick 1993, 1998; Tanaka et al. 1998; Schulze-Makuch et al. 2007) such as is identified at Warrego Rise, Claritas Rise, and volcanoes along rift systems located in the Thaumasia highlands (Dohm et al. 2001a, b, 2009a). Mountains having crustal roots with plumes interacting with groundwater reservoirs could yield significant mineral reserves similar to those terrains in the western United States, Australia, Africa, and other parts of Earth.

Before terrestrial mining begins, the locations of significant deposits must be identified. This involves using advanced remote sensing and in-situ techniques, including geochemical analysis of soils and rocks (e.g., visible/infrared spectra, Raman/ultraviolet spectra, gas chromatography/mass spectrometry) or geophysical surveys (e.g., magnetic, gravitational, electromagnetic fields), all of which help to determine if there is a sufficient mineral deposit in that location to warrant mining [CRIRSCO (*Committee for Mineral Reserves International Reporting Standards*) *website*]. Once a preliminary investigation is completed, laboratory analyses are required for more in-depth testing to determine whether the concentration of materials are adequate for economic extraction. Mining on Mars would likely follow a similar course.

Based on current knowledge, Martian mining could be divided into three categories: (1) in-situ resource utilization of water and its components of hydrogen and oxygen; (2) in-situ resource utilization of construction materials; and (3) extraction of mineral commodities and non-renewable resources.

20.4.1 In-Situ Resource Utilization of Water

Water is an essential requirement for life, both on and off Earth. Beyond the need of water for survival, the next most immediate relevance is to provide the oxygen needed for life support, as well as hydrogen and oxygen for creating in-situ rocket propellant and other fuel. Presently, spacecraft leaving Earth must carry with them all the propellant and life support needed for both the outgoing and return trips. Once on the surface, propellant for the return trip could be created by extracting hydrogen

and oxygen from ice in the near surface. Additionally, water resources may provide fuels for permanent mining operations. Polar ice deposits, glaciers, and polygonal patterned ground all suggest extensive, readily available water ice deposits on Mars (Christensen 2006), and possible salty, groundwater resources (Ojha et al. 2015). While these water reservoirs have been identified (see Sects. 20.2.3 and 20.2.4), it has yet to be determined if they are present in recoverable volumes sufficient for the habitation of Mars. The utilization of water-rich environments of Mars may face a planetary protection problem or contamination of potential biological niches. Future mining of water resources on Mars will have to take astro-biological aspects into account during extraction.

20.4.2 In-Situ Resource Utilization of Construction Materials

In the future, humans will journey to, land on, and explore the surface of Mars. Once the astronauts arrive on Mars, they will need to prepare for a stay on the surface, either short term or long term, before they can return home. It can be presumed that the crewed ships could serve as temporary habitats. However, for long-term deployment, astronauts will need to construct habitats to live, sleep, and work in. Bringing building materials from Earth to construct these habitats would not be feasible in terms of mass and cost. For example, to launch one kilogram of material to low Earth orbit and then send it to Mars would be economically expensive and logistically very difficult. The preferable way to build habitats would be from the in-situ materials. On Mars, the most abundant material available for building is Martian “soil”, or regolith, consisting primarily of pulverized basaltic rock that ranges from dust- to gravel- to boulder-sized in scale. A major challenge is to develop workable technologies to convert this regolith into a usable substance, suitable for in-situ construction.

On Earth, the most common building material is concrete, made from a combination of limestone (calcium carbonate), clay, and water, but large deposits of calcium carbonate have not been identified on Mars to date. The question that then arises is whether something different on Mars could serve as a substitute for construction materials. Observations by recent orbiters and landers have demonstrated that Mars has significant sulfur in its soil (Rieder et al. 1997), which has been used on Earth in the formulation of some concrete (Omar and Issa 1994; Casanova 1997; Grugela and Toutanji 2008). Additive manufacturing (AM) is the industrial production name for 3D printing, a computer-controlled process that creates three-dimensional objects by depositing materials, usually in layers, and may be the answer for future space missions. For example, NASA’s Perseverance rover, landed on Mars on February 18, 2021, includes 11 metal parts made with 3D printing. In addition, ESA and NASA have been working on this new technique that could allow the first humans on Mars to 3D print everything from tools to temporary housing out of a tough rubber-like material, using only Martian dust (Jakus et al. 2017).

20.4.3 Extraction of Mineral Commodities

A commodity is a physical good that can be bought or sold and traded on the open market. Mineral (hard) commodities are natural resources like oil, gold, and rubber that can be mined and extracted. On Earth, some of the major mineable commodities are chromium, cobalt, copper, gold, lead, zinc, nickel, silver, manganese, and iron ore. The question arises as to whether these elements and/or minerals have formed on Mars. Minerals identified on Mars to date from the numerous landed missions are primarily associated with mafic rocks; these include olivine, pyroxene, and plagioclase feldspar. Additional elements identified include iron-forming minerals (e.g., hematite and magnetite), sulfate, gypsum, clay, and opal. It is possible that Mars contains ore deposits that would be very useful for mining due to the abundance of volcanic rocks (West and Clarke 2010).

Rare earth elements (REE) could be abundant on Mars. Yttrium, lanthanum, and the other 15 minerals which make up the group known as rare earth elements are necessary for the development of key components found within many modern technologies, such as computers, smartphones, wind turbines, hybrid cars, and cruise missiles (Balaram 2019). Although relatively abundant in terrestrial deposits, extraction requires laborious and waste-intensive processing to separate these elements from the surrounding rock. Because these elements are so vital to modern-day technology, corporations are now searching other solar system bodies beyond the Earth's reaches for new deposits, whether it be from the Moon, nearby asteroids, or Mars. *Mining Technology* reported in 2017 that "there are areas on Mars, especially large igneous provinces (LIP), volcanoes and impact craters that hold significant potential for nickel, copper, iron, titanium, platinum group elements and more." To date, no deposits have been found on Mars that would justify the high cost of transport to Earth. However, the presence of LIPs, among other potential mining targets, supports the potential for future mining prospects on Mars.

20.5 Conclusion

Off-Earth mining of rocky planets and small planetary bodies, such as asteroids, could be operational in the relatively near term. Investigations of Mars, derived from orbital, landing, and roving missions, have revealed many traits similar to Earth. Similarities include: (1) a recorded ancient dynamo/magnetosphere; (2) a relatively thick ancient atmosphere; (3) ancient hydrological cycling, including rivers, lakes, oceans, glaciers, and ice sheets; (4) dynamic magmatic and tectonic activity; (5) magma plume manifestation among groundwater reservoirs (e.g., hydrothermal alteration); (6) mountain-composing, potentially felsic rock materials; and (7) a varied mineralogic composition. Mars' similarities to Earth, along with evidence for extant water ice, suggests that in-situ materials could be used for life support, energy/fuel,

construction of habitats, and possible extraction of rare and exotic commodities. This makes Mars a major target for human colonization, exploration, and mining.

Acknowledgements We would like to thank Dr Vic Baker and Dr Goro Komatsu for their feedback, which improved the quality of this manuscript. Part of the research was funded by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). © 2022. All rights reserved.

References

- Acuña, M.H., E.P. Connerney, P. Wasilewski, et al. 2001. Magnetic field of Mars: summary of results from the aerobraking and mapping orbits. *Journal of Geophysical Research* 106: 23,403–23,417. <https://doi.org/10.1029/2000JE001404>.
- Acuña, M.H., J.E.P. Connerney, N.F. Ness, et al. 1999. Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793. <https://doi.org/10.1126/science2845415790>.
- Acuña, M.H., J.E.P. Connerney, P. Wasilewski, et al. 1998. Magnetic field and plasma observations at Mars: initial results of the Mars Global Surveyor Mission. *Science* 279: 1676–1680.
- Anderson, R.C., J.M. Dohm, M.P. Golombek, et al. 2001. Significant centers of tectonic activity through time for the western hemisphere of Mars. *Journal of Geophysical Research* 106: 20563–20585. <https://doi.org/10.1029/2000JE001278>.
- Anderson, R.C., J.M. Dohm, A.F.C. Haldemann, et al. 2004. Tectonic histories between Alba Patera and Syria Planum, Mars, Mars. *Icarus* 171: 31–38.
- Anderson, S., and R.E. Grimm. 1998. Rift processes at the Valles Marineris, Mars: constraints from gravity on necking and rate-dependent strength evolution *Journal of Geophysical Research E: Planets*, 103 (E5): 11113–11124.
- Anderson, R.C., J.M. Dohm, J.P. Williams, S.J. Robbins, S. Andrew, M.P. Golombek, and J.F. Schroeder. 2019. Unraveling the geologic and tectonic history of the Memnonia-Sirenum region of Mars: implications on the early formation of the Tharsis rise. *Icarus* 332: 132–150.
- Andrews-Hanna, J.C. 2012a. The formation of Valles Marineris: 1. Tectonic architecture and the relative roles of extension and subsidence. *Journal of Geophysical Research* 117(E03006). <https://doi.org/10.1029/2011JE003953>.
- Andrews-Hanna, J.C. 2012b. The formation of Valles Marineris: 2. Stress focusing along the buried dichotomy boundary. *Journal of Geophysical Research* 117(E04009). <https://doi.org/10.1029/2011JE003954>.
- Andrews-Hanna, J.C. 2012c. The formation of Valles Marineris: 3. Super-isostasy, stress, sedimentation, and subsidence. *Journal of Geophysical Research* 117(E06002). <https://doi.org/10.1029/2012JE004059>.
- Arkani-Hamed, J. 2004. Timing of the Martian core dynamo. *Journal of Geophysical Research* 109. <https://doi.org/10.1029/2003JE002195>.
- Baker, V.R., and J.B. Patridge. 1986. Small Martian valleys: Pristine and degraded morphology. *Journal of Geophysical Research* 91: 3561–3572.
- Baker, V.R. 1982. *The channels of Mars*. Austin: University of Texas Press.
- Baker, V. 2001. Water and the Martian landscape. *Nature* 412: 228–236.
- Baker, V.R. 1978. The Spokane Flood controversy and the Martian outflow channels. *Science* 202: 1249–1256.
- Baker, V.R., S. Maruyama, and J.M. Dohm. 2007. Tharsis superplume and the geological evolution of early Mars. In *Superplumes: Beyond plate tectonics*, ed. Yuen, D.A., Maruyama, S., Karato, S.-I., Windley, B.F., 507–522. Springer.

- Baker, V.R., R.G. Strom, V.C. Gulick, et al. 1991. Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature* 352: 589–594.
- Balaram, V. 2019. Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact, *Geoscience Frontiers*, 10(4), 2019. ISSN 1285–1303: 1674–9871.
- Bandfield, J.L., V.H. Hamilton, and P.R. Christensen. 2000. A global view of Martian surface composition from MGS-TES. *Science* 287: 1626–1630.
- Baratoux, D., H. Samuel, C. Michaut, et al. 2014. Petrological constraints on the density of the Martian crust. *Journal of Geophysical Research* 119: 1707–1727. <https://doi.org/10.1002/2014JE004642>.
- Barnhart, C. J., S. Tulaczyk, E. Asphaug, E. R. Kraal, J. Moore, et al. 2005. Ice-Ridge pile up and the genesis of martian “Shorelines”. *Lunar and Planetary Science XXXVI, Part 2*.
- Bibring, J.P., Y. Langevin, J.F. Mustard, F. Poulet, R. Arvidson, A. Gendrin, B. Gondet, N. Mangold, P. Pinet, and F. Forget. 2006. Global mineralogical and aqueous mars history derived from OMEGA/Mars express data. *Science* 312 (5772): 400–404. <https://doi.org/10.1126/science.1122659>.
- Bishop, J.L., M. Parente, C.M. Weitz, et al. 2009. Mineralogy of Juventae Chasma: sulfates in the light-toned mounds, mafics in the sand, and opal in the plains. *Journal of Geophysical Research* 114(E00D09). <https://doi.org/10.1029/2009JE003352>.
- Borg, L., J.N. Connelly, L.E. Nyquist, et al. 1999. The age of the carbonates in martian meteorite ALH84001. *Science* 286: 90–94.
- Bourke, M.C., et al. 2010. Extraterrestrial dunes: an introduction to the special issue on planetary dune systems. *Geomorphology* 121: 1–14.
- Boynton, W.V., W.C. Feldman, I. Mitrofanov, et al. 2004. The Mars odyssey gamma-Ray spectrometer instrument suite. *Space Science Reviews* 110: 37–83.
- Boynton, W.V., W.C. Feldman, S.W. Squyres, et al. 2002. Distribution of hydrogen in the near-surface of Mars: evidence for subsurface ice deposits. *Science* 297: 81–85.
- Boynton, W., D.W. Ming, S.P. Kounaves, et al. 2009. Evidence for calcium carbonate at the Mars Phoenix landing site. *Science* 325: 61–64. <https://doi.org/10.1126/science.1172768>.
- Brückner, J., G. Dreibus, R. Rieder, and H. Wanke. 2001. Revised data of the Mars pathfinder alpha proton X-ray spectrometer: geochemical behavior of major and minor elements. In: *Abstracts of the 32nd lunar and planetary science conference*, Houston, Texas, 12–16 March 2001.
- Brückner, J., G. Dreibus, R. Gellert, S. Squyres, H. Wänke, A. Yen, and Jutta Zipfel. 2008. Mars exploration rovers-chemical composition by APXS. In *The martian surface: composition, mineralogy, and physical properties*, 58–101.
- Bryan, W.B. 1973. Wrinkle ridges as deformed surface crust on ponded mare lava. In *Proceedings of the lunar science conference*, 4th, 93–106.
- Burt, D.M., and L.P. Knauth. 2003. Electrically conducting, Ca-rich brines, rather than water, expected in the Martian subsurface. *Journal of Geophysical Research* 108 (E4): 8026. <https://doi.org/10.1029/2002JE001862>.
- Cabrol, N.A., and E.A. Grin. 1999. Distribution, classification, and ages of Martian impact crater lakes. *Icarus* 142: 160–172. <https://doi.org/10.1006/icar.19996191>.
- Cabrol, N.A., and E.A. Grin. 2001. The evolution of lacustrine environments on Mars: Is Mars only hydrologically dormant? *Icarus* 149: 291–328. <https://doi.org/10.1006/icar.20006530>.
- Cabrol, N.A., E.A. Grin, R. Landheim, et al. 1998. Duration of the Ma’adim Vallis/Gusev crater hydrogeologic system, Mars. *Icarus* 133: 98–108.
- Cabrol, N.A., D.D. Wynn-Williams, D.A. Crawford, and E.A. Grin. 2001. Recent aqueous environments in Martian impact craters: an astrobiological perspective. *Icarus* 154: 98–113.
- Carter, J., and F. Poulet. 2013. Ancient plutonic processes on Mars inferred from the detection of possible anorthositic terrains. *Nature Geosciences* 6: 1008–1012. <https://doi.org/10.1038/ngeo1995>.
- Carr, M.H., and J.W. Head. 2003. Oceans on Mars: an assessment of the observational evidence and possible fate. *Journal of Geophysical Research* 108(E5). <https://doi.org/10.1029/2002JE001963>.

- Casanova, I. 1997. Feasibility and application of sulfur concrete for lunar base development: a preliminary study. In *28th annual lunar and planetary science conference*, p. 209. March 17–21, 1997, Houston, TX.
- Chan, M.A., B. Beutler, W.T. Parry, et al. 2004. A possible terrestrial analogue for haematite concretions on Mars. *Nature* 429: 731–734.
- Christensen, P.R. 2006. Water at the poles and in permafrost regions of Mars. *Elements* 3 (2): 151–155. <https://doi.org/10.2113/gselements.2.3.151>.
- Clifford, S.M., and T.J. Parker. 2001. The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* 154: 0–79.
- Connerney, J.E.P., M.H. Acuña, N.F. Ness, et al. 2005. Tectonic implications of Mars crustal magnetism. *Science* 102: 14970–14975.
- Connerney, J.E.P., M.H. Acuña, P. Wasilewski, et al. 2001. The global magnetic field of Mars and implications for crustal evolution. *Geophysical Research Letters* 28: 4015–4018.
- Connerney, J.E.P., M.H. Acuña, P.J. Wasilewski, et al. 1999a. The global magnetic field of Mars and implications for crustal evolution. *Science* 284: 790–793.
- Connerney, J.E.P., M.H. Acuña, P.J. Wasilewski, et al. 1999b. Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
- Costard, F., A. Séjourné, K. Kelfoun, S. Clifford, F. Lavigne, I. Di Pietro, et al. 2017. Modeling tsunami propagation and the emplacement of thumbprint terrain in an early Mars ocean. *Journal of Geophysical Research: Planets* 122: 633–649.
- Costard, F., A. Séjourné, A. Lagain, J. Ormö, J.A.P. Rodriguez, S. Clifford, et al. 2019. The Lomonosov crater impact event: A possible mega-tsunami source on Mars. *Journal of Geophysical Research: Planets* 124: 1840–1851.
- Craddock, R.A., and A.D. Howard. 2002. The case for rainfall on a warm, wet early Mars. *Journal of Geophysical Research* 107 (E11): 5111. <https://doi.org/10.1029/2001JE001505>.
- Crown, D.A., K.H. Price, R. Greeley, R. et al. 1992. Geologic evolution of the east rim of the Hellas basin Mars. *Icarus* 100: 1–25. [https://doi.org/10.1016/0019-1035\(92\)90014-X](https://doi.org/10.1016/0019-1035(92)90014-X).
- Crown, D.A., and R. Greeley. 2007. Geologic map of MTM –30262 and –30267 quadrangles, Hadriaca Patera region of Mars. *US Geological Survey Report SIM 2936* (1:1,004,000).
- Di Achille, G., and B.M. Hynek. 2010. Ancient ocean on Mars supported by global distribution of deltas and valleys. *Nature Geoscience* 3(7): 459–463. <https://doi.org/10.1038/ngeo891>
- Di Prieto, I.D., et al. 2021. Evidence of mud volcanism due to the rapid compaction of martian tsunami deposits in southeastern Acidalia Planitia, Mars. *Icarus*, 354. <https://doi.org/10.1016/j.icarus.2020.114096>.
- Dohm, J.M., et al. 2015. Geological and hydrological histories of the Argyre province, Mars. *Icarus* 253: 66–98.
- Dohm, J.M., R.C. Anderson, J.-P. Williams, et al. 2009a. Claritas rise: Pre-Tharsis magmatism. *Journal of Volcanology and Geothermal Research* 185: 139–156.
- Dohm, J.M., V.R. Baker, W.V. Boynton, et al. 2009b. GRS evidence and the possibility of ancient oceans on Mars. *Planet Spa Sci* 57: 664–684. <https://doi.org/10.1016/jps.2008.10.008>.
- Dohm, J.M., J.-P. Williams, R.C. Anderson, et al. 2009c. New evidence for a magmatic influence on the origin of Valles Marineris, Mars. *Journal of Volcanology and Geothermal Research* 185: 12–27.
- Dohm, J.M., J.C. Ferris, V.R. Baker, et al. 2001a. Ancient drainage basin of the Tharsis region, Mars: Potential source for outflow channel systems and putative oceans or paleolakes. *Journal of Geophysical Research* 106: 32,942–32,958. <https://doi.org/10.1029/2000JE001468>.
- Dohm, J.M., K.L. Tanaka, and T.M. Hare. 2001b. Geologic map of the Thaumasia region of Mars. USGS Miscellaneous Investigations Series Map I-2650 (1:5,000,000).
- Dohm, J.M., S. Maruyama, V.R. Baker, R.C. Anderson. 2007. Traits and evolution of the Tharsis superplume, Mars. In *Superplumes: beyond plate tectonics* ed. Yuen, D.A., Maruyama, S., Karato, S.-I., Windley, B.F., 523–537. Springer.
- Dohm, J.M., H. Miyamoto, S. Maruyama, et al. 2013. Mars evolution. In *Mars: evolution, geology, and exploration*, ed. Fairén, A.G., 1–33. Nova Science Publishers.

- Dohm, J.M., S. Maruyama, M. Kido, V.R. Baker. 2018. A possible anorthositic continent of early Mars and the role of planetary size for the inception of Earth-like life. *Geoscience Frontiers* 9(4):1085–1098.
- Dohm, M., and K.L. Tanaka. 1999. Geology of the Thaumasia region, Mars: plateau development, valley, and magmatic evolution. *Planet Spa Sci* 47: 411–431. [https://doi.org/10.1016/S0032-0633\(98\)00141-X](https://doi.org/10.1016/S0032-0633(98)00141-X).
- Dohm, J.M., et al. 2016. Non-unique systems of features on Mars and Earth: possible telltale signatures of ancient dynamic lithospheric mobility including plate tectonism. In: *Abstracts of the 47th lunar and planetary science conference, Houston, Texas*, 21–25 March 2016.
- Dohm, J.M. et al. 2011. An inventory of potentially habitable environments on Mars: geological and biological perspectives. In *Analogues for planetary exploration: geological society of America special paper* 483, ed. Garry, W.B., and Bleacher, J.E., 317–347. ISBN: 9780813724836. [https://doi.org/10.1130/2011.2483\(21\)](https://doi.org/10.1130/2011.2483(21)).
- Dohm, J.M. 2017. Summarized evolution of Mars. In: *Abstracts of the 48th lunar and planetary science conference, Houston, Texas*, 12–16 March 2017, Abstract 1962.
- Dundas, C. M., S. Byrne, A. S. McEwen, M. T. Mellon, M. R. Kennedy, I. J. Daubar, L. Saper. 2013. HiRISE observations of new impact craters exposing martian ground ice. *Journal of Geophysical Research* 119:109–127.
- Ehlmann, B.L., J.F. Mustard, and S.L. Murchie. 2010. Geologic setting of serpentine deposits on Mars. *Geophysical Research Letters* 37(L06201). <https://doi.org/10.1029/2010GL042596>.
- Ehlmann, B.L., J.F. Mustard, S.L. Murchie. et al. 2008. Orbital identification of carbonate-bearing rocks on Mars. *Science* 322(5909):1828–1832.
- El Maarry, M.R., J.M. Dohm, G.A. Marzo, et al. 2012. Searching for evidence of hydrothermal activity at Apollinaris Mons, Mars. *Icarus* 217: 297–314.
- Fairén, A.G., and J.M. Dohm. 2004. Age and origin of the lowlands of Mars. *Icarus* 168: 277–284. <https://doi.org/10.1016/j.icarus.2003.11.025>.
- Fairén, A.G., J.M. Dohm, V.R. Baker, et al. 2003. Episodic flood inundations of the northern plains of Mars. *Icarus* 165: 53–67. [https://doi.org/10.1016/S0019-1035\(03\)00144-1](https://doi.org/10.1016/S0019-1035(03)00144-1).
- Fairén, A.G., D. Fernández-Remolar, J.M. Dohm, et al. 2004. Inhibition of carbonate synthesis in acidic oceans on early Mars. *Nature* 431: 423–426. <https://doi.org/10.1038/nature02911>.
- Farrell, W.M., et al. 2005. Detecting sub-glacial aquifers in the north polar layered deposits with Mars Express/MARSIS. *Geophysical Research Letters* 32(11). <https://doi.org/10.1029/2005GL022488>.
- Farrell, W.M., P.H. Smith, G.T. Delory, et al. 2004. Electric and magnetic signatures of dust devils from the 2000–2001 MATADOR desert tests. *Journal of Geophysical Research* 109(E03004). <https://doi.org/10.1029/2003JE002088>.
- Feldman, W.C., T.H. Prettyman, S. Maurice, et al. 2004. Global distribution of near-Surface hydrogen on mars. *Journal of Geophysical Research* 109(E9). <https://doi.org/10.1029/2003JE002160>.
- Ferrill, D.A., D.Y. Wyrick, A.P. Morris, et al. 2004. Dilational fault slip and pit chain formation on Mars. *GSA Today* 14(10). <https://doi.org/10.1130/1052-5173>.
- Ferris, J.C., J. M. Dohm, V.R. Baker, and T. Maddock. 2002. Dark slope streaks on Mars: Are aqueous processes involved? *Geophysical Research Letters* 29(10): 128-1–128-4. <https://doi.org/10.1029/2002GL014936>.
- Formisano, V., S. Atreya, T. Encrenaz, et al. 2004. Detection of methane in the atmosphere of Mars. *Science* 306: 1758–1761. <https://doi.org/10.1126/science.1101732>.
- Gierasch, P.J. 1974. Martian dust storms. *Reviews of Geophysics* 12 (4): 730–734.
- Gierasch, P.J., and R.M. Goody. 1973. A model of a Martian great dust storm. *Journal of Atmospheric Science* 30(2): 169–179.
- Golombek, M.P., et al. 2003. Selection of the Mars Exploration Rover landing sites. *Journal of Geophysical Research* 108: 8072. <https://doi.org/10.1029/2003JE002074>,E12.

- Golombek, M.P., F.S. Anderson, M.T. Zuber. 2001. Martian wrinkle ridge topography: evidence for subsurface faults from MOLA. *Journal of Geophysical Research* 106(E10): 23,811–23,821. <https://doi.org/10.1029/2000JE001308>.
- Grant, J.A., T.J. Parker. 2002. Drainage evolution of the margaritifer sinus region, Mars. *Journal of Geophysical Research* 107(E9): 4-1–4-19. <https://doi.org/10.1029/2001JE001678>.
- Greeley, R., and J.E. Guest. 1987. Geologic map of the eastern equatorial region of Mars. USGS Miscellaneous Investigations Series Map I-1802B (1:15,000,000).
- Grugela, R.N., and H. Toutanji. 2008. Sulfur concrete for lunar applications-Sublimation concerns. *Advances in Space Research* 41(1): 103–112. *Bibcode:2008AdSpR..41..103G*. <https://doi.org/10.1016/j.asr.2007.08.018>.
- Gulick, V.C. 1993. Magmatic intrusions and hydrothermal systems: Implications for the formation of small Martian valleys. PhD Thesis, University of Arizona, Tucson.
- Gulick, V.C. 1998. Magmatic intrusions and a hydrothermal origin for fluvial valleys on Mars. *Journal of Geophysical Research* 103: 19,365–19,387. <https://doi.org/10.1029/98JE01321>.
- Haberle, R.M. 2015. Solar system/sun, atmospheres, evolution of atmospheres/planetary atmospheres: Mars. In *Encyclopedia of atmospheric sciences*, 2nd edn. ed, North, Gerald, R., Pyle, John, Zhang, Fuqing, 168–177. Academic Press. <https://doi.org/10.1016/b978-0-12-382225-3.00312-1>. ISBN 9780123822253.
- Halevy, I., W.F. Woodward, and J.M. Eiler. 2011. Carbonate in the Martian meteorite Allan Hills 84001 formed at 18 ± 4 °C in a near-surface aqueous environment. *PNAS* 108: 16895–16899.
- Hall, J.L., S.C. Solomon, and J.W. Head. 1986. Elysium Region, Mars: tests of lithospheric loading models for the formation of tectonic features. *Journal of Geophysical Research* 91(B11): 11377–11392. <https://doi.org/10.1029/JB091iB11p11377>.
- Halliday, A.N., H. Wänke, J.-L. Birck, and R.N. Clayton. 2001. Accretion, composition and early differentiation of Mars. *Space Science Reviews* 96: 197–230.
- Hauber, E., M. Grott, and P. Kronberg. 2010. Martian rifts: structural geology and geophysics. *Earth and Planetary Science Letters* 294: 393–410. <https://doi.org/10.1016/j.epsl.2009.11.005>.
- Hayward, R.K., K.F. Mullins, L.K. Fenton, et al. 2007. Mars global digital dune database and initial science results. *Journal of Geophysical Research* 112(E11007). <https://doi.org/10.1029/2007JE002943>.
- Head III, J.W., H. Hiesinger, M.A. Ivanov, M.A. Kreslavsky, S. Pratt, and B.J. Thomson. 1999. Possible ancient oceans on Mars: evidence from Mars orbiter laser altimeter data. *Science* 10286: 2134–2137.
- Hiesinger, H., and J. Head. 2002. Topography and morphology of the Argyre Basin, Mars: implications for its geologic and hydrologic history. *Planetary and Space Science* 50 (10): 939–981. [https://doi.org/10.1016/S0032-0633\(02\)00054-5](https://doi.org/10.1016/S0032-0633(02)00054-5).
- Holt, J.W., A. Safaeinili, J.J. Plaut, J.W. Head, R.J. Phillips, R. Seu, S.D. Kempf, P. Choudhary, D.A. Young, N.E. Putzig, D. Biccari, Y. Gim, et al. 2008. Radar sounding evidence for buried glaciers in the southern mid-latitudes of Mars. *Science* 322: 1235–1238.
- Howard, A.D., J.M. Moore, and R.P. Irwin III. 2005. An intense terminal epoch of widespread fluvial activity on early Mars: 1 Valley network incision and associated deposits. *Journal of Geophysical Research* 110(E12S14). <https://doi.org/10.1029/2005JE002459>.
- Irwin III, R.P., A.D. Howard, T.A. Maxwell. 2004. Geomorphology of Ma'adim Vallis, Mars, and associated paleolake basins. *Journal of Geophysical Research* 109(E12009). <https://doi.org/10.1029/2004JE002287>.
- Jakus, A.E., et al. 2017. Robust and elastic lunar and martian structures from 3D-printed regolith Inks. *Science and Reports* 7: 44931. <https://doi.org/10.1038/srep44931>.
- Jakosky, B. M., M. Slipiski, M. Benna, P. Mahaffy, M. Elrod, R. Yelle, S. Stone, N. Alsaeed. 2017. Mars' atmospheric history derived from upper-atmosphere measurements of 38Ar/36Ar. *Science* 355:1408–1410.
- Karasozen, E, J.C. Andrews-Hanna, J.M. Dohm, and R.C. Anderson. 2016. The formation of the South Tharsis Ridge Belt and implications for the early evolution of Tharsis on Mars. *Journal of Geophysical Research* 121. <https://doi.org/10.1002/2015JE004936>.

- Kargel, J., and R. Strom. 1991. Terrestrial glacial eskers: analogs for martian sinuous ridges. In *Abstracts of the 22nd lunar and planetary science conference, Houston, Texas*, 18–22 March 1991.
- Kargel, J.S. 2004. *Mars: a warmer wetter planet*. Praxis-Springer, UK.
- Kargel, J.S., and R.G. Strom. 1992. Ancient glaciation on Mars. *Geology* 20 (1): 3–7. [https://doi.org/10.1130/0091-7613\(1992\)020%3c0003:AGOM%3e23CO;2](https://doi.org/10.1130/0091-7613(1992)020%3c0003:AGOM%3e23CO;2).
- Keck Institute for Space Studies. 2020. Space science opportunities augmented by exploration telepresence, A W. M. Keck Institute for Space Studies workshops, October 3–7, 2016 and July 7–13, 2017.
- Kieffer, H.H., P.R. Christensen, T.N. Titus. 2006. CO₂ jets formed by sublimation beneath translucent slab ice in Mars' seasonal south polar ice cap. *Nature* 442:793–796.
- Klingelhofer, G., R.V. Morris, B. Bernhardt, C. Schroder, D.S. Rodionov, P.A. de Souza Jr, R.E. Arvidson. 2004. Jarosite and hematite at Meridiani Planum from Opportunity's Mossbauer spectrometer. *Science* 306 (5702): 1740–1745.
- Komatsu, G., J.M. Dohm, and T.M. Hare. 2004. Hydrogeologic processes of large-scale tectono-magmatic complexes in Mongolia–southern Siberia and on Mars. *Geology* 32: 325–328. <https://doi.org/10.1130/G202372>.
- Komatsu, G., G.G. Ori, M. Cardinale, J.M. Dohm, V.R. Baker, D.A. Vaz, R. Ishimaru, N. Namiki, and T. Matsui. 2011. Roles of methane and carbon dioxide in geological processes on Mars. *Planetary and Space Science* 59: 169–181. <https://doi.org/10.1016/j.pss.2010.07.002>.
- Kreslavsky, M., and J. Head. 2002. Mars: nature and evolution of young latitude-dependent water-ice-rich mantle. *Geophysical Research Letters* 29(15). <https://doi.org/10.1029/2002GL015392>.
- Kraal, E. R., E. Asphaug, J. M. Moore, and R. D. Lorenz, 2006. Quantitative geomorphic modeling of martian bedrock shorelines. *Journal of Geophysical Research* 111:E3. <https://doi.org/10.1029/2005JE002567>.
- Langlais, B., M.E. Purucker, and M. Manda. 2004. Crustal magnetic field of Mars. *Journal of Geophysical Research* 109(E2). <https://doi.org/10.1029/2003JE002048>.
- Lauro, S.E., E. Pettinelli, G. Caprarelli, et al. 2021. Multiple subglacial water bodies below the south pole of Mars unveiled by new MARSIS data. *Nature Astronomy* 5: 63–70. <https://doi.org/10.1038/s41550-020-1200-6>.
- Lemmon, Mark, M. Wolff, Myron Smith, R. Clancy, D. Banfield, Geoffrey Landis, A. Ghosh, Peter Smith, N. Spanovich, B. Whitney, Patrick Whelley, R. Greeley, Stefan Thompson, J. Bell, and S. Squyres. 2005. Atmospheric imaging results from the mars exploration rovers: spirit and opportunity. *Science (new York, NY)* 306: 1753–1756. <https://doi.org/10.1126/science.1104474>.
- Lillis, R.J., H. Frey, M. Manga, et al. 2008. An improved crustal magnetic field map of Mars from electron reflectometry: highland volcano magnetic history and the end of the martian dynamo. *Icarus* 194: 575–596. <https://doi.org/10.1016/j.icarus.200709032>.
- Lucchitta, B.K. 1981. Mars and earth: comparison of cold-climate features. *Icarus* 45: 264–303.
- Lundin, R., S. Barabash, H. Andersson, M. Holmström, Grigoriev Alexander, M. Yamauchi, Sauvaud Jean-André, A.O. Fedorov, Budnik Elena, J.-J. Thocaven, J. Winningham, R. Frahm, J. Scherrer, Sharber John, Asamura Kazushi, H. Hayakawa, Coates Andrew, D. Linder, C. Curtis, and P. Bochsler. 2004. Solar wind-induced atmospheric erosion at Mars: first results from ASPERA-3 on Mars express. *Science (new York, NY)* 305: 1933–1936. <https://doi.org/10.1126/science.1101860>.
- Mahaffy, P.R., C.R. Webster, S.K. Atreya, H. Franz, M. Wong, P.G. Conrad, D. Harpold, J.J. Jones, L.A. Leshin, H. Manning, T. Owen, R.O. Pepin, S. Squyres, M. Trainer, and M.S.L. Science Team. 2013. Abundance and isotopic composition of gases in the martian atmosphere from the curiosity rover. *Science* 19 Jul 2013: 263–266.
- Malin, M.C., and M.H. Carr. 1999. Groundwater formation of martian valleys. *Nature* 399 Feb 18;397(6720): 589–591. <https://doi.org/10.1038/17551>. PMID: 10050852.
- McEwen, A.S., L. Ojha, and C.M. Dundas. 2011. Seasonal flows on warm martian slopes. *Science* 333 (6043): 740–743. <https://doi.org/10.1126/science.1204816>.

- McKinnon, W.B. 1981. Application of ring tectonic theory to Mercury and other system bodies. In *Multi-ring basins, proceedings of the lunar and planetary science*, ed. P.H. Schultz, R.B. Merrill, 12A, 259–273. New York: Pergamon Press.
- McSween, H.Y., Jr., G.J. Taylor, and M.B. Wyatt. 2009. Elemental composition of the martian crust. *Science* 324 (5928): 736–739. <https://doi.org/10.1126/science.1165871>.
- Mège, D., and P. Masson. 1996. A plume tectonics model for the Tharsis Province, Mars. *Planetary and Space Science* 44: 1499–1546.
- Melosh, H.J., and W.B. McKinnon. 1988. The tectonics of mercury. In *Mercury*, ed. F. Vilas, C.R. Chapman, and M.S. Matthews, 374–400. Tucson: University of Arizona Press.
- Metzger, S.M., and N. Lancaster. 1996. Terrestrial dust devils; implications for the surface of Mars. In *Abstracts of the 27th Lunar and Planetary Science Conference*, Houston, Texas, 18–22 March 1996.
- Mittelholz, A., et al. 2020. Timing of the Martian dynamo: new constraints for a core field 4.5 and 3.7 Ga ago. *Science Advances* 6(18): eaba0513. <https://doi.org/10.1126/sciadv.aba0513>.
- Michalski, J.R., P.B. Niles. 2010. Deep crustal carbonate rocks exposed by meteor impact on Mars. *Nature Geoscience* 3:751–755. <https://doi.org/10.1038/ngeo971>.
- Morris, R.V., S.W. Ruff, R. Gellert, et al. 2010. Identification of carbonate-rich outcrops on Mars by the spirit rover. *Science* 329: 421–424. <https://doi.org/10.1126/science.1189667>.
- Morris, R.V., et al. 2006. Mössbauer mineralogy of rock, soil, and dust at Meridiani Planum, Mars: opportunity's journey across sulfate-rich outcrop, basaltic sand and dust, and hematite lag deposits. *Journal of Geophysical research* 111:E12515.
- Mouginot, J., A. Pommerol, W. Kofman, et al. 2010. The 3–5 MHz global reflectivity map of Mars by MARSIS/Mars express: implications for the current inventory of subsurface H₂O. *Icarus* 210: 612–625. <https://doi.org/10.1016/j.icarus.201007003>.
- Murchie, S., et al. 2007. Compact reconnaissance imaging spectrometer for Mars (CRISM) on Mars Reconnaissance Orbiter (MRO). *Journal of Geophysical Research* 112, E05S03. <https://doi.org/10.1029/2006JE002682>.
- Newsom, H.E., G.E. Brittelle, C.A. Hibbitts, L.J. Crossey, and A.M. Kudo. 1996. Impact crater lakes on Mars. *Journal of Geophysical Research*, 101(E6): 14951–14955. <https://doi.org/10.1029/96J E01139>.
- Nimmo, F., and K.L. Tanaka. 2005. Early crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161.
- Ojha, L., M.B. Wilhelm, S.L. Murchie, A.S. McEwen, J.J. Wray, J. Hanley, M. Massé, and M. Chojnacki. 2015. Spectral evidence for hydrated salts in recurring slope lineae on Mars, *Nature Geoscience*. 8(11): 829–832. <https://doi.org/10.1038/ngeo2546>. (Bibcode:2015NatGe...8..829O. S2CID 59152931).
- Okubo, C.H., and R.A. Schultz. 2003. Thrust fault vergence directions on Mars: A foundation for investigating global-scale Tharsis-driven tectonics. *Geophysical Research Letters* 30: 2154. <https://doi.org/10.1029/2003GL018664.22>.
- Okubo, C.H., R.A. Schultz, and G.S. Stefanelli. 2004. Gridding Mars orbiter laser altimeter data with GMT: effects of pixel size and interpolation methods on DEM integrity. *Computers & Geosciences* 30 (2004): 59–72.
- Okubo, C.H., and Richard A. Schultz. 2006. Variability in early Amazonian Tharsis stress state based on wrinkle ridges and strike-slip faulting. *Journal of Structural Geology* 28 (12): 2006.
- Okubo, C.H., and Richard A. Schultz. 2004. Mechanical stratigraphy in the western equatorial region of Mars based on thrust fault-related fold topography and implications for near-surface volatile reservoirs. *GSA Bulletin* 116(5–6): 594–605. <https://doi.org/10.1130/B25361.1>.
- Omar, H.A., and M. Issa. 1994. Production of lunar concrete using molten sulfur. In *Engineering, construction, and operations in space IV: Space '94; proceedings of the 4th international conference, Albuquerque, New Mexico* ed. Rodney G. Galloway & Stanley Lokaj (eds.), February 26–March 3, 1994. (2. New York: Amer. Soc. Of Civil Engineers, pp. 952–959. ISBN 0872629376).

- Parker, T.J., D.S. Gorsline, R.S. Saunders, D.C. Pieri, and D.M. Schneeberger. 1993. Coastal geomorphology of the Martian northern plains. *Journal of Geophysical Research* 98: 11061–11078.
- Parker, T.J., J.A. Grant, B. J. Franklin, and J.W. Rice. 2001. A comparison of MOC and MOLA observations of northern plains “contacts” with coastal landforms of the Bonneville basin, Utah. In *Abstracts of the 32nd Lunar and Planetary Science conference*, Houston, Texas, 12–16 March 2001.
- Parker, T.J., R.S. Saunders, and D.M. Schneeberger. 1989. Transitional morphology in the west Deuteronilus Mensae region of Mars: implications for modification of the lowland/upland boundary. *Icarus* 82: 111–145.
- Parker, T.J., D.M. Schneeberger, D.C. Pieri, and R.S. Saunders. 1987. Geomorphic evidence for ancient seas on Mars. In *MECA symposium on Mars: evolution of its climate and atmosphere*, LPI Tech Rept 87–01, 9698, Lunar and Planet Inst, Houston, Tex.
- Parker, T.J., and D.S. Gorsline. 1991. Where is the source for Uzboi Vallis, Mars? In: *Abstracts of the 22nd Lunar and Planetary Science conference*, 18–22. Houston, Texas, March 1991.
- Parker, T.J., and D.R. Currey. 2001. Extraterrestrial coastal geomorphology. *Geomorphology* 37:303–328.
- Plaut, J.B., A. Safaeinili, J.W. Holt, et al. 2009. Radar evidence for ice in lobate debris aprons in the mid-northern latitudes of Mars. *Geophysical Research Letter* 36(L02203). <https://doi.org/10.1029/2008GL036379>.
- Phillips, R.J., R. Seu, D. Biccari, et al. 2007. North polar deposits on Mars: new insights from MARSIS, SHRAD, and other MRO instruments. In: *Abstracts of the 38th lunar and planetary science conference*, Houston, Texas, 12–16 March 2007.
- Quantin, C., P. Allemand, N. Mangold, et al. 2005. Fluvial and lacustrine activity on layered deposits in Melas Chasma, Valles Marineris, Mars. *Journal of Geophysical Research* 110(E12S19). <https://doi.org/10.1029/2005JE002440>.
- Ramirez, R. M., and R. A. Craddock, 2018. The geological and climatological case for a warmer and wetter early Mars. *Nature Geoscience* <https://doi.org/10.1038/s41561-018-0093-9>.
- Rieder, R., T. Economou, H. Wänke, A. Turkevich, J. Crisp, J. Brückner, G. Dreibus, and H.Y. McSween Jr. 1997. The chemical composition of martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: preliminary results from the X-ray mode. *Science* 05 Dec 1997: 1771–1774.
- Rivoldini, A., et al. 2011. Geodesy constraints on the interior structure and composition of Mars. *Icarus* 213: 451–472.
- Rodríguez, J.A.P., S. Sasaki, R.O. Kuzmin, et al. 2005. Outflow channel sources, reactivation, and chaos formation, Xanthe Terra, Mars. *Icarus* 175: 36–57. <https://doi.org/10.1016/j.icarus.2004.0025>.
- Rodríguez, J.A.P., A.G. Fairén, K.L. Tanaka, M. Zarroca, R. Linares, T. Platz, and G. Komatsu. 2016. Tsunami waves extensively resurfaced the shorelines of a receding, early Martian ocean. *Scientific Reports* 6 (25106). <https://doi.org/10.1038/srep25106>.
- Rodríguez, J. A. P., A. G. Fairén, K.L. Tanaka, M. Zarroca, R. Linares, T. Platz, G. Komatsu, H. Miyamoto, J. Kargel, J. Yan, V. Gulick, K. Higuchi, V.R. Baker, N. Glines. 2016. Tsunami waves extensively resurfaced the shorelines of a receding, early Martian ocean. *Scientific Reports* 6(25106). <https://doi.org/10.1038/srep25106>.
- Schultz, R.A., and J. Lin. 2001. Three-dimensional normal faulting models of the Valles Marineris, Mars, and geodynamic implications. *Journal of Geophysical Research* 106 (B8): 16549–16566. <https://doi.org/10.1029/2001JB000378>.
- Schultz, R.A., and T.R. Watters. 2001. Forward mechanical modeling of the Amenthes Rupes thrust fault on Mars. *Geophysical Research Letters* 28 (24): 4659–4662.
- Schultz, R.A., C.H. Okubo, and S.J. Wilkins. 2006. Displacement-length scaling relations for faults on the terrestrial plane. *Journal of Structural Geology* 28 (2006): 2182–2193.
- Schulze-Makuch, D., J.M. Dohm, C. Fan, A.G. Fairén, J.A.P. Rodríguez, V.R. Baker, and W. Fink. 2007. Exploration of hydrothermal targets on Mars. *Icarus* 189: 308–324.

- Scott, D.H., M.G. Chapman, J.W. Rice Jr., and J.M. Dohm. 1991a. New evidence of lacustrine basins on Mars: Amazonis and Utopia Planitiae. In *Proceeding of LPSC 22nd*, 53–62.
- Scott, D.H., J.W. Rice Jr., and J.M. Dohm. 1991b. Martian paleolakes and waterways: Exobiological implications. *Origins of Life and Evolution of the Biosphere* 21(3): 189–198. <https://doi.org/10.1007/BF01809447>.
- Scott, D.H., J.M. Dohm, and D.J. Applebee. 1993. Geologic map of science study area 8, Apollinaris Patera region of Mars. USGS Miscellaneous Investigations Series Map I-2351 (1:500,000).
- Scott, D.H., J.M. Dohm, and J.W. Rice Jr. 1995. Map of Mars showing channels and possible paleolake basins. USGS Miscellaneous Investigations Series Map I-2461 (1:30,000,000).
- Scott, D.H., J.M. Dohm, and J.R. Zimbleman. 1998. Geologic map of Pavonis Mons volcano, Mars. USGS Miscellaneous Investigations Series Map I-2561 (1:1,000,000).
- Scott, D.H., and K.L. Tanaka. 1986. Geologic map of the western equatorial region of Mars US Geological Survey Miscellaneous Investigations Series Map I-1802-A (1:15,000,000).
- Scott, D.H., and J.R. Zimbleman. 1995. Geologic map of Arsia Mons volcano, Mars. USGS Miscellaneous Investigations Series Map I-2480 (1:1,000,000).
- Smith, P.H., L.K. Tamppari, R.E. Arvidson, et al. 2009. H₂O at the Phoenix landing site. *Science* 325: 58–61.
- Smith, D.E., et al. 2001. Mars orbiter laser altimeter: experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research* 106 (E10): 23689–23722. <https://doi.org/10.1029/2000JE001364>.
- Smith, D.E., W.L. Sjogren, G.L. Tyler, G. Balmino, F.G. Lemoine, A.S. Konopliv, et al. 1999. The gravity field of Mars: results from mars global surveyor. *Science* 94–97. (01 Oct 1999).
- Squyres, S.W., R.E. Arvidson, S. Ruff, R. Gellert, V. Morris, D.W. Ming, L. Crumpler, J.D. Farmer, D.J. Des Marais, A. Yen, S.M. McLennan, W. Calvin, J.F. Bell, B.C. Clark, A. Wang, T.J. McCoy, M.E. Schmidt, and P.A. de Souza. 2008. Detection of silica-rich deposits on Mars. *Science* 320: 1063–1067.
- Squyres, S.W., J.P. Grotzinger, R.E. Arvidson, et al. 2004. In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* 306:1709–1714. <https://doi.org/10.1126/science.1104559>.
- Strom, R.G. 1972. Lunar mare ridges, rings and volcanic ring complexes. In *The Moon*, ed. S.K. Runcorn, H.C. Urey, Symposium of International Astronomical Union, vol 47, pp. 187–215. Dordrecht, The Neth.: D. Reidel
- Sullivan, R., et al. 2001. Mass movement slope streaks imaged by the Mars Orbiter Camera. *Journal of Geophysical Research* 106 (E10): 23607–23633.
- Tanaka, K.J., J.A. Skinner, J.M. Dohm, et al. 2014. Geologic map of Mars. USGS Investigations Map 3292 (1:20,000,000), pamphlet 43 p.
- Tanaka, K.L., J.A. Skinner, T.M. Hare, T. Joyal, and A. Wenker. 2003. Resurfacing history of the northern plains of Mars based on geologic mapping of Mars Global Surveyor data. *Journal of Geophysical Research* 108: 8043. <https://doi.org/10.1029/2002JE001908.E4>.
- Tanaka, K.L., J.M. Dohm, J.H. Lias, and T.M. Hare. 1998. Erosional valleys in the Thaumasia region of Mars: hydrothermal and seismic origins. *Journal of Geophysical Research* 103: 31407–31419.
- Turbet, M., and F. Forget. 2019. The paradoxes of a late Hesperian Mars Ocean. *Scientific Reports* 9(5717).
- Vaucher, J., D. Baratoux, N. Mangold, et al. 2009a. The volcanic history of central Elysium Planitia: Implications for martian magmatism. *Icarus* 204: 418–442.
- Vaucher, J., D. Baratoux, M.J. Toplis, et al. 2009b. The morphologies of volcanic landforms at central Elysium Planitia: evidence for recent and fluid lavas on Mars. *Icarus* 200: 39–51.
- Viviano-Beck, C.E., S.L. Murchie, A.W. Beck, and J.M. Dohm. 2016. Compositional and structural constraints on the geologic history of eastern Tharsis Rise, Mars. *Icarus*. <https://doi.org/10.1016/j.icarus.2016.09.005>.
- Viviano-Beck, C.E., F.P. Seelos, S.L. Murchie, et al. 2014. Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars. *Journal of Geophysical Research* 119: 1403–1431. <https://doi.org/10.1002/2014JE004627>.

- Watters, T.R., M.S. Robinson, and A.C. Cook. 1998. Topography of lobate scarps on Mercury: new constraints on the planet's contraction. *Geology* 26: 991–994.
- Watters, T.R., and M.S. Robinson. 1999. Lobate scarps and the Martian crustal dichotomy. *Journal of Geophysical Research* 104 (E8): 18981–18990. <https://doi.org/10.1029/1998JE001007>.
- Watters, T.R., P. McGovern, and R.P. Irwin III. 2007. Hemispheres apart: the crustal dichotomy on Mars. *Annual Review of Earth and Planetary Sciences* 35: 621–652.
- Watters Thomas, R. 2004. Elastic dislocation modeling of wrinkle ridges on Mars. *Icarus* 171 (2): 284–294.
- Weitz, C., et al. 2003. Geology of the Melas Chasma candidate landing site for MER. *Journal of Geophysical Research* 108(E12). <https://doi.org/10.1029/2002JE002022>.
- West, M., and J. Clarke. 2010. Potential martian mineral resources: mechanisms and terrestrial analogs. *Planetary and Space Science* 58: 574–582.
- Whaler, K.A., and M.E. Purucker. 2005. A spatially continuous magnetization for Mars. *Journal of Geophysical Research* 110(E09001). <https://doi.org/10.1029/2004JE002393>.
- Wray, J., S. Hansen, J. Dufek, G.A. Swayze, S.L. Murchie, F.P. Seelos, J.R. Skok, R.P. Irwin III, M.S. Ghiorso. 2013. Prolonged magmatic activity on Mars inferred from the detection of felsic rocks. *Nature Geoscience* 6: 1013–1017. <https://doi.org/10.1038/ngeo1994>
- Wichman, R.W., and P.H. Schultz. 1989. Sequence and mechanisms of deformation around the Hellas and Isidis impact basins on Mars. *Journal of Geophysical Research* 94 (B12): 17333–17357. <https://doi.org/10.1029/JB094iB12p17333>.
- Wilkins, S.J., and R.A. Schultz. 2003. Cross faults in extensional settings: stress triggering, displacement localization, and implications for the origin of blunt troughs at Valles Marineris, Mars. *Journal of Geophysical Research* 108: 5056. <https://doi.org/10.1029/2002JE001968>, E6.
- Withers, P., and G.A. Neumann. 2001. Enigmatic ridges on the plains of Mars. *Nature* 410: 652.
- Wray, J., et al. 2013b. Prolonged magmatic activity on Mars inferred from the detection of felsic rocks. *Nature Geoscience* 2013b. <https://doi.org/10.1038/NNGEO1994>.
- Wyatt, M.B., and H.Y. McSween Jr. 2002. Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars. *Nature* 417 (6886): 263–266. <https://doi.org/10.1038/417263a>.
- Wyrick, D., D.A. Ferrill, A.P. Morris, et al. 2004. Distribution, morphology, and origins of martian pit crater chains. *Journal of Geophysical Research* 109(E6). <https://doi.org/10.1029/2004JE002240>.
- Yin, A. 2012a. An episodic slab-rollback model for the origin of the Tharsis rise on Mars: implications for initiation of local plate subduction and final unification of a kinematically linked global plate-tectonic network on Earth. *Lithosphere* 4: 553–593.
- Yin, A. 2012b. Structural analysis of the Valles Marineris fault zone: possible evidence for large-scale strike-slip faulting on Mars. *Lithosphere* 4 (4): 286–330. <https://doi.org/10.1130/L1921>.
- Zuber, M.T., et al. 2000. Internal structure and early thermal evolution of Mars from Mars global surveyor topography and gravity. *Science* 287(5459): 1788–1793. <https://doi.org/10.1126/science.287.5459.1788>. (10 Mar 2000).

Chapter 21

Robotic Deployment and Installation of Payloads on Planetary Surface



Ashitey Trebi-Ollennu, Khaled Ali, Cristina Sorice, Won Kim, Steven Myint, Omair Khan, Philip Bailey, Hallie Abarca, Robert G. Deen, Jeng Yen, Justin N. Maki, Grace Lim, Nythi Udomkesmalee, and Jeffrey Umland

Abstract The InSight (Interior Exploration using Seismic Investigations, Geodesy and Heat Transport) mission is a Discovery Program lander to investigate the internal structure of Mars and the differentiation of the terrestrial planets (Banerdt et al. in *Space Sci Rev* 215:22 2018). The InSight flight system is a close copy of the Mars Phoenix Lander and comprises a lander, cruise stage, heatshield and backshell. The

A. Trebi-Ollennu (✉) · K. Ali · C. Sorice · W. Kim · S. Myint · O. Khan · P. Bailey · H. Abarca · R. G. Deen · J. Yen · J. N. Maki · G. Lim · N. Udomkesmalee · J. Umland
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
e-mail: ashitey@jpl.nasa.gov

K. Ali
e-mail: Khaled.S.Ali@jpl.nasa.gov

C. Sorice
e-mail: Cristina.E.Sorice@jpl.nasa.gov

W. Kim
e-mail: Won.S.Kim@jpl.nasa.gov

S. Myint
e-mail: Steven.Myint@jpl.nasa.gov

O. Khan
e-mail: Mohammed.O.Khan@jpl.nasa.gov

P. Bailey
e-mail: Philip.Bailey@jpl.nasa.gov

H. Abarca
e-mail: Hallie.E.Gengl@jpl.nasa.gov

R. G. Deen
e-mail: bob.deen@jpl.nasa.gov

J. Yen
e-mail: Jeng.Yen@jpl.nasa.gov

J. N. Maki
e-mail: Justin.N.Maki@jpl.nasa.gov

G. Lim
e-mail: Grace.Lim@jpl.nasa.gov

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_21

InSight science payload includes a seismometer, a wind and thermal shield, a heat flow probe and a precision tracking system to measure the size and state of the core, mantle and crust of Mars. InSight is NASA's first successful precision robotics instrument placement and release on another astronomical body since Apollo. This operations breakthrough enabled NASA's InSight lander to detect a 'marsquake', a faint trembling of Mars's surface on 6 April 2019, 128 Martian days after its landing on Mars. This is the first quake detected on an astronomical body other than Earth or the Moon. This chapter describes the operations of the robotics instrument deployment systems (IDS) that successfully deployed the InSight science payload to the surface of Mars, and the planning and command sequence generation process used for its successful deployment. Among its recommendations, the chapter identifies technology gaps in the operations of in-situ manipulators for planetary exploration.

21.1 Introduction

The InSight (Interior Exploration using Seismic Investigations, Geodesy and Heat Transport) mission is a Discovery Program lander to investigate the internal structure of Mars and the differentiation of the terrestrial planets (Banerdt et al. 2018). The InSight flight system is a close copy of the Mars Phoenix lander and comprises a lander, cruise stage, heatshield and backshell. The InSight science payload includes a seismometer (SEIS), (Longonné et al. 2019) and wind and thermal shield (WTS), a heat flow probe (Heat Flow and Physical Properties Package, HP3, Spohn et al. 2018) and a precision tracking system (RISE), (Folkner et al. 2018) to measure the size and state of the core, mantle and crust of Mars.

InSight is NASA's first successful precision robotics instrument placement and release on another astronomical body since Apollo. This operations breakthrough enabled NASA's InSight lander to detect the first known 'marsquake', a faint trembling of Mars's surface on 6 April 2019, 128 Martian days after landing on Mars on 26 November 2018. This is the first quake detected on an astronomical body other than Earth or the Moon. This chapter describes the operations of the Robotics Instrument Deployment Systems (IDS) that successfully deployed the InSight science payload to the surface of Mars. In addition, the chapter describes the IDS planning and command sequence generation process used for the successful deployment of SEIS, WTS and HP3 on the surface of Mars. The paper concludes with recommendations based on the experience gained from InSight IDS operations. This includes technology gaps identified in the operations of in-situ manipulators for planetary exploration.

N. Udomkesmalee
e-mail: nythi.udomkesmalee@jpl.nasa.gov

J. Umland
e-mail: Jeffrey.W.Umland@jpl.nasa.gov

NASA-JPL, 4800 Oak Grove Dr, M/S 157-205, Pasadena, CA 91109, USA

21.2 Robotic System

The InSight Instrument Deployment System (IDS) consists of the Instrument Deployment Arm (IDA), scoop, five-finger “claw” grapple, motor controller, arm-mounted Instrument Deployment Camera (IDC), lander-mounted Instrument Context Camera (ICC), and control software (Fig. 21.1). IDS is responsible for the first precision robotics instrument placement and release (seismometer and heat flow probe instruments) on a planetary surface. These instruments will enable scientists to perform the first comprehensive surface-based geophysical investigation of Mars. Table 21.1 list IDS driving requirements.

21.2.1 Instrument Deployment Arm (IDA)

The InSight IDA is a refurbished flight robotic arm from the Mars Surveyor 2001 lander mission (Bonitz et al. 2000). The IDA is a four degrees-of-freedom backhoe design manipulator with a 1.8 m reach that provides the following motion: yaw (shoulder azimuth, joint 1) and three pitch joints (shoulder elevation, elbow, and wrist, joints 2 through 4, respectively). The IDA links are made of titanium. During normal operations the IDA actuators are capable of generating 35, 120, 65, and 10.5 N-meters of torque at the joint output for joints 1 through 4, respectively. The

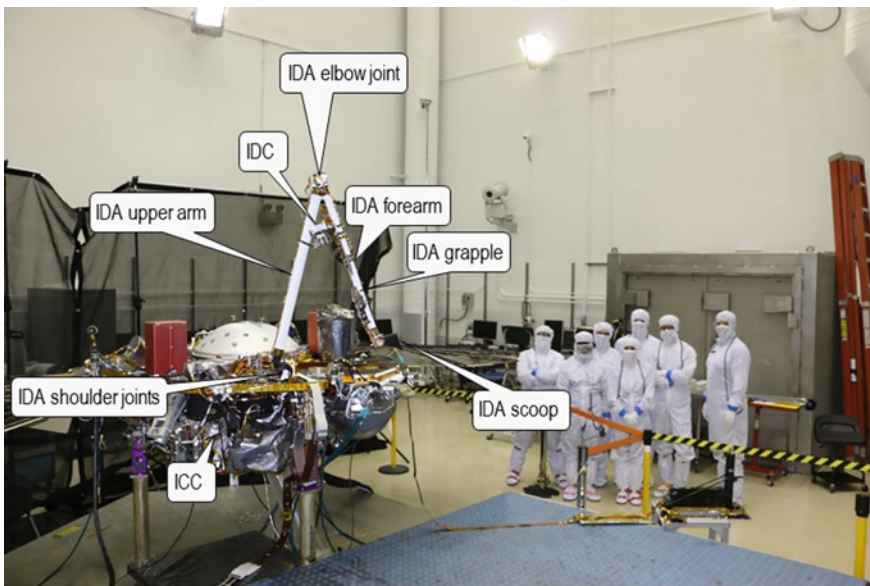


Fig. 21.1 InSight Mars lander with IDS elements labeled

Table 21.1 IDS driving requirements

	IDS driving requirements
1	The IDA shall operate for 4 h and for 111 Martian sols plus 10 h for ground testing
2	The IDS shall positively capture and retain deployable elements, including under loss of power, until placement on the surface is confirmed
3	The IDS shall deploy elements to the surface with the Lander deck tilted 15 degrees w.r.t. gravity
4	The IDS shall have a total mass of less than or equal to 9.41 kg
5	The IDS shall be able to lift a mass of up to 9.5 kg
6	The IDS shall acquire context images of the IDA workspace within the FOV of the InSight Context Camera (ICC)
7	The InSight Deployment Camera (IDC) field of view shall include the IDA end effector during capture and disengagement from deployable elements
8	The IDA shall have repeatability of 0.005 m
9	The IDA shall position the end effector with an absolute error of less than or equal to 0.015 m
10	The IDS shall determine the IDC imaging baseline to within 0.0028 m
11	The IDS shall acquire images of the Lander deck and Solar panels within the field of view of the IDC

IDA can lift and deploy a 9 kg payload on Mars (33 N) at 1.65 m distance. The force the IDA end effector can exert is configuration dependent, but the average force is typically about 80 N.

Each joint has a temperature sensor and heater and includes a dust seal to prevent contamination of the motor and gearbox. The IDA is designed to withstand expected environmental temperatures from -110° to $+70^{\circ}$ °C, in a CO₂ atmosphere, with pressure as low as 5 Torr. Each of the IDA joints consists of a brushed DC motor with two-stage speed planetary gears and a harmonic drive at the output (except the wrist, which has a bevel gear at the output of the planetary gears). The IDA joints do not have mechanical braking systems but employ a dynamic braking system where actively shorting the motor leads slows the motor until magnetic detents capture the rotor. The magnetic detents are sized to provide the appropriate holding torque to assure no slippage while the IDA is powered off. Each joint has two position sensors: encoders on the joint input motor shaft and potentiometers at the joint output load shaft. Each joint is equipped with two mechanical hardstops at the end of their range of travel. The encoder counters are initialized based on potentiometer data or by running each joint up against their respective mechanical hardstops.

The IDA end effector consists of a five-finger “claw” grapple hanging on an umbilical cable, a scoop, and forearm-mounted camera IDC (closer to the elbow joint) facing the IDA end effector.

Two thermal characterization tests were performed on the IDS subsystem in a 13-foot sensor chamber at the Raytheon El Segundo Integrated Test Laboratory (ITL), California (Fig. 21.2). During the test the IDA heaters were characterized and

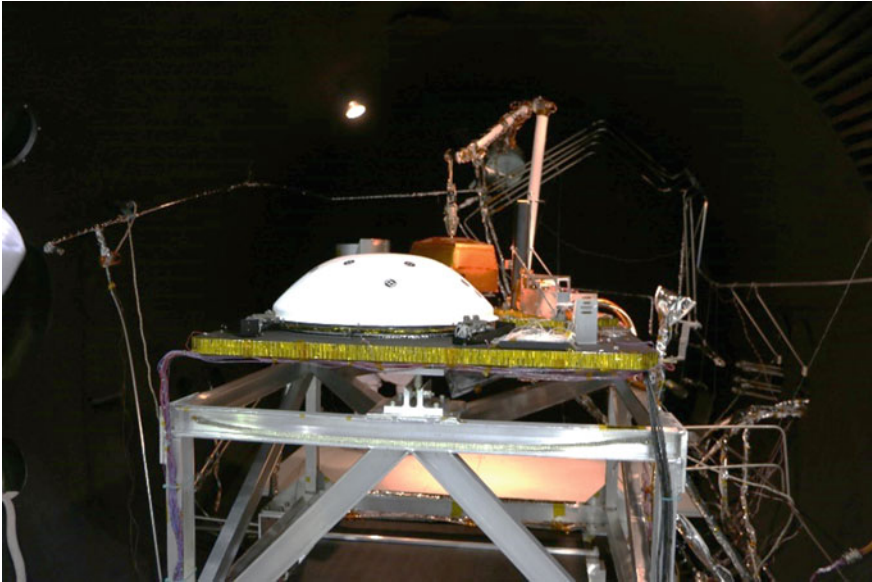


Fig. 21.2 IDS subsystem set-up in a 13-foot sensor chamber at the Raytheon El Segundo Integrated Test Laboratory (ITL), California for thermal characterization test

IDA functional qualification was successfully performed at proto-flight operational temperature. In addition, IDA stop-and-hold torques were characterized at various temperatures.

21.2.2 IDA End Effector Grapple

The grapple is a five-finger “claw” and hangs by an umbilical at the IDA end effector. The grapple is designed with five fingers to assure proper self-alignment and be position error tolerant while closing the grapple fingers around the spherical cap grapple hooks on the payloads. The grapple fingers are forced closed. The grapple fingers are opened by a single high-output paraffin (HOP) actuator that slowly heats up and melts the wax that pushes a rod out to open the fingers. When the fingers are fully open (as shown in Fig. 21.3), a limit contact switch trips and turns the grapple HOP heater power off. As the grapple HOP cools down in the ambient temperature, the grapple fingers slowly close passively without any actuation. The grapple design is robust against unexpected power loss because power is required to open the fingers. The grapple umbilical provides the necessary compliance (unactuated additional 2 DOF for the 4 DOF IDA) for engaging and deploying the payloads on tilted lander and uneven terrain. The grapple is stowed against the IDA forearm such that it does not obstruct the IDC FOV (shown in Fig. 21.3a). However, when the grapple is

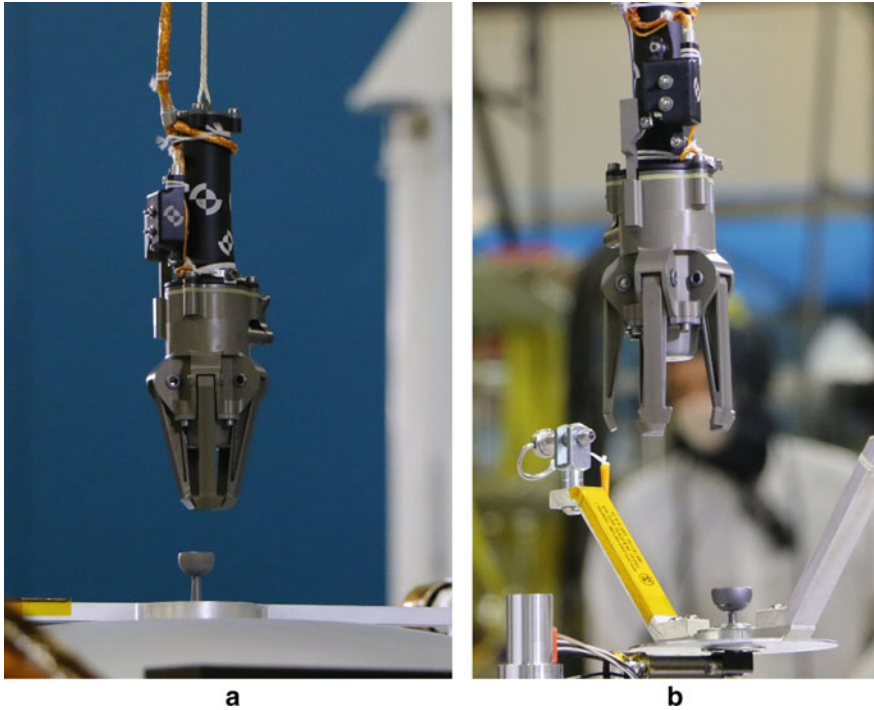


Fig. 21.3 Flight grapple with **a** fingers closed and **b** fingers opened

unstowed (shown in Fig. 21.4b) it hangs in the FOV of the IDC such that the IDC images can capture the opening of grapple fingers and the engagement of spherical cap grapple hooks on the payload.

During deployment the grapple is unstowed, hanging from the IDA end effector by an umbilical cable. The IDA can position the grapple to capture the payload’s spherical cap grapple hook, lift, and place SEIS, WTS and HP3 on the Martian surface. The grapple can be stowed using the IDA in a “ball-and-cup” maneuver to the grapple restraint mechanism on the IDA forearm (shown in Fig. 21.4a).

21.2.3 IDA End-Effector Scoop

The scoop consists of a single chamber with a front blade and a secondary blade on the bottom side (as shown in Fig. 21.5). The scoop’s front and secondary blades can be used to excavate materials (by digging or scraping) and collect materials excavated in the IDA workspace. The scoop will enable soil mechanics experiments for inferring mechanical properties of the Martian soil at the landing site using the IDA housekeeping data (motor currents, scoop position, etc.) to estimate the

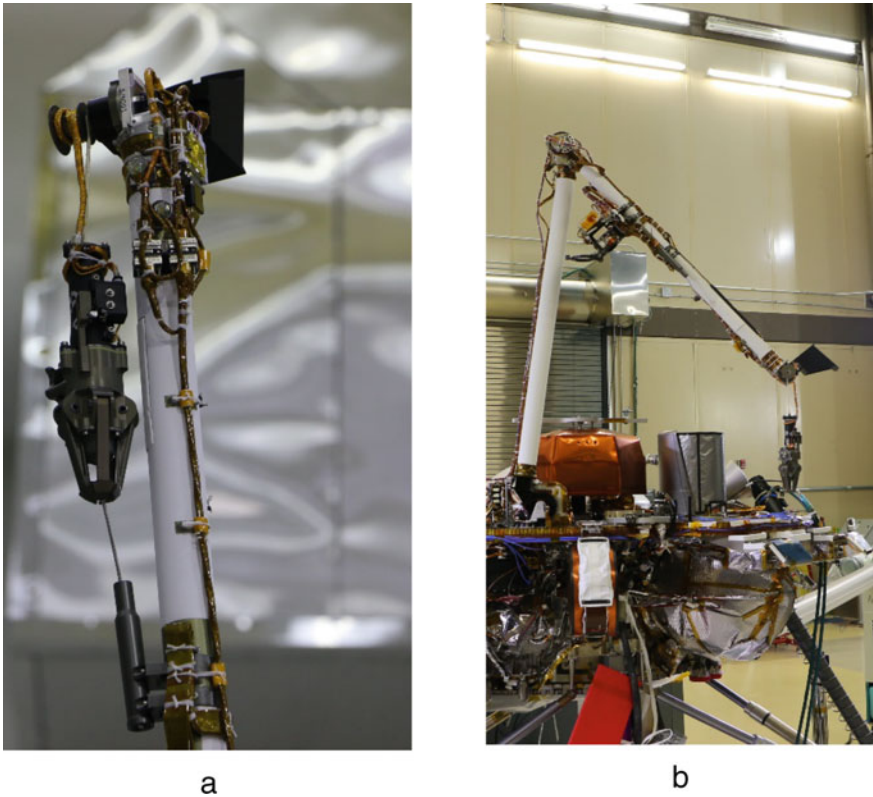


Fig. 21.4 **a** Stowed flight grapple and **b** Grapple unstowed from grapple pre-deployment restraint

scoop's applied force. The scoop is not required for nominal instrument deployment operations.

21.2.4 *IDS Cameras*

To assist in the deployment of the payloads, the robotic arm is equipped with two cameras: the Instrument Deployment Camera (IDC) mounted on the robotic arm and the Instrument Context Camera (ICC) mounted on the lander body underneath the top deck (Fig. 21.6) (Maki et al. 2018). The primary objectives of the IDC and ICC are to: (1) document the state of the lander and surrounding terrain; (2) support terrain assessment for the selection of the SEIS and HP3 instrument deployment locations; (3) facilitate and document the deployment activities; (4) monitor the location and state of the instruments post-deployment; and (5) measure and monitor atmospheric dust opacity (Banfield et al. 2020). The IDC has a FOV of $45^\circ \times 45^\circ$ and an angular

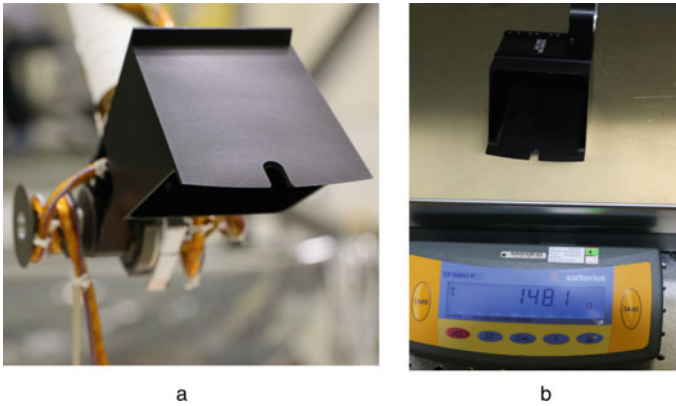
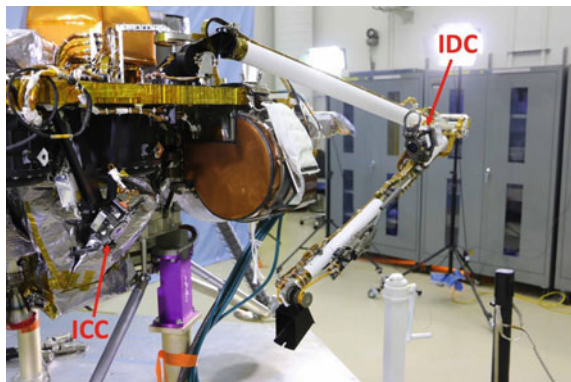


Fig. 21.5 **a** InSight scoop mounted on the IDA and **b** front view of the scoop before installation on the IDA

resolution of 0.82 mrad/pixel at the center of the image. The ICC is mounted to the lander and will acquire wide-angle views of the instrument deployment activities. The ICC has a FOV of $124^\circ \times 124^\circ$ and an angular resolution of 2.1 mrad/pixel at the center of the image. The IDC and ICC cameras are flight spare engineering cameras from the Mars Science Laboratory (MSL) mission. The InSight project upgraded the inherited cameras from single-channel greyscale to red/green/blue (RGB) color by replacing the detector with a Bayer-pattern version of the same 1024×1024 -pixel detector. Stereo IDC image pairs, acquired by moving the arm between images, are critical for characterizing the topography of the instrument deployment workspace, a 4×6 -m area located in front of the lander. Images from the cameras are processed using software from previous Mars surface missions, with several new image products developed for InSight to support instrument placement activities.

Fig. 21.6 Camera locations on the InSight lander. The ICC can be seen on the left, mounted to the lander, and the IDC is on the right, mounted to the forearm section of the robotic arm. The distance from the IDC to the scoop at the end of the arm is approximately 0.6 m



21.2.5 IDA Motor Controller

The IDA motor controller consists of two printed-circuit boards located in the lower payload electronics box (PEB) and provides power conditioning, motor voltage control and drivers, grapple heater drivers, joint encoder counting, and analog-to-digital conversion of potentiometer voltages, temperature sensor voltages, motor currents, and heater current. The PEB provides the interface to the lander command and data handling (C&DH) computer over a serial link. Firmware running on the IDA motor controller microprocessor provides for low-level motor command execution to move the joints to the specified positions, grapple heater command execution, analog-to-digital calibration, and sensor monitoring (Trebi-Ollennu et al. 2018).

21.3 Robotics Flight Software

The IDA flight software (FSW) provides both control of the IDA hardware and visibility of IDA hardware and software state. Running on the on-board flight computer, IDA FSW communicates with the IDA motor controller. It provides the interface for IDS ground operators to control the IDA through the motor controller. Telemetry data from the IDA FSW provides the hardware and software state to the ground operators. The IDA FSW inherits from and builds upon the Phoenix robotic arm FSW (Bonitz et al. 2008). The IDA FSW provides the following specific capabilities (Trebi-Ollennu et al. 2018):

- interface with external entities, including other spacecraft FSW components and the IDA PEB
- expansion of high-level IDA commands from the command sequencer into low-level IDA actions
- motion control of the IDA
- control of the grapple
- fault sensing, recovery, and safing
- collision prevention between the IDA, lander, and science instruments
- visibility of the IDS state in telemetry.

The lander sequencing engine sends sequenced IDA commands to the IDA FSW, one at a time. The IDA FSW responds to these commands, taking action as appropriate. While handling a command, the IDA FSW may communicate with external entities, such as the IDA motor controller or power-switching FSW.

Most IDA movement commands specify a single motion of the arm or action of the IDA FSW. This could be to move the IDA to a specific position, using the most direct path, to set a parameter, or to turn a heater on or off. After executing the activity, the IDA FSW sets a flag to let the command sequence in the sequencing engine know it is ready for the next command.

Other higher-level IDA FSW commands are more complicated, building on those low-level commands. These commands execute multiple actions, potentially including multiple IDA motions, before informing the sequencing engine of the command completion. Examples of these include commands to initialize the motor controller, calibrate the motors, dig a trench, and scrape the Martian surface.

21.3.1 IDA Motion Commands and Motor Control

There are multiple types of IDA commands to specify single motions. These different categories of motion commands allow the movement to be specified in terms of joint angles, Cartesian coordinates, and time durations for motion.

Joint commands specify the IDA motion in terms of the joint angles. These commands can specify the goal position as either absolute joint angles or relative offsets from the current joint angles. Joint commands can move either single joints or all IDA joints at the same time.

Cartesian commands specify the IDA motion in terms of Cartesian coordinates and an approach angle for the current end effector, or tool. These commands can specify the goal position using absolute or relative coordinates. Absolute coordinates can be in the IDA frame, which has its origin at the base of the IDA. Relative coordinates can be in the IDA frame or in one of many tool frames attached to the IDA end-effectors. As the IDA moves, the orientation of the tool frames changes relative to the IDA frame. Some of these tool frames are rigidly fixed to the IDA links, while others hang such that their z-axis is always the direction of gravity.

Timed motion commands are relative motion commands that specify a direction and duration of time to move each joint. The speed of the joint motion is set with another command prior to the motion command.

Both joint and Cartesian move commands can be commanded in a “guarded” manner. Guarded moves allow the IDA to safely contact other objects in its workspace. Normally, if the motor controller or IDA FSW detects excessive motor currents or joint torques, the FSW will “safe” itself, which means it will stop all motion, shut itself down, abort the currently executing command sequence, and wait for ground operators to tell it what to do next. If the motion is a guarded move command, however, the motion will stop after detecting high currents or torques, but the IDA FSW will not safe itself. Instead, it will inform the sequencing engine that the motion completed successfully, thus proceeding to the next command in the sequence.

For each motion command of any type, the IDA FSW breaks it into multiple via points, where each via point is a set of joint angles or Cartesian pose between the start and goal pose. One at a time, the FSW generates a joint velocity profile for each via point and passes the via-point encoder angles and a voltage that corresponds to the desired velocity to the motor controller. The motor controller closes the loop on the encoders to move the joints to the desired angles. During each control cycle, the IDA FSW monitors the motor state and uses a software-based PID controller to

compute a new voltage to send to the motor controller. All via points in a motion are checked for limit violations and potential collisions before beginning motion to the first via point.

The IDA FSW chooses via points based on whether the motion is intended to be joint interpolated or linear interpolated (straight line). The FSW uses joint-interpolated motion for joint move commands. Ground operators can choose between joint-interpolated and linear-interpolated motion for Cartesian move commands.

21.3.2 Grapple Control

To open the grapple (Fig. 21.7), the IDA FSW turns on one of a pair of redundant grapple heaters, pushing open the grapple fingers. The heat from the grapple’s HOP actuator will damage the grapple if the actuator is left powered for too long. To avoid damage to the grapple, the IDA FSW monitors the grapple finger limit switches and the actuator temperature to determine when to power off the HOP actuator heaters. When the limit switches indicate the grapple fingers are fully open, or if the actuator becomes too hot, the FSW powers off the heaters. Additionally, the command to open the grapple has a timeout argument, and if the timeout is exceeded, the FSW will power off the grapple heater.

Ground operators use a spacecraft command to inform the IDA FSW of the state of the grapple. This grapple phase indicates whether the grapple is safe to open, stowed on the side of the IDA forearm, stowed for launch, or grappling one of the deployable payloads. If the grapple is grappling a payload, the grapple phase indicates which payload. Only spacecraft commands sent by ground operators can change the value of the grapple phase, and they are used to indicate both whether it is safe to open the grapple and which payload, if any, is grappled.

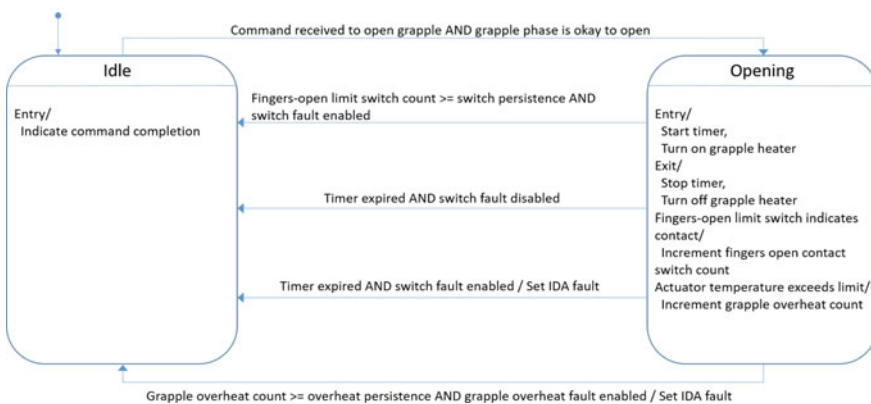


Fig. 21.7 IDA FSW grapple opening control

The grapple phase serves two purposes. First, it provides a level of protection against accidentally opening the grapple when it is not safe to do so, such as when a payload is suspended off the ground. The grapple phase must indicate that it is safe to open the grapple or the IDA FSW will reject commands to do so. Second, the grapple phase is used to indicate the grapple payload, and thus the suspended mass, when determining the deflection of the IDA due to gravity.

21.3.3 Kinematics and Deflection

For most IDA end-effectors, the IDA FSW computes the forward and inverse kinematics in the ordinary manner from the joint angles. The grapple, however, requires special treatment.

The grapple hangs on a flexible tether, suspended just prior to the fourth (wrist) joint of the IDA. The grapple and any suspended payload hang in the direction of gravity. The IDA FSW computes the forward kinematics for the grapple and suspended payloads by first computing the forward kinematics for the IDA at the grapple attachment point. Because the attachment point is prior to the fourth joint, this requires only the first three joint angles. Then it adds the length of the grapple, plus grapple cable, plus any suspended payload, times the unit gravity vector, to the grapple attachment point to determine the origin of the grapple frame or the frame for the suspended payload. The z-axis for the grapple or payload frame is set to point in the direction of gravity, and the x-axis points radially away from the base of the IDA.

To compute the inverse kinematics, the IDA FSW reverses the process. First, it determines the position of the grapple attachment point by subtracting the length of the grapple, cable, and suspended payload, times the gravity vector, from the position of the grapple or payload frame. Then, it computes the first three joint angles of the IDA using the standard inverse kinematics, ignoring the fourth joint.

Because the links of the IDA are not rigid and deflect under the weight of the IDA itself plus any suspended payload, the IDA FSW compensates for this deflection. When the FSW receives a command to move to an absolute Cartesian pose, it computes the deflection as a transformation of the end-effector frame. It uses a model of the stiffness and mass properties of each IDA link and the gravitational vector. It also uses the configuration of the grapple and mass of any suspended payload, as indicated by the grapple phase.

Once the IDA FSW has determined the deflection, it compensates by changing the commanded end-effector frame by the inverse of the deflection transformation. This compensated pose is substituted for the commanded goal pose prior to generating the via-point sequences to move the arm. Gravity deflects the IDA and payload back to the commanded goal pose.

21.3.4 *Fault Protection*

During IDA activity, the IDA FSW checks for off-nominal behavior of the hardware and software. The typical fault response is to stop all motors and heaters, including the grapple heater, announce a fault, generate a telemetry file with the recent history of the IDA hardware and software state, and then transition to a faulted state. In this state, the IDA FSW will not execute any further commands except commands to exit the fault state. Normal command sequences do not include these commands, so ground operators must send them after determining the problem.

21.4 Deployment Workspace Analysis

Figure 21.8 shows a top view of the deployment workspaces for SEIS and WTS. The coordinates are specified in the IDA frame as the reference frame to be used for surface operations on Mars. The IDA frame origin is at the IDA arm base, fixed on the lander deck with the x-axis towards workspace and the z-axis down perpendicular to the lander deck. The nominal height of the lander deck is 1.05 m, which makes the z-coordinate of the level surface 1.05 m in the IDA frame. In Fig. 21.8, the lander deck and its nearby footpads touching the surface are represented by circles of 97.8 cm and 14.5 cm radii, respectively. The white area including all inner overlapping regions is the kinematically reachable SEIS/WTS payload grapple-hook workspace, where the IDA grapple holding the payload grapple hook can reach and perform instrument placement on the level surface with nominal lander deck height. The SEIS/WTS grapple-hook workspace is bounded by (1) an outer circular boundary constrained by kinematic reachability of the arm for both SEIS and WTS placements, (2) an inner circular boundary constrained by collision prevention between WTS (larger than SEIS) and the lander structure, and (3) side boundaries constrained by collision prevention between the arm and the lander deck. The gray area including all inner overlapping regions is the SEIS footprint boundary (19.8-cm radius circle) workspace. The blue area including all inner overlapping regions is the WTS footprint boundary (50.8-cm radius circle) workspace.

The green zone is the nominal IDA grapple workspace for SEIS and WTS deployments. Its outer circular boundary is reduced from the kinematically reachable workspace by several constraints: (1) manipulability avoiding near singularity regions, (2) torque limits, (3) arm joints back-drive (IDA does not have mechanical brakes), (4) SEIS tether length, and (5) payload recapture for relocation contingency. The radius of the outer circular boundary of the green zone is 1.65 m from the arm base. Its side boundaries are confined by the yellow and pink zones. In the yellow zone, the ICC field of view is partially occluded. In the pink zone, WTS deployment at higher height over SEIS requires more maneuvering to handle collision prevention.

To minimize the effect of the noise contributions of the lander, scientists prefer to place the SEIS as far away from the lander footpads as possible. One such location is

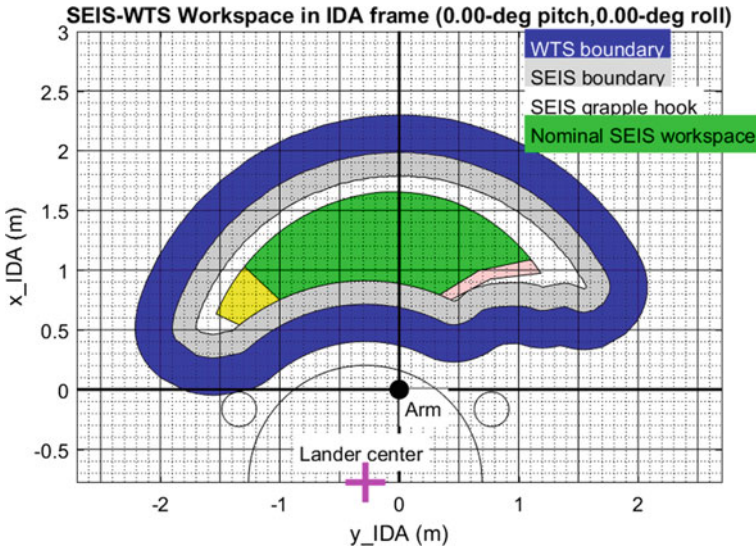


Fig. 21.8 SEIS-WTS deployment workspaces for level lander with level surface in IDA frame coordinates. The green zone is the nominal IDA grapple workspace for SEIS and WTS deployments. Its outer boundary’s radial distance from the arm base is 1.65 m. In the yellow zone, ICC view partially occluded. In the pink zone, WTS deployment at higher height over SEIS requires more maneuvering to handle collision prevention

(1.65 m, 0 m) in xy coordinates at the intersection of the outer boundary of the green zone and the x-axis projected on the surface. Another location is (1.59 m, 0.44 m) along the tether peel direction.

Figure 21.9 shows the deployment workspaces for HP3. They are very similar to those for the SEIS and WTS deployments above. The blue area including all inner overlapping regions is the HP3 footprint boundary (31.2-cm radius circle) workspace. Scientists prefer to place HP3 as far away from the lander as possible, and more than 1 m away from SEIS.

Lander tilt has a significant effect on the payload deployment workspaces and must be considered. Figure 21.10 shows SEIS-WTS workspaces on the IDA frame xy plane for four different lander tilt cases with lander (IDA frame origin) height of 1.05 m from the level surface and lander footpads. Note that the workspaces and footpads on the surface are moving together, relative to the lander deck boundary. For positive lander pitch of Fig. 21.10 (a), the maximum x-coordinate in IDA frame for the nominal deployment was reduced to 1.45 m, versus 1.65 m for the level lander. For negative lander pitch of (b), it increased to 1.8 m, but the positive-y workspace zone shrank due to an arm collision issue. For positive lander roll of (c), the positive-y workspace zone increased while the negative-y workspace zone decreased. For negative lander roll of (d), the opposite trend happened.

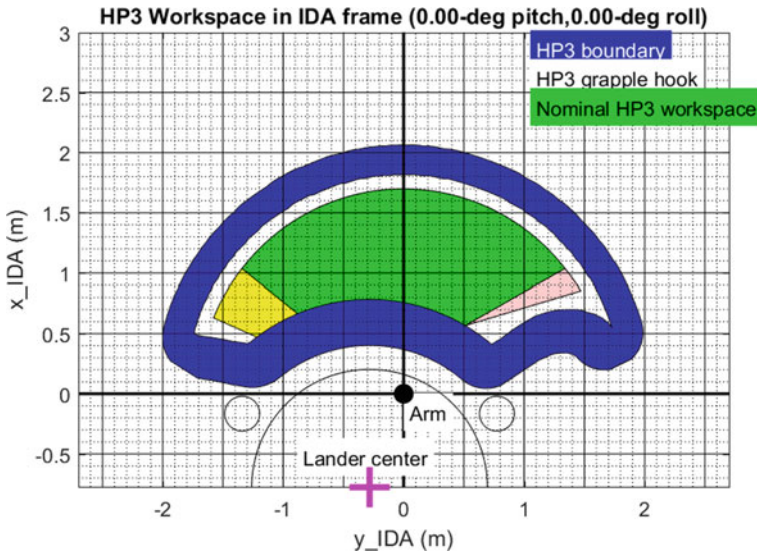


Fig. 21.9 HP3 deployment workspaces for level lander with level surface

Beyond the workspace reachability constraints, successful instrument deployments require knowledge of the 3-D workspace terrain to select the deployment site.

21.5 Workspace Imaging, Terrain Mosaic, and Site Selection

Prior to the instrument deployment, we need to know the workspace terrain in 3-D space, so we can model how the instruments interact with it. This requires a map, or digital elevation model (DEM) of the workspace, which is created by analyzing stereo images of the terrain acquired by the IDC. In order to minimize stereo baseline error, IDC stereo pairs are acquired by moving one arm joint only—the shoulder joint—while keeping all the other joints constant. The stereo overlap between left and right images is 80%, enabling generation of a workspace DEM comparable to those for the MER, PHX and MSL missions. IDC workspace imaging is done in event-driven mode with a sequence structure to capture images of the robotic arm’s workspace in several tiers, starting with an inner tier close to the base of the lander and moving progressively outward. Only the IDA azimuth joint angle is changed within a tier. To move from one tier to the next, the IDA elbow joint angle is changed. Stereo images are identified using a unique 32-bit label, called the image ID, assigned to each image, which assigns each image as “left” or “right” in the stereo pair and specifies a number which is used to pair the images (Trebi-Ollennu et al. 2012).

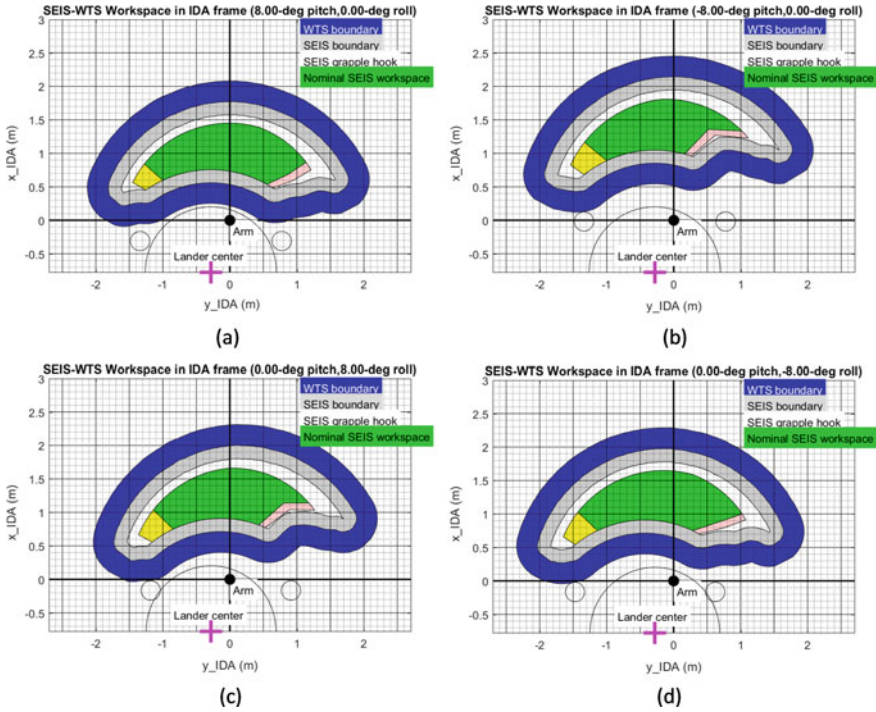


Fig. 21.10 SEIS-WTS deployment workspaces for tilted lander with level surface **a** 8-deg pitch and 0-deg roll lander **b** -8-deg pitch and 0-deg roll lander **c** 0-deg pitch and 8-deg roll lander, and **d** 0-deg pitch and -8-deg roll lander

The left eye and right eye image data products are used by a ground tool to perform stereo ranging. The range data from each pair of images is combined into a composite point cloud, which is then used to generate a 3-D model, also called a terrain mesh (see Fig. 21.11). Additional image data products are generated from the IDC stereo mosaic to support other science and operational goals (Abarca et al. 2018).

The workspace was imaged in three phases. First, an ICC context image was acquired on Sol 4. That provided an overview and was used for preliminary discussions. The primary workspace mosaic was acquired on Sol 12. This mosaic consisted of 56 images containing 26 stereo pairs in three tiers and four extra images between tiers for improved registration. This primary workspace mosaic was used for the bulk of the Instrument Site Selection Working Group (ISSWG). After preliminary deployment locations for the instruments were determined, high-resolution mosaics of just those two locations were acquired on Sol 16, consisting of 24 frames each. High resolution in this case means putting the camera closer to the terrain (1.2 m instead of 1.5 m). This could not be done earlier because of safety concerns; the workspace below the deck had to be proven to be free of obstacles (using the primary workspace mosaic) before the arm could go below the deck, as is required for the higher-resolution images.

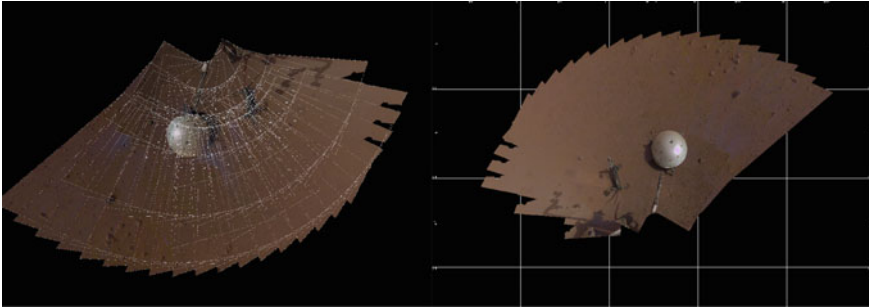


Fig. 21.11 Sol 110 post-instrument deployment vertically projected mosaic showing the location of the SEIS and HP3 instruments. Left: The white lines show the overlap of the 56 images that make up the workspace mosaic. Right: The resulting bundle-adjusted workspace mosaic

The pre-landing plan for workspace imaging and processing is fully described in (Abarca et al. 2018). Surface operations matched the pre-launch plan remarkably well, and the primary workspace mosaic was acquired and downlinked more quickly than anticipated. The actual landing site turned out to be much more benign than any of the testbed runs leading up to the mission.

As expected, the accuracy of telemetered arm positions was not sufficient to create reliable stereo products. The IDA requirement was to be able to determine stereo baseline accuracy to 2.8 mm. That amount of error in baseline knowledge translates to a theoretical ~ 2 cm of range error, which is too much to support accurate analysis of instrument placement. Therefore, the pointing knowledge of the arm must be corrected before the DEM can be computed. This is accomplished using a bundle adjustment procedure based on image tie points (Abarca et al. 2018). This process, developed before landing, worked very well, with mismatch of derived terrain in overlapping areas between frames measured at an average of 1.9 mm, up to 3.9 mm (compared to an average of 5.0 mm, up to 10.0 mm without bundle adjustment).

The high-resolution mosaics were similarly co-registered via bundle adjustment, additionally tying them to the “base map” created by the primary workspace mosaic as a control network. The process was repeated several times during instrument deployment with high-resolution mosaics co-registered to the base map on subsequent sols 35, 44, 58, 59, 62, 85, 182, 227, 230, and 240.

21.5.1 Deployment Image Products

A series of derived instrument placement products were created to show the instrument tilt, roughness of terrain under the instrument, and delta tilt between the SEIS and WTS at any point in the workspace, as well as an overall “goodness map” that combined them all, using thresholds for each instrument versus the instrument’s requirements. These products were generated on the workspace and high-resolution

mosaics. The algorithms behind them are fully described in (Abarca et al. 2018). They were utilized by both the science and engineering teams to decide on the instrument deployment locations as part of the Instrument Site Selection Working Group (ISSWG) process. It was determined after landing that the delta-tilt product was not particularly useful operationally, but all others products were used as expected based on pre-landing testing.

The instrument tilt product analyzed, for each pixel, what the minimum and maximum tilt of the instrument would be if placed at that pixel (meaning, the grapple point was directly above that position in the map). This required modeling the interaction of the feet of the instrument with the terrain, taking into account the clocking angle as well as possible sinkage of feet into the terrain. Clocking angle means rotation of the instrument about its axis. Clocking of the InSight instruments is fairly well constrained by the tether, but there is no active control over clocking angle so a range of possibilities was analyzed. Also the feet have broad pads with a spike on the bottom; the foot could then sit on either the pad or the spike depending on the firmness of the material (e.g., rock vs. loose regolith). The minimum and maximum tilt angles across all combinations of clocking and sinkage were gathered for each pixel and compared to thresholds based on the instrument requirements. A “goodness” band indicated whether the tilt met the thresholds and is shown as green (good), yellow (marginal), or red (bad) in Fig. 21.12.

Instrument tilt should not be confused with terrain slope; terrain slope is an intrinsic feature of the terrain, while instrument tilt is a measure of how the instrument interacts with that terrain.

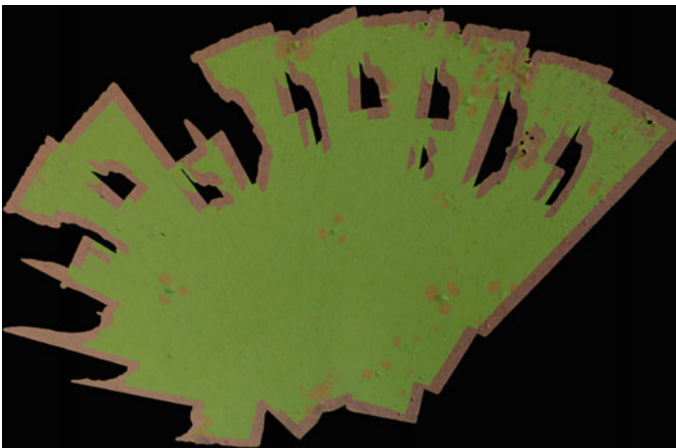


Fig. 21.12 Sol 12 orthorectified projection of the InSight Deployment Workspace with SEIS “instrument tilt” overlaid. Green pixels represent locations that pass all of the instrument placement criteria, orange represents one violated requirement, and red means at least two criteria are violated

Delta tilt looked at the difference in possible tilts between the SEIS and WTS; since the SEIS goes inside the WTS it is important to make sure they do not interact (touch).

Instrument roughness is a measure of the roughness of the terrain where it interacts with the instrument. Roughness generally is defined as the maximum peak-to-peak excursion of the terrain above or below a plane. There are two types: feet and body.

Feet roughness is a measure of roughness directly underneath the feet. A rough area for a foot means placing the foot there is unreliable; it could perch on a pebble or sink into a hole. As with tilt, the roughness was measured across a range of clock angles and sinkages.

Body roughness is more interesting; it is a measure of the terrain underneath the instrument. It measures only excursions above the plane of the feet (e.g., rocks); valleys or holes are ignored. This ensures the instrument is not “high-centered”, i.e., placed over a rock where it sits on the belly rather than the feet. That would be particularly troublesome for the seismometer, which depends on good foot contact with the ground.

The roughness product also had a goodness band that indicates whether the requirements were met at that location. Green indicates both body and feet are within thresholds, yellow indicates one is out of spec, and red indicates both. See Fig. 21.12. (Fig. 21.13).

These maps—tilt, delta tilt, roughness, as well as the workspace reachability discussed in Sect. 21.4—were then combined into a single “goodness” map, which indicated where all criteria were met (in green). See Fig. 21.14.

The goodness map was the primary screening tool used by the ISSWG to determine instrument placement (Golombek et al. 2018). It provided a quick overview of



Fig. 21.13 Sol 12 orthorectified projection of the InSight Deployment Workspace with WTS “instrument roughness” overlaid. Green pixels represent locations that pass all of the instrument placement criteria, orange indicates one requirement has been violated, and red means at least two criteria are violated

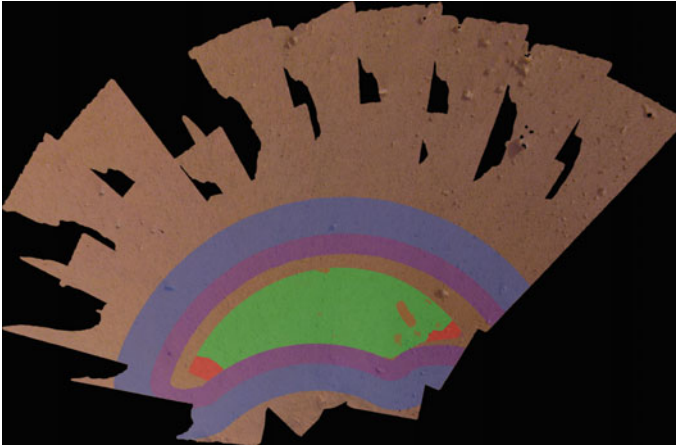


Fig. 21.14 Sol 12 orthorectified projection of the InSight Deployment Workspace with arm reachability and HP3 “goodness” overlaid. Green pixels represent locations that pass all of the instrument placement criteria, orange indicates one requirement has been violated, and red means at least two criteria are violated. Purple and blue colors indicate the extent of feet and other constraints

locations that were safe according to the analysis, with backup materials in all the other products to determine details for areas that might be marginal. As it turned out, the terrain was virtually all green across the entire workspace.

21.6 Payload Localization

On InSight, stereo imagery captured by moving the IDA was designed as the primary method for localization (Bailey et al. 2020). However, several constraints (both physical and planning related) on the system made capturing stereo data regularly difficult for day-to-day operations planning during deployment. For example, through the deployment of each instrument, the IDA was physically grappled to the instrument and thus could not capture the requisite stereo baseline. However, a measurement of the deployed location was desired as a check to know if it was safe to ungrapple. On the planning side, it was often difficult or impossible to schedule stereo imaging activities with proper lighting, or there was not enough downlink data volume available for stereo. As a result, robust monocular localization techniques were developed to allow the team to localize each instrument in the workspace with respect to the arm as well as trend their positions over time. The primary methods of monocular localization used were fiducial localization and grapple localization.

Fiducial markers were placed on each instrument and visible in the IDC field of view while the arm was grappled to the instrument; on the ground, we used this IDC image to calculate the pose of each instrument. The fiducial markers on the SEIS and WTS were “MSL-style” single-point fiducials, which provide a single measured

point on the instrument relative to the instrument origin. Unlike SEIS and WTS, the HP3 SSA visibility in the IDC field of view was much less favorable while grappled. As a result, many of the “MSL-style” fiducials on the HP3 were often not visible. To mitigate this, AprilTags were placed on the HP3. AprilTags consist of a rotationally asymmetrical pattern such that seeing a single tag in an image provided five points for position and orientation measurements. These methods of fiducial localization had an absolute error bound of about 1.5 cm and 1.5 degrees of accuracy. The sources of error on the localization were partly due to the relatively low resolution of the IDC, but had more to do with the kinematics of the IDA. For example, the computed IDC position error included propagated errors in robotic arm joint knowledge and in the robotic arm deflection model compensation. While this magnitude of error was acceptable with respect to mission requirements, the IDS team refined the measurements by calculating the bias locally to reduce risk on high-precision activities.

The method for calculating the bias for both stereo and monocular measurements was called “grapple localization.” This activity was a qualitative check of an instrument position using the arm. After each deployment and un-grappling, the IDS team placed the grapple at a known offset of 4 cm above the measured position of the instrument’s grapple hook. The team captured IDC images at this position and at 2 cm below this position. These images would give a clear qualitative indication of localization offset with respect to the arm frame location. The error in the direction perpendicular to the IDA would be very clear based on the alignment with the grapple hook and the grapple in the image. The error in the radial direction was visible in the alignment of the instrument and grapple shadows. Finally, the instrument height would be calculated using the distance from the grapple to the hook compared to the known baseline of the 2 cm move down between images. Thus, using grapple localization, the IDS team could calculate a local bias and refine future measurements. These methods became crucial for the WTS deployment in particular, where the margin for error on the deployment was very tight in order to prevent collision with the SEIS and LSA (Fig. 21.15).

21.7 Robotics Operation Tools

21.7.1 *Robot Sequencing and Visualization Program*

RSVP (Robot Sequencing and Visualization Program) is used to assist operators in planning InSight’s arm motions and instrument deployments. Prior to uplinking commands to Mars, it is necessary to simulate and visualize the predicted motions of the arm for safety and correctness.

RSVP is composed of two main components. RoSE (Robot Sequence Editor) allows operators to write command sequences. HyperDrive simulates the sequences and visualizes the resulting arm motions.

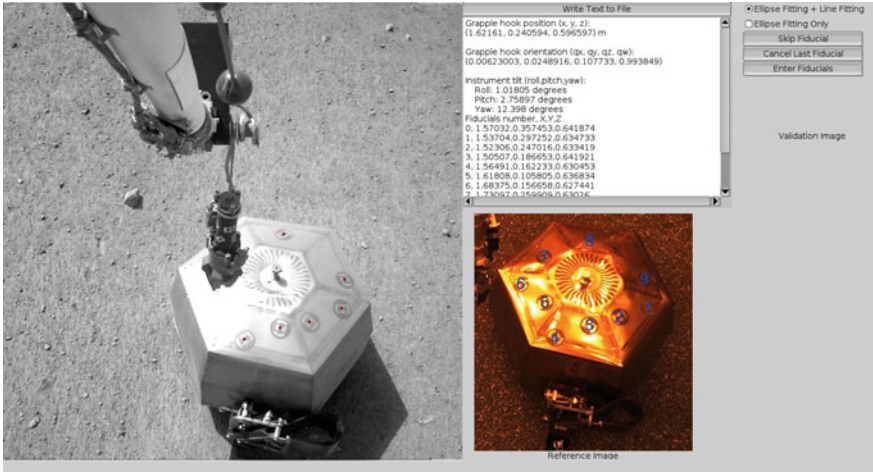


Fig. 21.15 Fiducial Localization Software with detected points remapped to the image

RSVP was first developed for driving the Sojourner rover. Newer versions of the software are used for many of the current and past Mars robotic missions. InSight’s version of RSVP shares common multi-mission code with Mars Science Laboratory’s version of RSVP. InSight’s arm simulation is derived from the arm simulation used on the Phoenix Mars mission (Yen et al. 2004) (Fig. 21.16).

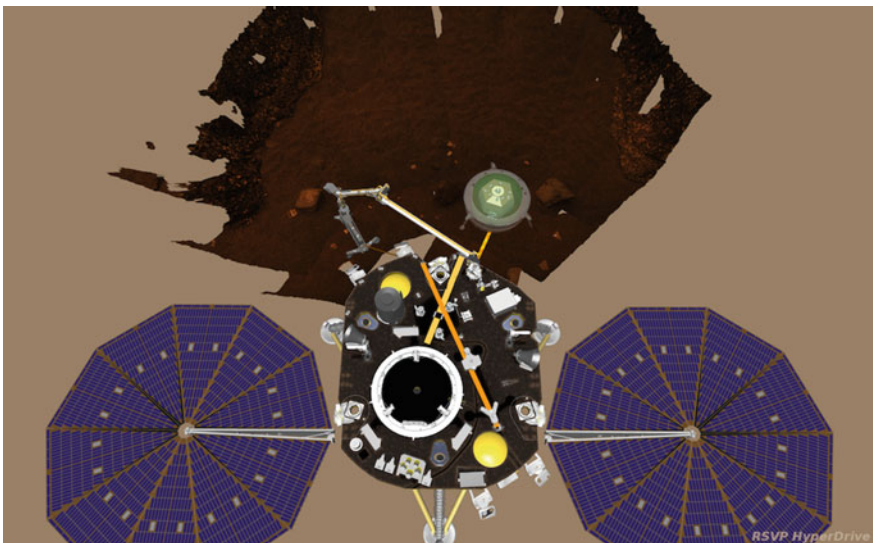


Fig. 21.16 InSight RSVP with instruments deployed

21.7.2 Simulation of Commands Using Flight Software in the Loop

RoSE is configured with knowledge of InSight’s command dictionary, allowing operators to construct syntactically correct commands. These commands are stored in the Robot Markup Language (RML) format. But safely and correctly operating the arm requires more than mere syntactic correctness. This is where HyperDrive comes into play. HyperDrive is the companion program to RoSE. The two programs send messages to each other via inter-process communication (Fig. 21.17).

HyperDrive parses the arm commands from RoSE and runs them through its embedded InSight arm flight software. HyperDrive stubs out the lower-level hardware interfaces that would normally go to actuators on the arm. Instead, HyperDrive captures the kinematic motions of the arm, which it then visualizes in 3D for the user to evaluate and iterate on. Collision faults and other faults reported by flight software are reported to the user, allowing them to resolve these issues on the ground before uplinking the command sequences to Mars (Fig. 21.18).

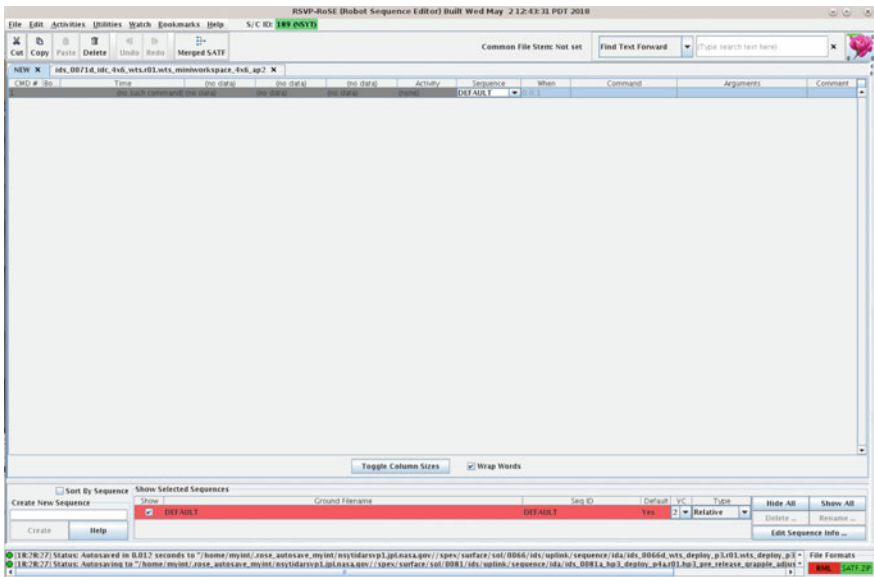


Fig. 21.17 RoSE user interface

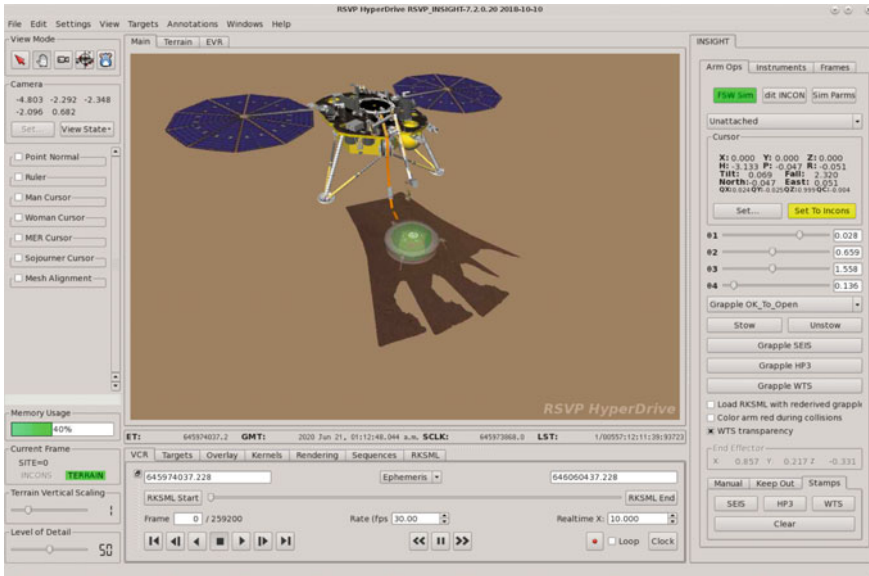


Fig. 21.18 HyperDrive user interface

21.7.3 Collision Volumes

In addition to the aforementioned collision checking during simulation of the command sequences, HyperDrive supports an interactive collision volume visualization. As the user moves the arm, HyperDrive can display the collision volumes for any given arm pose in real time. It does this by querying flight software for its collision volume information. HyperDrive also queries flight software for the collision state and displays objects that are in collision in red (Fig. 21.19).

21.7.4 Instrument Simulation

Building on top of the flight software-informed arm motions, HyperDrive simulates the motions of InSight’s grapple, instruments, and tethers. Based on the grapple state in flight software, HyperDrive derives the position and the open/closed state of the grapple.

HyperDrive uses a catenary model for its simulation of the tethers connecting the SEIS and the HP3 to the lander. This gives the operators a rough estimate of where to expect the tether during each instrument deployment. To simulate the instrument/terrain interaction, HyperDrive fits the instrument plane to the local terrain patch (Fig. 21.20).

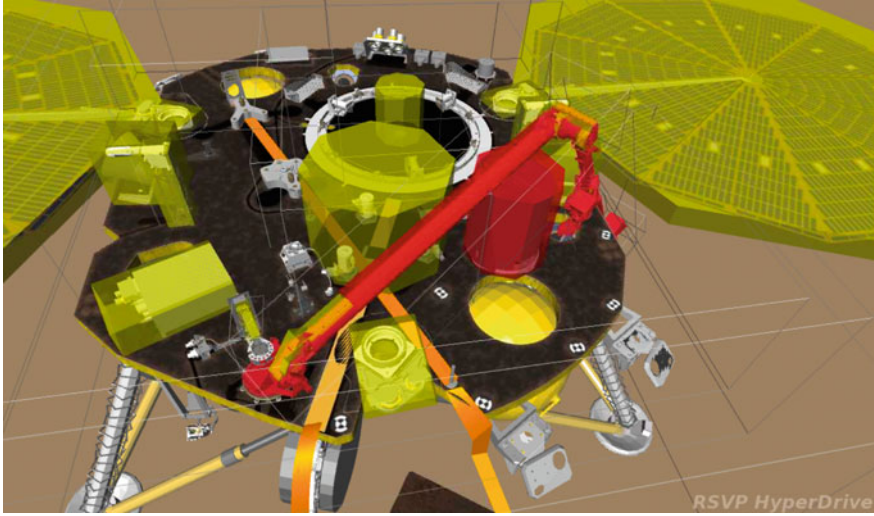


Fig. 21.19 Collision volumes in HyperDrive

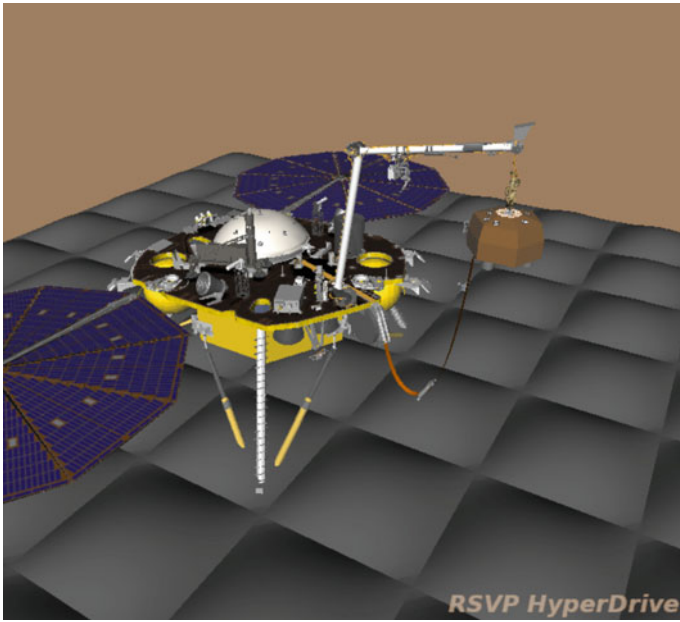


Fig. 21.20 Grapple in HyperDrive

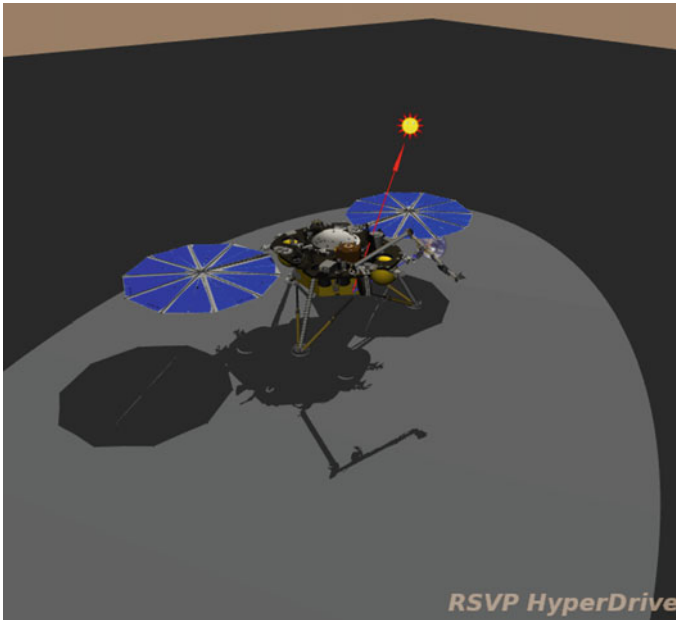


Fig. 21.21 Shadows in HyperDrive

21.7.5 Shadow Modeling

Using the InSight mission’s SPICE (Spacecraft, Planet, Instrument, C-matrix, Events) kernels, HyperDrive supports visualizing the location of the Sun with respect to the lander. It renders the lander’s shadow on the terrain to assist operators in constructing imaging commands (Fig. 21.21).

21.7.6 SEIS Tether Catenary Modeling

To successfully deploy the seismometer on the Martian surface, the team needed to consider the dynamics of the entire instrument system, which included its tether (Sorice et al. 2020). During SEIS deployment, Part 3 (as is described in Sect. 21.8), the robotic arm placed SEIS on the surface of Mars. In this deployment, key components of the tether were considered: the tether storage box (TSB), the field joint (FJ), the pinning mass (PM), and the load shunt assembly (LSA). Once the tether is peeled, the tether hangs from two points: one point at the exit of the TSB mounted on the lander, and the other point at the LSA loop where it attaches to the instrument. It is at this point that the tether takes on the catenary shape, and our quasi-static model

is interpreted to help us predict the position of the pinning mass and the field joint along the tether as we move the SEIS away from the lander and over the terrain.

When deploying SEIS in the deployment testbed and with the flight hardware, the team determined that the configuration of the tether had a significant impact on the SEIS instrument placement accuracy. After observing these effects, the team decided to develop a model to predict which part of the tether would touch the terrain first, given knowledge of the height of the terrain from the lander to the desired deployment target. By modeling the tether height throughout the deployment trajectory, we could control the deployment of the tether itself and utilize the tether–terrain interactions to our advantage. We chose to use a catenary equation rather than an arbitrary polynomial equation for the ease of computation, verification, and implementation in our simulation and visualization software. Additionally, the tether’s stiffness and bending behavior made it unlikely that we could use a single polynomial equation across the entire deployment workspace.

In deploying our payloads, we commanded the robotic arm to a series of waypoints, each followed by a set of images with the IDC and ICC. During the deployment sequence, the arm (and the payload) remain still during the duration of the image acquisitions (roughly 1.5–2 min) before moving on to the next commanded position. By choosing to model the tether and solve the equation at these specific configurations along the trajectory of SEIS deployment, we took a quasi-static approach and ignored inertial effects. This simplified the problem while still remaining relevant for understanding the movement and position of the tether during deployment (Fig. 21.22).

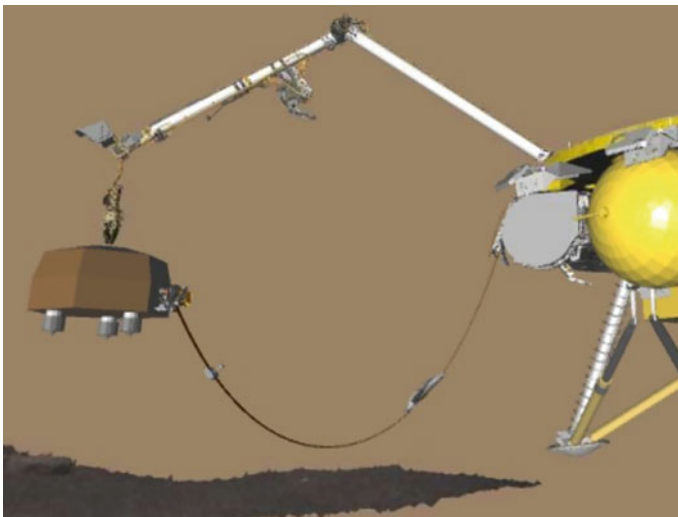


Fig. 21.22 RSVP simulation with suspended SEIS tether over terrain mesh

21.8 Surface Operations Results

21.8.1 SEIS Deployment

Each payload deployment (lift from the lander deck to placement on surface) consists of four parts (Part 1, Part 2, Part 3, and Part 4). After completion of each part there is a ground-in-the-loop GO/NO GO decision to execute the following part. In addition, there can be no intervening motion of the IDA between the four parts. On Sol 18 SEIS deployment Part 1 was successfully executed on Mars. It entailed moving the unstowed grapple to the SEIS teach point, leaving the grapple with about 4 cm between the bottom of the grapple fingers and the top of the SEIS grapple hook (see Fig. 21.23). Four IDC images were acquired at this IDA pose to document and verify successful alignment needed for the GO/NO GO decision for SEIS deployment Part 2.

SEIS deployment Part 2 (SEIS capture) was successfully executed on Sol 20. The SEIS deployment Part 2 sequence opened the grapple fingers, moved the IDA grapple end effector 4 cm toward the SEIS grapple hook, and then waited for the grapple fingers to close around the SEIS grapple hook. Grapple telemetry and IDC images were used to confirm successful SEIS capture (see Fig. 21.23).

On Sol 22 SEIS deployment Part 3 was successfully executed on Mars. First the SEIS launch restraint frangibolts were actuated successfully and SEIS was lifted 30 cm above the SEIS launch restraint pedal stools. Then, the SEIS tether, which was held to the lander deck and TSB by hook-and-loop fasteners, was peeled off. This was followed by the IDA extracting additional SEIS tether from the TSB up to the SEIS tether chock. The IDA then successfully placed the SEIS on the surface of Mars.

SEIS deployment Part 4 (SEIS release) was successfully executed on Sols 24 and 25. On Sol 24 the grapple cable slack was reduced by moving the IDA grapple end effector 1 cm up from the SEIS grapple hook. On Sol 25 the grapple was opened, and moved 10 cm up from the SEIS grapple hook to release the SEIS, marking the first ever successful precision robotics instrument placement and release on another astronomical body.

From Sols 26 to 36 the SEIS science operations team successfully performed SEIS health checkouts. On Sol 37 the remaining SEIS tether in the TSB was released by opening the TSB door. This was followed by IDC documentation of the SEIS tether configuration on the surface of Mars. On Sol 40 the SEIS load shunt assembly (LSA) frangibolt was actuated successfully but the LSA failed to meet the minimum separation gap that is intended to dampen the effects of tether thermoelastic noise on the seismic measurements. A successful minimum separation of the LSA is required for SEIS to meet its performance requirements.

From Sols 41 to 50 we acquired additional IDC images of the SEIS tether pinning mass (PM), LSA, and SEIS tether configuration on the surface on Mars in preparation for using the scoop to separate the LSA to meet the minimum separation gap required for SEIS to meet its performance requirements. On Sols 56 to 59 we successfully used the IDA scoop to move the PM via the grapple hook to assure successful LSA

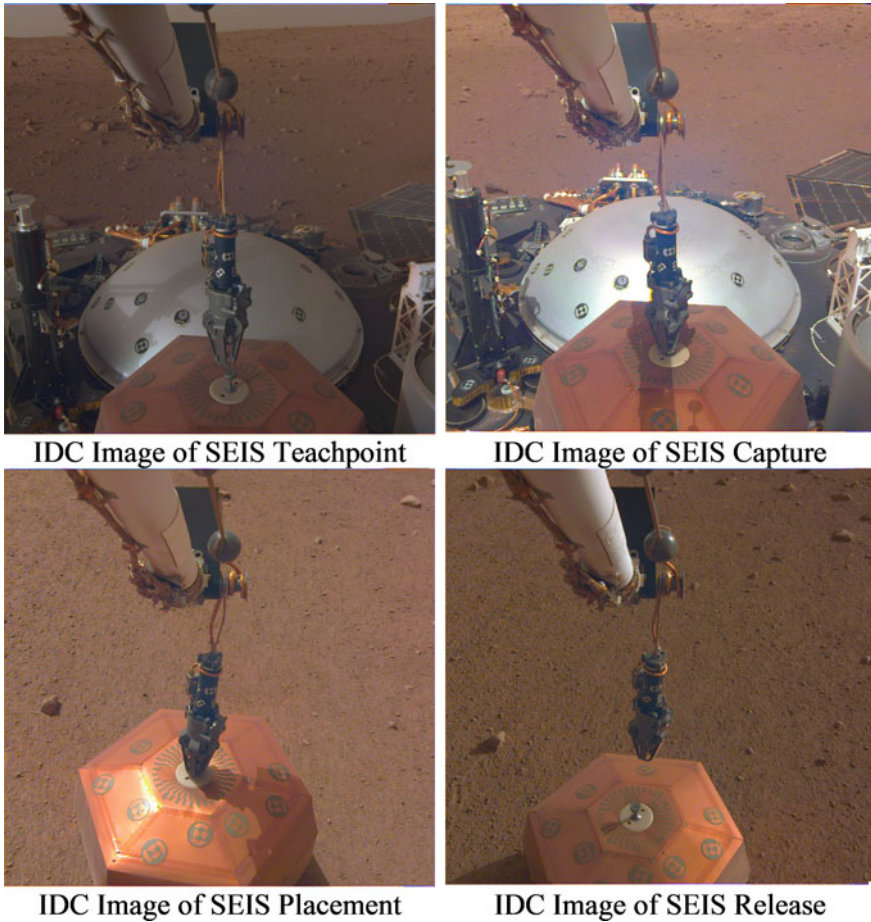


Fig. 21.23 IDC images showing successfully SEIS Deployment Parts 1–4 on the surface of Mars

minimum separation (see Fig. 21.27). From Sols 60 to 62 the SEIS science operations team successfully performed SEIS health checkouts in preparation for placing WTS over the SEIS.

21.8.2 WTS Deployment

The wind and thermal shield (WTS) is a cover that goes over the SEIS (Fig. 21.24). It has a rounded hollow aluminum cap supported by three feet. The feet were folded up beneath the cap when the WTS was on the lander deck (Fig. 21.25), and springs forced

them to unfold when the WTS was lifted from the deck. There is a compliant bellows-like aluminized-Kapton skirt, weighted down by steel scale mail and chainmail, which descends from the hard cap. The mail pulls the skirt down and makes it conform to the undulations of the Martian terrain and SEIS tether. The purpose of the WTS is to reduce noise in the seismic data collected by the SEIS. It does this by shielding the SEIS from the Martian wind and by reducing the magnitude of temperature swings throughout the Martian day.

The deployment of the WTS required accurate and precise placement. The WTS dome is not much larger than the combined SEIS remote warm electronics box (RWEB) (Fig. 21.26), which is the outer body of the SEIS, and the SEIS load shunt array (Fig. 21.27), where the tether attaches. To preserve the integrity of the SEIS data, the WTS must not contact the SEIS when placed over it. Therefore, there is not much margin for error in WTS placement. Error in WTS placement is a combination of IDA positioning error, error in SEIS location knowledge, error in surface slope knowledge, and error introduced by WTS movement by wind.

A WTS do-not-exceed (DNE) envelope was defined to account for up to 4 cm of placement uncertainty. If the chosen WTS placement location does not cause the SEIS to penetrate the WTS DNE, then the WTS should be guaranteed not to contact the SEIS. Potential differences in instrument tilt, due to undulations in the terrain, can shrink the range of placement locations that do not cause a WTS DNE violation.

The choice of WTS placement location was based on knowledge of the SEIS location and knowledge of the Martian surface slope in the area surrounding the SEIS. A relative offset, along the SEIS tether in the direction of the lander, was added to accommodate the opening of the load shunt array (LSA) (Fig. 21.27). Based on the maximum expected LSA opening distance and a maximum expected WTS placement error of 3 cm, a WTS/SEIS offset of 5 cm was chosen. This provides



Fig. 21.24 The wind and thermal shield (WTS) on the Martian surface on Sol 394. The SEIS is underneath the WTS, and the SEIS tether can be seen protruding from underneath the WTS skirt

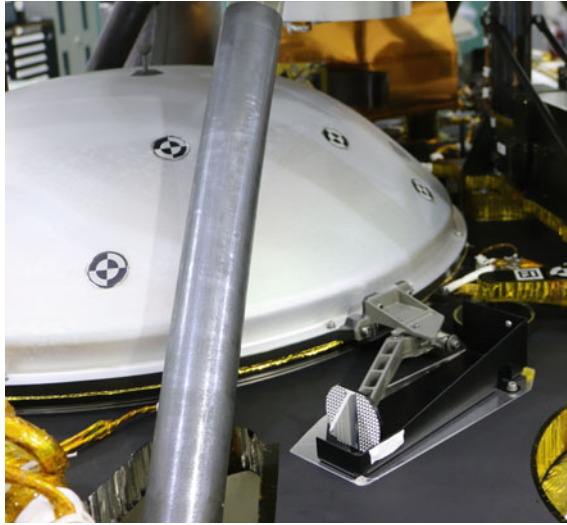


Fig. 21.25 The WTS on the lander deck before launch. The feet were folded to the side when the WTS was on the deck. They were pushed open by springs when the WTS was lifted

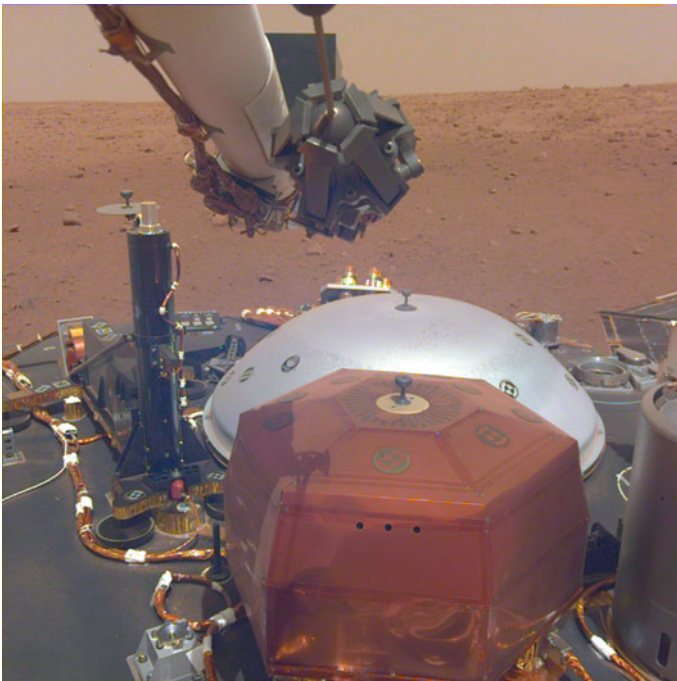


Fig. 21.26 The SEIS (foreground), WTS (behind the SEIS), and HP3 (to the left of the WTS) on the lander deck. The WTS is in its folded-leg configuration. The relative sizes of the SEIS and WTS can be seen

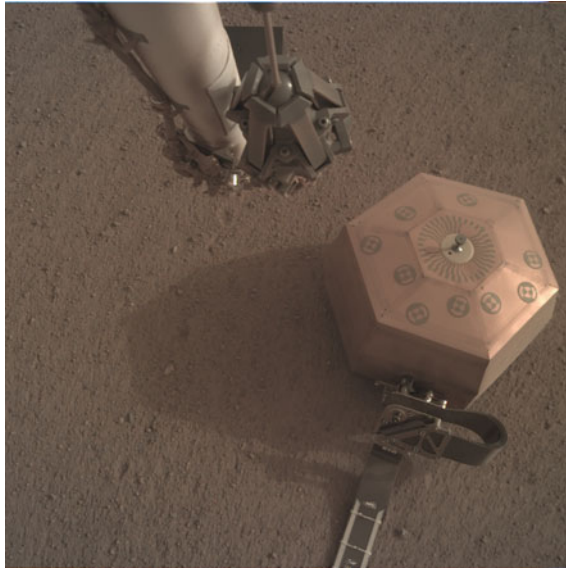


Fig. 21.27 The SEIS on the Martian surface with the LSA opened. Both the body of the SEIS and the LSA must fit beneath the WTS without contacting it

a margin of 3.5 cm of space between the far side of the SEIS and the WTS and a margin of 0.5 cm between the LSA and the WTS. The WTS/SEIS offset was chosen to allow more margin between the RWEB and the WTS than between the LSA and the WTS, because contact with the RWEB would cause a greater degradation to the quality of the scientific SEIS data.

The surface slope can cause an offset of the WTS as it is set down, where one foot contacts the ground before the others and then the WTS pivots around this foot. There was no control over the clocking, or rotation, of the WTS during deployment. The grapple and WTS grapple hook are designed to minimize clocking, but the compliant grapple cable still allows some clocking. Therefore, there was no control over the exact positions of the WTS feet on the surface. Because the positions of the feet on the Martian surface determine the effective surface slope in the WTS deployment location, this uncertainty in clocking added to the uncertainty in deployed WTS tilt, which in turn, affects the relative SEIS/WTS separation.

The temperature and wind sensors (TWINS) instrument, which is part of the auxiliary payload sensor system (APSS) (Banfield et al. 2020), provides wind speed data in the environment around the lander. This data was used to help choose a time of day for WTS deployment that was not excessively windy. Because the grapple cable is compliant, excessive winds could cause undesired movement of the WTS during deployment, which would affect the accuracy of the placement.

The IDA actuators were not designed to hold heavy loads such as the WTS, which has a mass of 9.5 kg, in an outstretched pose while not powered. In between each individual motion of the IDA, the motors are powered off. This means that the motors may back-drive while under a heavy load such as the WTS. Back-drive during WTS deployment contributes to WTS placement error. To minimize the amount of IDA back-drive, the WTS was deployed when the motors were cold, and therefore there was more friction in the joints. The IDA joints needed to be less than or equal to -20° Celsius to prevent back-drive, with colder temperatures preferred. However, the IDA has minimum temperature constraints during use to prevent damage to the motors. Additionally, there must be enough sunlight to allow useful images to be taken with the cameras during deployment. These restrictions created two time windows in which the WTS could be deployed. One was in the morning and one in the late afternoon. After consideration of the wind speeds and the timing of the available communications windows, the morning deployment window was chosen.

Like the SEIS and HP3 deployments, the WTS deployment was divided into four main parts, with an additional grapple positioning adjustment that had to be performed prior to executing the fourth part. Each of these parts, plus the grapple positioning adjustment, required ground-in-the-loop confirmation of success before proceeding to the next part. Therefore, each part had to be executed in a planning cycle of its own. There could be no intervening motion of the IDA between these deployment parts.

WTS deployment Part 1 was executed on Sol 63 of the mission. On this sol, before starting the WTS deployment, we commanded a move to put the grapple above the SEIS grapple hook to confirm our localization of the SEIS. To start the deployment, we moved the IDA to our “deck-ready-in” pose, a neutral pose from which the IDA can easily move the grapple to above any of the three instrument grapple hooks. Then we moved the grapple to a position such that the grapple frame was 5 cm above the WTS grapple hook frame. This leaves about 4 cm between the bottom of the grapple and the top of the grapple hook. We took an IDC image here and disabled the collision checking in the IDA FSW between the grapple fingers and the WTS grapple hook. Then we moved the grapple down 2 cm closer to the grapple hook and took another IDC image. We moved up 2 cm and then down 2 cm again, taking images (see Fig. 21.28) at each of the poses. We used the IDC images and the IDA joint angles to confirm that the grapple was in the correct location for the capture in the next deployment part. The images at different heights above the grapple hook assist in determining whether the grapple is sufficiently aligned over the grapple hook, because we do not have stereo vision and cannot move side to side for this determination. On Sol 64, in preparation for deployment Part 3, we monitored the IDA joint temperatures at the chosen deployment time, to verify that the temperatures were still cold enough.

WTS deployment Part 2 was executed on Sol 65. We opened the grapple, moved the grapple down 4 cm, and then let it close around the WTS grapple hook. We used IDC images, grapple telemetry showing correct grapple operation, and IDA joint angles to confirm that the IDA was in the correct pose and had successfully captured the WTS grapple hook.

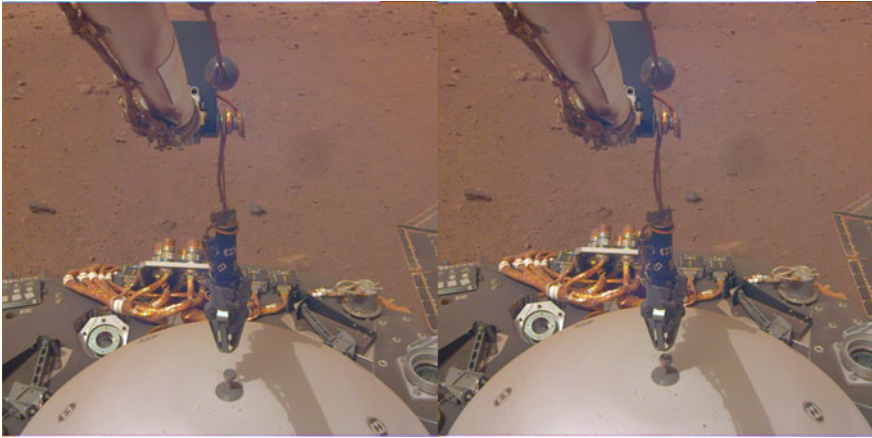


Fig. 21.28 The grapple above the WTS grapple hook during WTS deployment Part 1. In the left image, the bottom of the grapple fingers are approximately 4 cm above the top of the WTS grapple hook. In the right image, they are 2 cm above the top of the grapple hook, and this is the final position after the completion of WTS deployment Part 1

WTS deployment Part 3 was executed on Sol 66. In this part, we moved the WTS from the deck to the Martian surface. First, we actuated frangibolts which held the WTS to the deck, freeing the WTS. Then, we lifted the WTS 18 cm in the direction opposite the pull of gravity. This lifted it to its standup position, such that the legs extended to their upright position supporting the weight of the WTS. Then, we lifted it another 43 cm in the anti-gravity direction, such that it was high enough for us to swing it over the rest of the deck without hitting anything. We then moved it a few centimeters outward (straight ahead in Fig. 21.29), so that it would not fully eclipse the UHF antenna when we moved it around to the IDA workspace. By rotating only the IDA azimuth joint, we moved the WTS to the front of the lander over the workspace. In a series of stair-step motions, we moved the WTS back in towards the base of the arm and downward towards the surface. Once the WTS was lower to the surface, but not so low that it would strike the SEIS, we moved the WTS to the high standoff position above the SEIS, directly over the chosen WTS deployment location. After lowering the WTS in the anti-gravity direction to the low standoff position, just above the height of the SEIS, we disabled the IDA FSW collision checking between the WTS and SEIS. Finally, we lowered the WTS to the Martian surface in multiple steps, taking IDC images along the way. Because there was some uncertainty in the height of the ground, we overdrove the final movement, meaning that we moved the grapple lower than necessary to place the WTS on the surface. This overdrive guaranteed that the WTS would not be partially suspended at the end of the deployment, even with uncertainty in the terrain height. We used IDC and ICC images, and IDA joint angles, to determine that the WTS was in the correct location and fully supported by the surface. We looked for slack in the grapple cable in the images to determine that the WTS was not still partially suspended (Fig. 21.30).

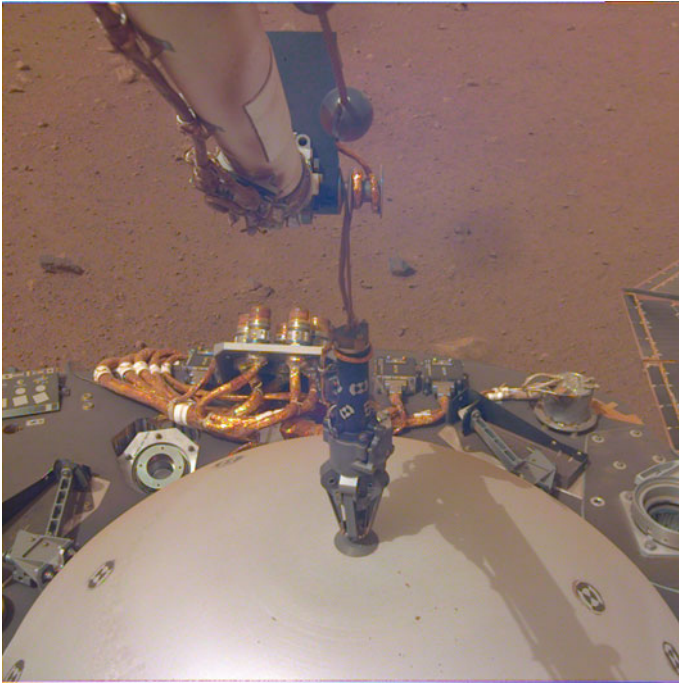


Fig. 21.29 The grapple closed around the WTS grapple hook at the completion of WTS deployment Part 2

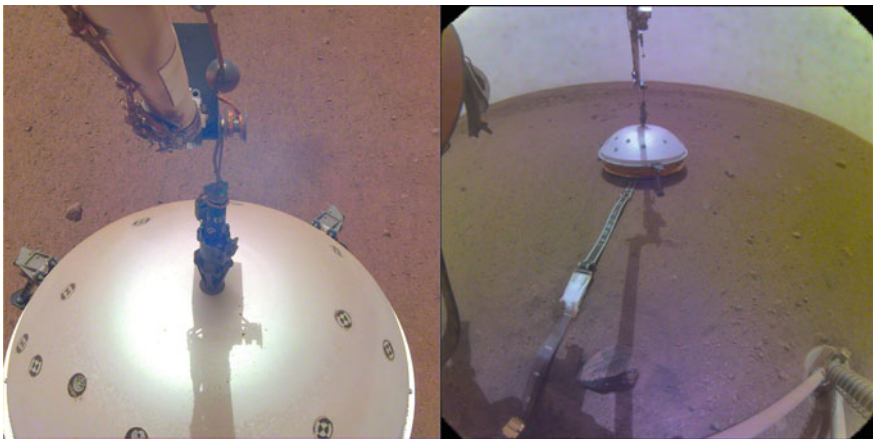


Fig. 21.30 The WTS and grapple after completion of WTS deployment Part 4a. The grapple has been moved up and over a little to reduce the slack in the grapple cable and center it over the WTS grapple hook

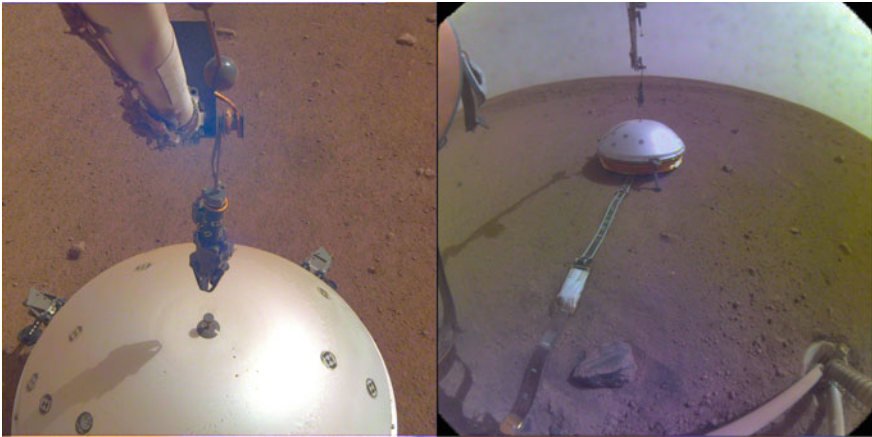


Fig. 21.31 The WTS and grapple after completion of WTS deployment Part 4. The grapple has released the WTS grapple hook and the IDA has moved upwards

WTS deployment Part 4a, or the grapple positioning adjustment, was executed on Sol 67. This part included moving the IDA slightly in order to pull out slack in the grapple cable and align the grapple over the WTS. This prepares the grapple for opening during WTS deployment Part 4. Before moving the IDA, we set the collision parameters in the IDA FSW to position the WTS collision model at our current best estimate for the WTS position. We moved the grapple up and over to center it over the grapple hook. While we wanted to lift the grapple up to remove the slack in the cable, we biased the motion to the lower portion of the grapple hook to ensure there was no upward force on the top of the grapple hook. To confirm the correct positioning, we looked at both IDC and ICC images (see Fig. 21.30), plus IDA joint angles.

WTS deployment Part 4 was executed on Sol 70. During this final part of WTS deployment, we moved the grapple up one more cm to reduce the slack in the grapple cable more, opened the grapple, and then moved the grapple up 10 cm to release the grapple hook. Then we restored the collision parameters in the IDA FSW so that it would again consider collisions between the grapple and the WTS. We used IDC and ICC images (see Fig. 21.31), grapple telemetry showing correct grapple operation, and IDA joint angles to confirm that the IDA was in the correct pose and had successfully released the WTS grapple hook.

The subsequent localization of the WTS indicated a ground position offset by 6.5 cm in the desired direction from the SEIS center point. Because we were trying to achieve a 5 cm offset, the deployed position was within 1.5 cm of the desired placement location, which is very good accuracy for this first-of-its-kind robotic stacking deployment on another planet. The SEIS science data showed an immediate improvement after the WTS was placed over the SEIS in WTS deployment Part 3.

Based on the SEIS data, the SEIS team was able to confirm that there is no contact between the WTS and SEIS.

21.8.3 HP3 Deployment

The Heat Flow and Physical Properties Package, or HP3, consists of a support structure assembly (SSA), a “mole” intended to penetrate into the Martian regolith, and a fixed radiometer. The mole is attached to the SSA via a science tether, and the SSA is attached to the lander via an engineering tether. The HP3 experiment is attempting to understand Mars’ subsurface heat flow and physical properties (Spohn et al. 2018). During the deployment phase, the HP3 was deployed last, on Sol 76. As described in Sect. 21.5, the HP3 placement location was selected earlier in the mission by the Instrument Site Selection Working Group. The final decision was based on the necessary engineering and science criteria as well as a desire to be far away from SEIS/WTS.

The surface operations team determined the deployment time of day by evaluating constraints such as robotic arm motor temperature requirements, engineering tether temperature requirements, and sunlight in the IDC field of view. Ultimately, 10.45 a.m. local mean solar time (LMST) proved to satisfy all of the constraints.

Like SEIS and WTS deployments before it, HP3 deployment and release was divided across five operational sols and five parts. Part 1, which consisted of moving the IDA’s grapple above the HP3 grapple hook, known as a teach point, took place on Sol 73. Part 2, or capturing the HP3, took place on Sol 74. Part 3, or lifting the HP3 from the lander deck and placing it down on the Martian surface, took place on Sol 76. Part 4a, or adjusting the grapple above the HP3 to be vertical, took place on Sol 79. Part 4, or releasing the grapple from the HP3, took place on Sol 83.

The HP3 engineering tether that connects the SSA to the lander was stored inside the SSA while it was bolted to the lander deck. During deployment, the tether unfurled and was pulled out of the SSA as the SSA was pulled further and further from its position on the deck. The robotic arm lifted the SSA and moved it across the lander deck and over the workspace (Fig. 21.32). At that point, we commanded the arm to an outstretched position over the lander deck which allowed us to further extract the engineering tether before bringing the SSA towards its designated placement site.

On Sol 87, the mole was released from the SSA via a frangibolt firing. On Sol 92, the first mole hammering cycle was commanded. The mole did not reach its target depth of 70 cm during this hammering cycle. The subsequent hammering tests were unsuccessful; the mole remained partially above ground and inside the SSA (Fig. 21.33). These unexpected hammering failures resulted in the creation of an Anomaly Resolution Team to resolve this issue. Mole recovery efforts are still ongoing.

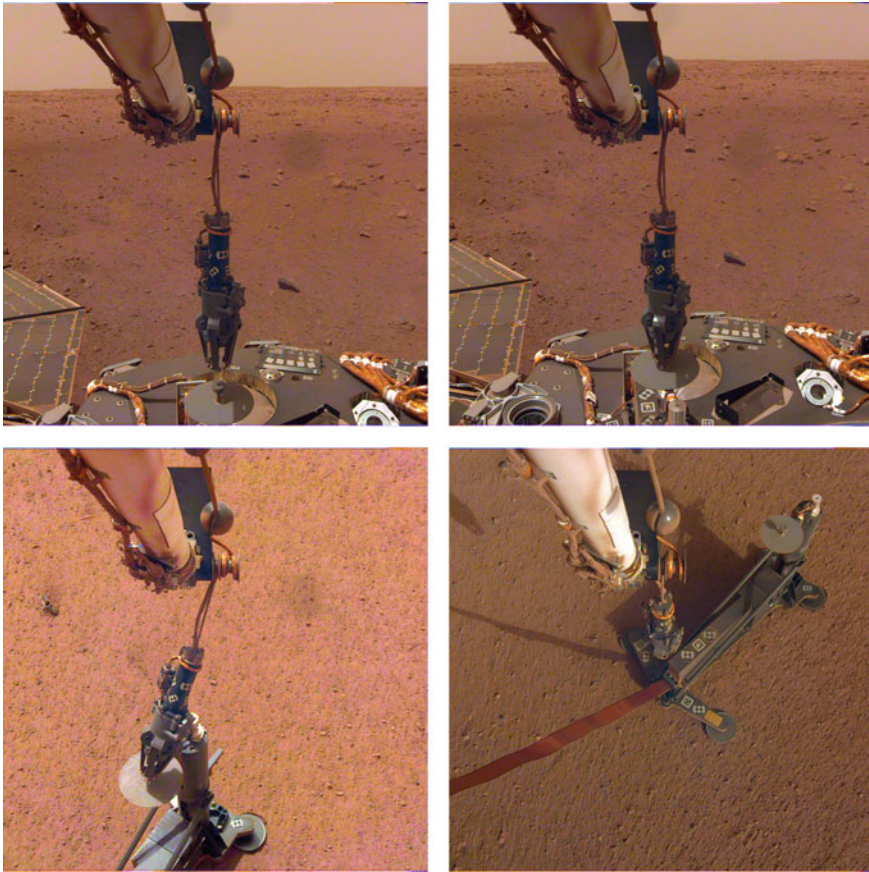


Fig. 21.32 (upper left) Grapple at HP3 Teach Point; (upper right) HP3 grappled; (bottom left) HP3 touchdown on Martian surface; (bottom right) HP3 SSA and engineering tether visible after grapple release

21.9 Summary and Conclusions

This chapter presented flight operations results for the InSight Robotics Instrument Deployment Systems (IDS) that successfully deployed SEIS, WTS and HP3 on the surface of Mars and enabled scientists to perform the first comprehensive surface-based geophysical investigation of Mars' interior structure. NASA's first successful precision robotics instrument placement and release on another astronomical body since Apollo has paved the way for future human precursor robotics planetary construction missions. In addition, the success of the IDS has paved the way for the development of autonomous manipulation as a key enabling technology for successful in-situ payload installation and geophysical science investigations in planetary exploration missions.

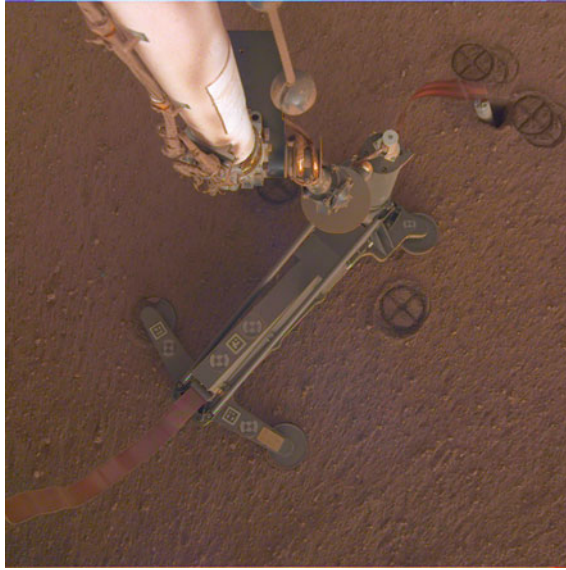


Fig. 21.33 IDC view on Sol 209, after the HP3 SSA was lifted and re-placed down behind the mole by the Anomaly Resolution Team to investigate the configuration of the mole in the Martian surface

Acknowledgements The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). © 2020. California Institute of Technology. Government sponsorship acknowledged.

References

- Abarca, H., et al. 2018. Image and data processing for InSight lander operations and science. *Space Science Reviews* 215: 22 (2019).
- Bailey, P., et al. 2020. Deployed instrument monocular localization on the InSight Mars lander. In *IEEE aerospace conference. Proceedings* (Cat. No.00TH8484), vol. 7, pp. 235–246 <https://doi.org/10.1109/AERO47225.2020.9172343>
- Banerdt, W.B., et al. 2018. The InSight mission, space science reviews. *Space Science Reviews* 215: 22 (2018).
- Banfield, D., A. Spiga, C. Newman, et al. 2020. The atmosphere of Mars as observed by InSight. *Nature Geoscience* 13: 190–198. <https://doi.org/10.1038/s41561-020-0534-0>.
- Bonitz, R.G., T.T. Nguyen, and W. S. Kim. 2000. The Mars surveyor '01 rover and robotic arm. *IEEE aerospace conference. Proceedings* (Cat. No.00TH8484), vol. 7, pp. 235–246
- Bonitz, R.G., et al. 2008. NASA Mars 2007 Phoenix lander robotic arm and icy soil acquisition device. *Journal of Geophysical Research* 113: E00A01. <https://doi.org/10.1029/2007JE003030>.
- Folkner, W., et al. 2018. The rotation and interior structure experiment on the InSight mission to Mars. *Space Science Reviews*. <https://doi.org/10.1007/s11214-018-0530-5>. (2018 this issue).
- Golombek, M., et al. 2018. Geology and physical properties investigations by the InSight lander. *Space Science Reviews* 214: 84.

- Lognonné, P., et al. 2019. SEIS: Insight's seismic experiment for internal structure of Mars. *Space Science Reviews* 215: 12.
- Maki, J.N., et al. 2018. The color cameras on the InSight lander. *Space Science Reviews* 214: 105.
- Sorice, C., P. Bailey, A. Trebi-Ollennu, W.S. Kim, and S. Myint. 2020. Catenary model of InSight SEIS tether for instrument deployment. In *2020 IEEE aerospace conference*, Big Sky, MT, USA, pp. 1–8. <https://doi.org/10.1109/AERO47225.2020.9172783>.
- Spohn, T., M. Grott, et al. 2018. The heat flow and physical properties package (HP3) for the InSight mission. *Space Science Reviews*, this issue.
- Trebi-Ollennu, A., et al. 2012. Lunar surface operation testbed (LSOT). In *2012 IEEE aerospace conference*, Big Sky, MT, pp. 1– 16. <https://doi.org/10.1109/AERO.2012.6187057>.
- Trebi-Ollennu, A., et al. 2018. InSight Mars lander robotics instrument deployment system. *Space Science Reviews* 214: 93.
- Yen, J., B. Cooper, F. Hartman, S. Maxwell, and J. Wright. 2004. Sequence rehearsal and validation on surface operations of the Mars exploration rovers. In *Proceedings of SpaceOps 2004*, Montreal, Canada.

Part V
Asteroids and Comets

Chapter 22

Asteroid Habitats—Living Inside a Hollow Celestial Body



Werner Grandl and Clemens Böck

Abstract The chapter discusses the possible use of near-Earth asteroids (NEAs) for mining, the space industry and human habitation. The building of self-sustaining habitats in space will be a crucial challenge in coming centuries. In the long run potential hazardous asteroids (PHAs) are a real threat to life on Earth; however, with the development of advanced propulsion systems like the Bussard fusion engine they can be intercepted and their trajectories modified. Various methods of deflection are discussed. Some NEAs or PHAs could be forced into an Earth orbit beyond the Moon and used for mining rare metals and rare Earth elements. When the mining process is complete the remaining hull of the asteroid is a suitable shield against cosmic rays and meteorites. The authors propose construction of rotating toroidal habitats inside these hollow celestial bodies. Natural sunlight is collected by an array of parabolic mirrors, beamed into the cave and distributed by a central cone with facet mirrors. By adjusting shutter elements a 24-hour cycle of day and night can be simulated. Oxygen and hydrogen are extracted from asteroid material during the mining process or in industrial plants in space, e.g., in the Lagrange points of the Earth–Moon system. Vertical farms and aquaculture could provide nutrition for around 2000 inhabitants of an asteroid colony. In the coming centuries some hundreds of thousands of people could settle in the Earth–Moon system and constitute a new society beyond planet Earth.

22.1 Introduction

The utilization of near-Earth asteroids (NEAs) for mining, and building human habitats inside some of these celestial bodies, will be a crucial challenge for coming centuries. NEAs, which are expected to contain resources like nickel–iron, platinum

W. Grandl (✉)

Consulting architect and civil engineer, Dr. Billroth-Strasse 6, 3430 Tulln, Austria

e-mail: archigran@gmx.at

C. Böck

Mechanical engineer, Alban Berg Gasse 2/1, 3430 Tulln, Austria

© Springer Nature Switzerland AG 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,

https://doi.org/10.1007/978-3-030-97913-3_22

763

group metals or rare-earth elements and have a typical bulk density of 2.5 g/cm^3 and a diameter of 100–500 m, will be selected for mining. It will be useful to modify the orbit of a target NEA from a heliocentric orbit into a geocentric orbit beyond the Moon (Grandl and Bazso 2013). Since 2016 we have known that Earth has a second temporary companion, the “quasi-satellite” 2016HO3. This small asteroid, with a maximum diameter of approx. 100 m, is orbiting Earth at an average distance of 60 times the distance of the Moon. 2016HO3 may be the first target to test asteroid mining technology. To change the orbit of NEAs it is necessary to develop advanced propulsion systems, e.g., Deuterium-Helium-3 fusion engines, to move the huge masses of these asteroids.

For the mining process an Earth orbit beyond the Moon should enable us to keep the rate of mining advance equal to the rate of cargo shipping between the NEA and the Earth–Moon system. To move the NEAs by remote control, unmanned *space tugs* with advanced propulsion types will be used. A pair of space tugs is docked to each asteroid using drilling anchors. The rocket engines of the tugs then apply the thrust forces for the manoeuvres. Once stabilized in Earth orbit beyond the Moon, the mining process is started along the major axis of the asteroid. A *manned space station* will be connected to the asteroid, carrying digging, conveying and processing machinery and storage modules.

The *active mining head* initially digs a *main central tunnel* of 8 m diameter to the center of the asteroid (Taylor et al. 1995). Then it excavates a spherical or cylindrical cave up to 50% of the NEA's volume. While the mass of the asteroid decreases constantly, its orbit is stabilized by the two space tugs. The processed material is transported by *unmanned cargo ships* to low-Earth orbit or to the Lagrange points of the Earth–Moon system for further industrial use. In the last phase of mining the inner surface of the cave is sintered by a robotic laser device.

After the end of the mining process, a *rotating human habitat* can be built inside the cave. The remaining shell of the asteroid will provide shelter against cosmic rays, solar flares and micrometeorites. The various materials produced by asteroid mining can be used for construction. Oxygen, hydrogen and carbon can be extracted from C-type and similar NEAs. Natural sunlight can be collected outside the asteroid's shell by parabolic mirrors and beamed into the cave through the central tunnel.

In a future scenario we can imagine dozens of *asteroid colonies* orbiting the planet Earth each with several thousand inhabitants (Grandl 2017).

22.2 A Short History of Space Colony Design

If we accept the survival of *homo sapiens* as an ultimate purpose, we have to stretch the concept of nature beyond the biosphere and have it comprehend also the cosmic evolution with all its dangers to survival. Global and cosmic deviations and incidents, be they on a smaller or larger scale, are among the most terminal menaces to the survival of most or all species on Earth.

To build self-sustaining space stations and colonies using lunar and asteroid resources will be a crucial step of human evolution; it will definitely establish human civilization in space. To be independent from Earth is a “life insurance” for the human species in case of global disasters caused by nature itself, like super-volcanoes, ice ages, asteroid impacts, novae and other cataclysmic events that we may expect in the far future.

Besides a successor to the ISS, a permanent lunar base will be the first step within the next decades. Probes to NEAs and the development of methods to deflect hazardous asteroids will be some of the next milestones in space technology. NEAs contain important elements such as iron, aluminum, uranium and gold. Some asteroids consist of carbon and water ice, some even contain rare-Earth elements. Future mining of asteroids will yield the raw material for an increasing space industry in Earth orbit as well as in the Lagrange points of the Earth–Moon system. Space mining and space industry will enable us to build advanced space stations and finally large rotating space colonies.

In 1926 the Russian space pioneer Konstantin Tsiolkovsky was the first to discuss the establishment of large colonies around the Earth (Gatland 1981). He designed a spinning conical habitat in which trees and plants could be grown. His proposal prefigured later and even current space colony concepts and pointed out the possibility of creating artificial gravity.

In 1928 Hermann Potočnik (pen name: Hermann Noordung), a Slovenian engineer and former officer in the Austro-Hungarian army, published an accurate design of a wheel-shaped orbital station called *Weltraumrad* (Fig. 22.1).

The rotating *Weltraumrad* was equipped with parabolic mirrors to use solar light for illumination and a small power station. It can be called the prototype design for many succeeding toroidal concepts (Mielke 1986).

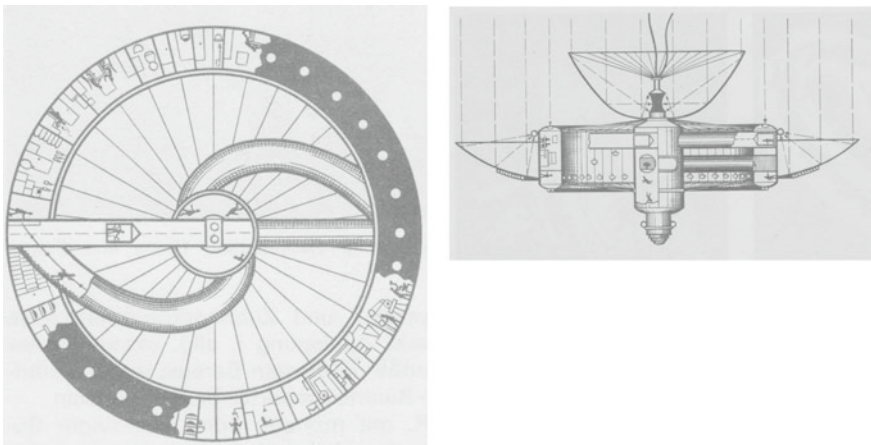


Fig. 22.1 Hermann Noordung’s (1928) *Weltraumrad* (space wheel) with a diameter of approx. 50 m and a maximum crew of 50

In 1948 the ideas of Tsiolkovsky and Noordung were recognized by H. Ross and R.A. Smith of the British Interplanetary Society. They envisaged a space station embodying wheel-shaped living quarters (30.5 m in diameter) supporting a big parabolic mirror to collect solar energy for a turbo-generator. In the early 1950s Wernher von Braun proposed a pneumatic torus as an initial orbital station. The wheel-shaped station was the first concept to use inflated structures in space. Based on von Braun's idea, NASA built experimental models in the 1960s but did not carry on the concept later. But in 1968 the famous movie *2001-A Space Odyssey*, based on the novel by British author Arthur C. Clarke, showed a wheel-shaped orbital station to the public (Fig. 22.2a).

During the 1960s NASA engineers developed various concepts of modular space stations based on the payload capacity of the Saturn V launcher, both zero-gravity and rotating low-gravity stations. Due to the decreasing NASA budget in the 1970s, all these projects were canceled and only the small SKYLAB mission was completed successfully. Meanwhile Russia (the former USSR) established the MIR orbital station, built of cylindrical modules and nodes, not unlike the former NASA concepts and somehow a precursor of the present ISS.

Despite low space budgets all over the world during the 1970s, some ambitious designs for advanced space colonies emerged from academia. In 1975 the *Stanford Torus* was developed by a group of engineers and students during the Ames–Stanford summer study. The essence of the design was a wheel-shaped habitat of largely metal construction spinning in an outer, non-rotating casing to which a shield of

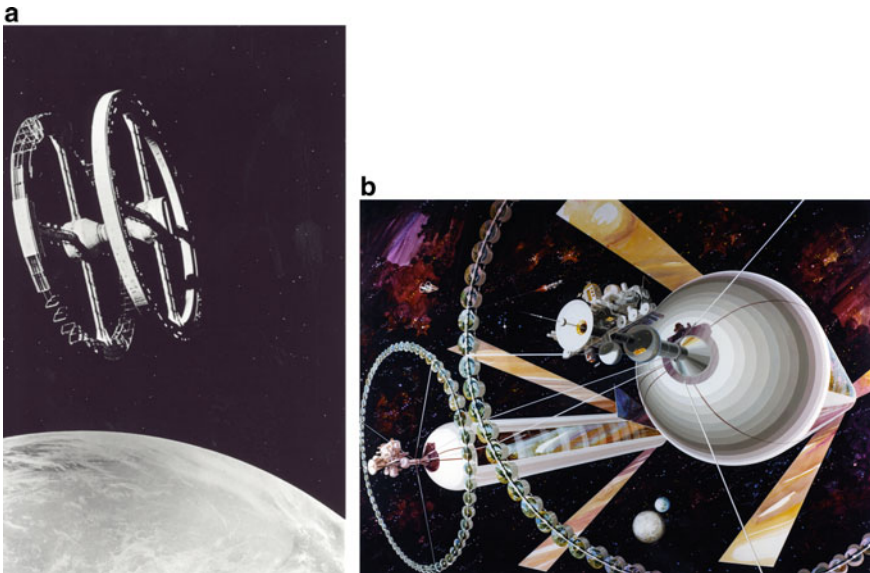


Fig. 22.2 **a** Rotating orbital station, *2001-A Space Odyssey* (1968). **b** The giant space colony, *Island 3*, by O'Neill (1975)

lunar rock has been applied for protection against radiation and meteorites. The design emphasized the use of extra-terrestrial materials, preferably from the Moon. “Mooncrete”—or lunar concrete—was to be produced by processing lunar rocks in a factory. The torus would have a diameter of 1.6 km and was considered to have a population of 10,000 people. The *Stanford Torus* would be sited in the stable Lagrange points L4 or L5 of the Earth–Moon system. The most amazing proposals for future space colonies were made by Gerard K. O’Neill, Space Studies Institute, Princeton, in 1975. Giant rotating cylinders entirely made of extra-terrestrial material, the biggest one called *Island 3*, would be 36 km long and have a diameter of 6.5 km (O’Neill 1976). A population of several hundred thousand living in an artificial bucolic landscape was proposed (Fig. 22.2b). Just like the Stanford Torus, the artificial islands in space would be sited in the Lagrange points L4 and L5. Natural illumination was considered to be provided by large adjustable mirrors, creating a 24-h cycle. O’Neill envisaged the first small *Island 1* (500 m in diameter) being built in the first decade of the twenty-first century. In view of the immense efforts and unpredictable problems it may be possible sometime in the twenty-second century. Nevertheless, O’Neill’s projects inspired one of the authors to design the Solar Arks in the 1990s in cooperation with Antonio Germano (Fig. 22.3).

In developing concepts for space colonies in the 1990s, we were strongly influenced by the studies of Gerard K. O’Neill. We have analyzed the feasibility and safety of those early utopian concepts (Germano and Grandl 1993). O’Neill’s concepts have been simple pressure vessels with single aluminum outside plating and huge glass

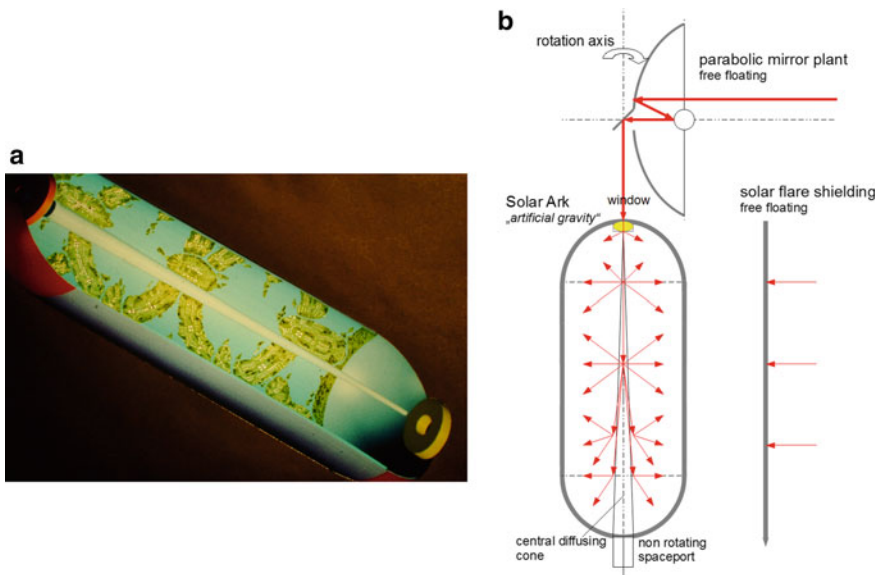


Fig. 22.3 a Solar Ark, longitudinal section (Germano and Grandl 1993). b Illumination of a Solar Ark

Fig. 22.4 Habitat structure inside the comet 107P/ Wilson-Harrington (design: Germano and Grandl 1995)



windows, exposing the artificial biosphere to cosmic rays and micrometeorites. The resulting design differs from the O'Neill concepts in the following ways. The hull of the colony consists of a space-frame structure, covered by an outer and an inner aluminum membrane. This structure stabilizes the cylinder without any internal air pressure. The outer membrane carries the thermal insulation layer and the meteorite- and cosmic ray shielding. The stable construction enables external rotation control thrusters to adjust rotation and the course of the colony. The structural cavity between the membranes is also necessary for climate control and acts as a heat exchanger. The sunlight is focused by a system of parabolic mirrors, beamed into the cylinder through a central window and distributed by a central cone (Fig. 22.3b). We propose this array for the asteroid habitats detailed in Sect. 22.5. Thus we avoid huge glass panels exposed to radiation and meteorites. Because we use natural sunlight for illumination, we called the proposed colonies Solar Arks should weigh 8 million tons length of 2300 m and 900 m diameter and have approx. 8 million tons. In the far future the largest Solar Arks could be 8 km long and 3.2 km in diameter with a mass of approx. 350 million tons.

The raw materials for construction and shielding of the Solar Arks are assumed to be taken from our Moon and from asteroids of the main belt. Considering the huge masses of material, it is clear that the use of a hollow NEA would be much easier and more feasible than to build a free-floating colony in space. In 1995 a paper on commercial resource development and utilization of the comet 107P/ Wilson-Harrington (also 1979 VA, 4015 W-H) was presented (Taylor et al. 1995). After finishing the mining process, the caverns of the comet would be used to build a rotating toroidal habitat (Fig. 22.4).

22.3 Near-Earth Asteroids (NEAs)

The number of detected NEAs has increased continuously during recent decades and their estimated population is approx. 23,000 objects. We now understand the role of asteroid impacts in the evolution of life on Earth (Alvarez et al. 1980). To ensure that mankind will survive as a species in the long run, we have to take the “asteroid threat”

Table 22.1 Geochemical groups for meteorites with typical elements occurring in mineral associations

Group	Elements (selection)
Siderophile	Fe, Ni, Co, Cu, Au, Pd, Pt, Os, Ir
Chalcophile	Fe, Ag, Cd, In, Th, Pb, Bi, and S, Se, Te
Lithophile	Rb, Cs, Be, Al, Sc, rare-Earth elements, Th, U, Ti, Nb, Ta, Cr, Mn

seriously. On one hand we will have to develop methods of detection and deflection for hazardous asteroids; on the other hand we can use these methods to modify their orbits and exploit their resources. Rare-Earth elements, rare metals like platinum group elements, etc. may be extracted more easily from NEAs than from terrestrial soil, without environmental pollution or political and social problems. Table 22.1 shows the possible resources of asteroids if we assume their mineralogy is similar to that of meteorites (Mittlefehldt 2003).

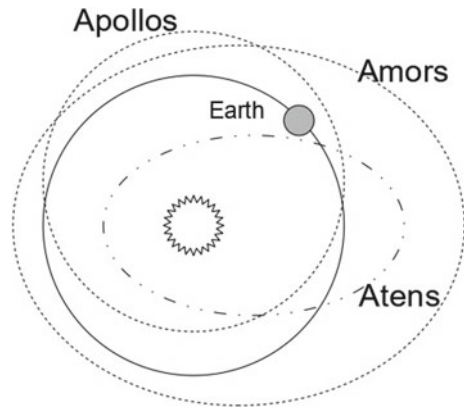
22.3.1 *Apollos, Atens and Amors*

There are three main classes of NEAs, the Amors, Apollos, and Atens (Shoemaker et al. 1979). The classes are defined on a dynamical basis, depending on the orbital elements of an asteroid, which are subject to changes over timescales of thousands to millions of years (Milani et al. 1989). The Amor class asteroids are a kind of intermediate population between main-belt and Earth-crossing asteroids. While they have typical semi-major axes (average distances from the Sun) that would place them in the main belt, nevertheless they cross the orbit of Mars and approach the Sun below 1.3 AU.¹ These distances still position them in a safe region where they do not present any immediate risk to Earth, unless they do not become Apollo-type NEAs. The Apollo group differs from the Amors in the minimum distances from the Sun, which does not exceed 1.017 AU. This value is the Earth's farthest point from the Sun, so that the Apollos may really intersect the Earth's orbit and could cause collisions. The third group, the Atens, comprises asteroids that have average distances of less than 1 AU, i.e., they move inside the Earth's orbit. Atens move out to distances above 0.983 AU, crossing the Earth orbit in its Sun-nearest point. Just like the Apollos they can cause collisions, but given their higher velocities they are more dangerous (Dvorak 1998). This fact is compensated by the relatively low number of Atens as compared to the Apollos, which represent currently the most numerous NEA group (Fig. 22.5).²

¹ Astronomical unit (AU) = average Sun–Earth distance, 150 million kilometers.

² <http://cneo.jpl.nasa.gov/stats/>.

Fig. 22.5 The three major classes of near-Earth asteroids



22.3.2 Potential Hazardous Asteroids (PHAs)

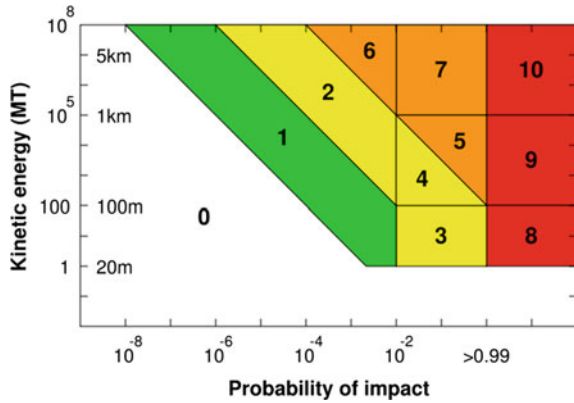
The Greek philosopher Plato (427 to 347 BCE) tells in one of his famous dialogues, “Timaios”, that in the past many disasters, “either caused by fire or water” would have devastated the Earth’s surface, killing the majority of humans. Although he is referring to old Egyptian priests for this information, he explains in the astonishing words of modern science: “This is not a legend, but really caused by the aberration of Earth-orbiting celestial bodies...” (Plato-347). Today it is becoming more and more evident that Plato was exactly describing PHA impacts that were probably on Torino Scale 8.

The Torino Scale (Fig. 22.6), created by Richard P. Binzel at the Massachusetts Institute of Technology (MIT), is a method for categorizing the impact hazard associated with near-Earth objects such as asteroids and comets. It is intended as a tool for astronomers and the public to assess the seriousness of collision predictions, by combining probability statistics and known kinetic damage potentials into a single threat value (Binzel 1999).

The Torino Scale uses categories from 0 to 10. A score of 0 indicates an object with a negligibly small chance of collision with Earth. A score of 1 indicates a pass near the Earth with no unusual level of danger. The levels from 2 to 4 merit the attention of astronomers. Current calculations give a 1% or greater chance of collision capable of localized destruction. Levels 5–6 are threatening, predicting close or very close encounters to Earth, but an impact is still uncertain. At level 7 international contingency planning is warranted, especially to determine urgently whether a collision will occur.

Levels 8–10 describe *certain collisions (red)*. A collision of level 8 or 9 causes “local” destruction by an impact and probably a tsunami. Such an event may occur between once per 60 years and once per several tens of thousands of years. At level 10 a collision is *certain*. It causes a global climatic catastrophe, extinguishing most of the living species on land and even human civilisation, and it may be the “end of

Fig. 22.6 The Torino Scale



days” for mankind. Such a cosmic disaster is expected once per 100,000 years or less often (Annals of the New York Academy of Science 1977).

22.3.3 Methods of Deflection

There are three possible techniques for deflecting PHAs:

- kinetic impact
- lateral detonation with chemical or nuclear explosives
- connecting rocket engines (a space tug) to the asteroid

At the Planetary Defense Conference, held 23–26 February 2004 in Orange County, California, by the American Institute of Aeronautics and Astronautics (AIAA), techniques for deflection of dangerous asteroids and comets were discussed.

Mark Barrera presented a study of using nuclear explosions for deflecting a NEO with today’s space technology.³ One of the example scenarios was a 200 m asteroid with only 11 years warning,... requiring use of current technology...the goal is to reduce probability of impact below 1 in 100,000. To achieve this goal we must deflect the error ellipsoid, which requires more than deflecting centreline of predicted orbit. Requires delta v of several cm/s depending on how early we can reach the asteroid and apply impulse. For nominal coupling of blast to object this plan requires a 1500 kg explosive package... all require heavy—lift launch vehicle, with multiple interceptors to improve system reliability...(Morrison 2004).

In view of this statement, it may be useful to deploy robotic interceptors in the Lagrange points L4 and L5 of the Earth–Moon system. Each interceptor should be built as a semi-autonomous robotic missile containing a chemical or nuclear explosive charge. When a hazardous object is detected, the “nearer” of the two interceptors will be alerted by the Earth station (or a future lunar base) and will start up rapidly.

³ M. Barrera, The Aerospace Corp., El Segundo, California.

When the interceptor is close enough to detect the hazardous NEO with its on-board telescope, it has to complete the mission autonomously. The missile approaches the object laterally using radar scanning and approach control. Immediately before the impact of the interceptor the blasting charge has to explode. The detonation applies a lateral force to the NEO, which deflects the celestial body and changes its path.

In 2019 ESA started the “Comet Interceptor Mission”. The goal is to launch three small spacecraft in 2029 and deploy them in the Lagrange point L2 of the Sun–Earth system. When a potential target comet is detected, one of the spacecraft tries to encounter the object, taking photos and measurements (www.cometinterceptor.space).

The more advanced technique would be to connect space tugs with the PHA to change its orbit. If the celestial body has been detected early enough, we could not only deflect it to avoid collision with Earth, but modify its track from solar orbit to a stable Earth orbit beyond the Moon (Sect. 22.4.2). In any case, high-precision astrometry will be necessary (Eggl et al. 2013).

22.4 Utilization of Asteroids

For mining and building a human habitat PHAs seem to be a favorable choice for the following reasons: they have frequent close encounters to Earth, the minimum orbit intersection distance is less than 0.05 AU (~7.5 million km) and they have diameters exceeding 150 m. In addition to the previous points, the velocity change required (ΔV) for a spaceship to reach them is typically below 12 km/s or less. We limit the candidate’s diameters to values between 150 and 500 m, as smaller objects may not be rewarding targets for mining, and much larger objects require vast amounts of energy and propellant for orbital maneuvers. By assuming a bulk density of 3 g/cm^3 we calculate an upper bound for the mass of these objects (many NEAs, for example, 1999 JU3 Ryugu, have much lower densities, of around 1.2 g/cm^3).

For the future exploitation of NEAs, we must elaborate standard probes and methods to investigate the asteroid we may choose for mining. After the NEA to be mined has been chosen, some preliminary steps are needed. In the first phase, the target needs to be well characterized and its physical properties need to be determined thoroughly. Whenever possible, ground-based observations (optical and radio measurements) will constrain the shape and rotational state of the asteroid; a spectrum of the target will enable the mission planners to derive the surface properties (spectral type). If the orbit of the NEA is known with sufficient accuracy, these measurements can be timed around the phase of minimal geocentric distance; otherwise, the next window of opportunity depends on the orbital period of the NEA, typically 0.7–3 years. As a consequence of these requirements a reasonable time scale seems to be seven years for the preliminary work. The Asteroid Mining Corporation Ltd. in the UK is currently developing a CubeSat-class satellite to analyze and determine potential asteroids for mining with an estimated launch date in 2023.

In the subsequent, second phase a small orbiter would be sent to the target NEA (preferably on a direct trajectory), imaging and mapping the surface as well as probing the gravity field, which is essential to determine the mass, density and porosity with a good level of accuracy. Density and porosity are important parameters to exclude asteroids which are very loosely bound rubble-piles that cannot be used for mining. This phase—disregarding the flight time—can be completed in one year (given the results of the Dawn mission at Vesta); eventually some sort of simple sample return process can follow. The total time allocated for this phase would be another seven years, including the preparations, flight time and active mission. The techniques for this second phase could be tested on our “quasi-satellite” 2016HO3.

In the third and main phase a group of space tugs (see Sect. 22.4.2) would approach the target and connect to it. The goal is to move the target asteroid into an Earth orbit where it can be mined safely by a manned mining station. The asteroid would be placed either into orbit beyond the lunar orbit, or into a libration orbit about the Earth–Moon Lagrange points L4 or L5. The duration of this final phase depends strongly on the orbital parameters and the mass of the asteroid, therefore only a rough assessment can be given here. Provided that the proposed propulsion systems are technically feasible, a mission time of 10–20 years is envisioned before the asteroid reaches an orbit around the Earth (Grandl and Bazzo 2013).

22.4.1 An Advanced Propulsion System

As is visible from Table 22.2, the masses of the proposed NEA targets are higher than 10^{10} kg. These huge masses are difficult to deal with using conventional propulsion systems. There is a clear need for advanced propulsion systems that are able to deliver high thrust and specific impulse. Such a system could be the Bussard Fusion System, also known as the quiet-electric-discharge (QED) engine (Bussard 1997, 2002).

This system uses electrostatic fusion devices to generate electrical power. The fuel consists of deuterium and Helium-3 which fuse to create Helium-4, plus protons

Table 22.2 List of NEA (PHA) objects as potential candidates for mining with their physical and orbital parameters. For the calculated mass value, a spherical shape with homogeneous density of 3 g/cm^3 has been assumed (Grandl and Bazzo 2013)

Designation	Diameter [m]	Mass [kg]	Semi-major axis [AU]	Eccentricity	Spectral type
2004 MN4	270	3.092×10^{10}	0.922	0.191	Sq
1982 DB	330	5.645×10^{10}	1.489	0.360	Xe
1998 SF36	330	5.645×10^{10}	1.324	0.280	S
2005 YU55	400	1.005×10^{11}	1.157	0.430	C
2008 EV5	450	1.431×10^{11}	0.958	0.084	S
1982 XB	500	1.963×10^{11}	1.835	0.446	S
1999 RQ36	493	1.882×10^{11}	1.126	0.204	C

Table 22.3 Energy and fuel mass required for transfer of an NEA into Earth orbit. The ΔV values were taken from a table of JPL (Grandl and Bazzo 2013)

Designation	Mass [kg]	Energy [J]	Fuel mass [kg]	DeltaV [km/s]
2004 MN4	3.092×10^{10}	1.161×10^{18}	3317	5.687
1982 DB	5.645×10^{10}	8.265×10^{18}	23,614	4.979
1998 SF36	5.645×10^{10}	6.164×10^{18}	17,611	4.632
2005 YU55	1.005×10^{11}	6.094×10^{18}	17,411	6.902
2008 EV5	1.431×10^{11}	2.778×10^{18}	7937	5.629
1982 XB	1.963×10^{11}	3.984×10^{19}	113,829	5.490
1999 RQ36	1.882×10^{11}	9.392×10^{18}	26,834	5.087

releasing 18.3 meV of energy per reaction. When the charged protons escape from confinement, their kinetic energy can be converted to electricity or be used directly as a plasma beam for generating thrust. The advantage of the Deuterium–Helium reaction is the low neutron production rate (via Deuterium–Deuterium reaction), as neutrons are unavailable for generating thrust; the disadvantage is that Helium-3 is rather rare on Earth. However, it is more abundant on the Moon (via solar-wind deposition), so that it first has to be gathered from there, which could increase the mission costs. For the reaction a specific energy of 3.5×10^{14} J/kg can be computed (Bussard 1997, 2002), i.e., orders of magnitude higher than for any existing propulsion system (Fraser 2013).

From Table 22.2 the differences in Kepler energy between the NEA’s current orbit and the Earth’s orbit can be calculated to estimate the amount of (specific) energy needed for the transfer (Roy 1988, Sect. 11.3). Then the energy needed is this difference multiplied by the object’s mass, where we have assumed a bulk density of 3 g/cm^3 as mentioned above. The given ΔV values were taken from a table by Benner (JPL)0.4⁴

Table 22.3 gives details of the fuel mass, which is estimated by dividing the energy column by 3.5×10^{14} J/kg for the De-He-3 reaction.

We take as an example the near-Earth object designated 2008 EV5, an Aten group asteroid with a mean diameter of 450 m that belongs to spectral type S (stony asteroid). Our mass estimate (using a bulk density of 3 g/cm^3) is 1.431×10^{11} kg. To bring it to an Earth-like orbit, the transfer energy required is estimated to be 2.778×10^{18} J.

$$2.778 \times 10^{18} \text{ (J)} / 3.5 \times 10^{14} \text{ (J/kg)} = 7937 \text{ kg (about 8 tons of He-3 fuel)}$$

If the orbit is arranged in such a way that the maximum approach to Earth is still a safe distance beyond the Moon’s orbit, but close enough to take advantage of a swing-by maneuver, then the remaining ΔV can be supplied through gravity assist from the Moon.

⁴ http://echo.jpl.nasa.gov/~lance/delta_v/.

22.4.2 Orbit Modification of Asteroids

To achieve orbit modification, it is necessary not only to develop and test the Bussard Fusion System but also to create an unmanned space tug which is propelled by Bussard engines and can carry the required fuel. All components of this spaceship should be designed for series production to reduce costs in the long run. Therefore, all parts should be modular and easy to assemble in low-Earth orbit (LEO). As a main supporting structure, we propose cylindrical modules built of aluminum trapezoid sheeting, stringers, frames and bulkheads. Considering possible heavy-lift rockets, which could be derived from existing launchers like Ariane 6, we may limit the size to 24 m length and 8 m diameter. The maximum payload for one module containing a fusion engine, a drilling anchor and technical equipment, should be approx. 20–25 tons.

In LEO four modules will be assembled to a space tug as shown in Fig. 22.7. The cylinders are filled with fuel in space to reduce the launch mass on Earth and because Helium-3 is extracted from lunar soil. This implies that a lunar base (Grandl 2012) has to be established before sending space tugs to the asteroids.

For the example target, the primary fuel mass is approx. 8000 kg (Table 22.3). But we have to add the fuel needed by the tugs to reach the NEA. We have also to consider some fuel reserve for adjusting manoeuvres after the final swing-by. For these additional thrusts to stabilize the NEA we need a second space tug, which can apply counterforces to the first tug. So we assume a total fuel mass of approx. 9000 kg for the 2008EV5 mission. The missions to some of the NEAs in Table 22.3 will require much more fuel. In those cases we can add external tanks between the modules.

Fig. 22.7 Asteroid Space Tugs (Grandl and Bazso 2013)

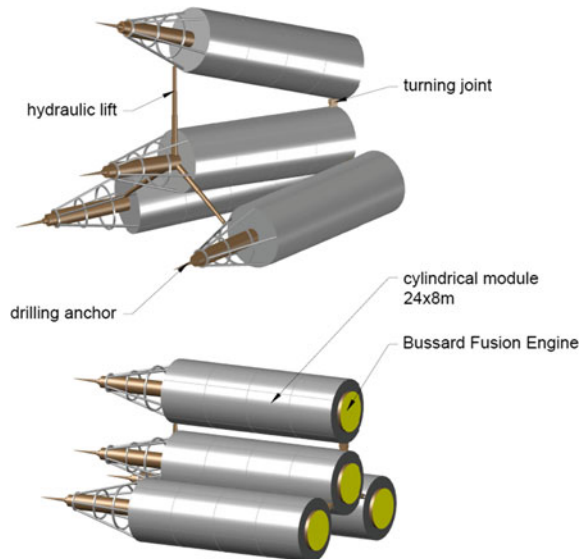
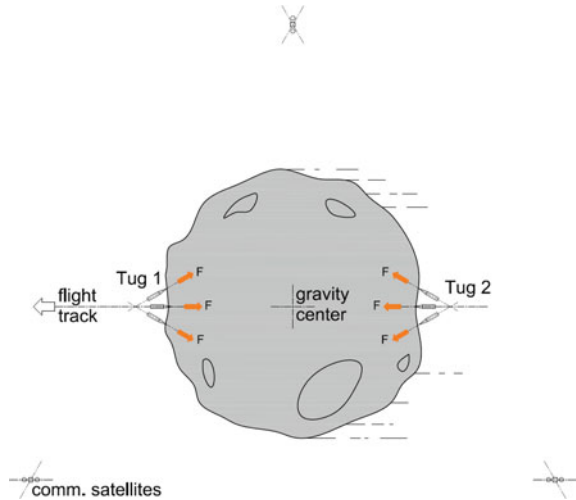


Fig. 22.8 Asteroid, guided by space tugs and communication satellites



Thus we send a couple of asteroid space tugs to the selected asteroid. When the tugs reach the NEA they deploy their drilling anchors and penetrate the target as shown in Fig. 22.8. Once fixed to the asteroid rock the tugs can apply forces in any direction by firing their engines. The first tug, which carries most of the fuel, applies the primary force for the orbital manoeuvres. The second one adjusts the flight track by short engine thrusts.

All the manoeuvres would be supervised by a mission-control center on Earth. Therefore the tugs are accompanied by three small communication satellites, which orbit the NEA and enable remote control from Earth or from the lunar base.

An asteroid can be moved from a solar orbit to an Earth orbit by various transfer mechanisms (Chobotov 1991).

- Hohmann transfer

Assuming an initially circular orbit for the NEA—this assumption is generally not met, as many NEAs have moderate to quite high eccentricities—a simple Hohmann transfer from the original NEA orbit to Earth orbit can be applied. If the engine bursts are considered to be instantaneous velocity “kicks”, the engines have to deliver peak performance (high thrust). Low-thrust engines, on the other hand, can gradually change the orbit and operate more efficiently, but the necessary change in velocity will be up to 141% higher, and the mission duration will be longer.

- Bi-elliptic transfer

A bi-elliptic transfer involves two elliptic transfer orbits, but only one-half of each is needed. The engines have to fire three times, first to leave the original orbit, second to change the elliptic transfer orbits, and finally to arrive at the final orbit. Under certain circumstances the bi-elliptic transfer requires less fuel than a Hohmann transfer, although the travel time is generally longer.

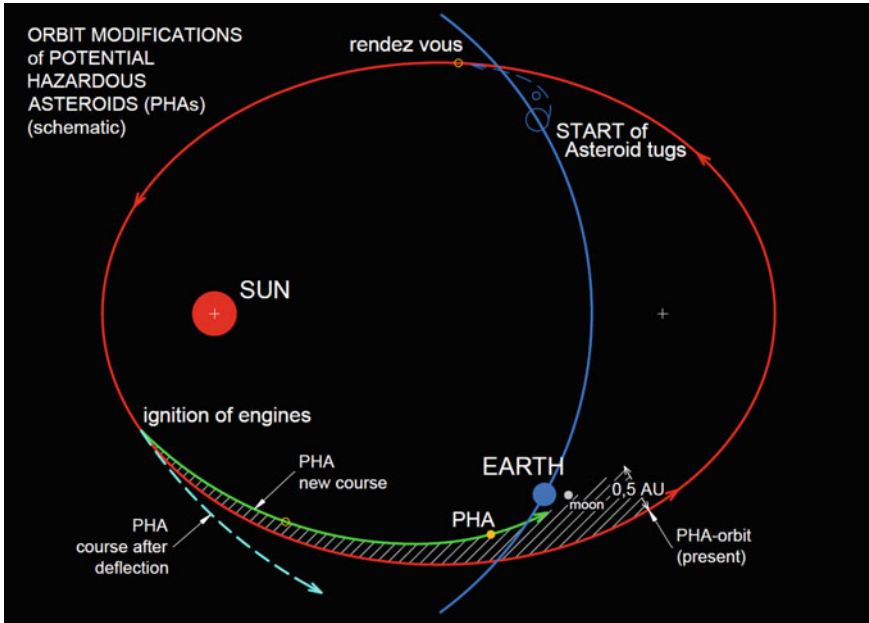


Fig. 22.9 Schematic drawing of possible PHA orbit modifications; light blue: deflection by chemical or nuclear blasting green: capture of a PHA by space tugs for mining

Figure 22.9 shows possible modifications of a PHA’s orbit: either deflection to avoid a possible collision with Earth, or “catch” the PHA and force it into an Earth orbit beyond the Moon for mining (Fig. 22.10).

22.4.3 Mining and Processing

When the PHA has been stabilized in Earth orbit beyond the Moon, a *manned mining station* is docked to the asteroid. The station is built of cylindrical modules and nodes. It contains a drilling machine, called the *Active Mining Head* (Taylor et al. 1995), conveying, processing machinery, storage and docking modules. Rotating habitat modules provide a small artificial gravity for the crew. Electric current is provided by solar panels and a small nuclear battery (Fig. 22.11).

Mining an asteroid will be very different from mining on Earth because of low or zero gravity. On one hand it is easier to dig tunnels and caves because one needs fewer and smaller ring beams, rock bolts, etc., to stabilize the rock. On the other hand, in a low-gravity environment it is much more difficult to convey the excavated rock particles, known as “muck”, which are chipped away by the Active Mining Head (AMH). On the asteroid’s surface these rock chips can float away and be lost easily. For this reason we prefer underground mining to open-cast mining. The

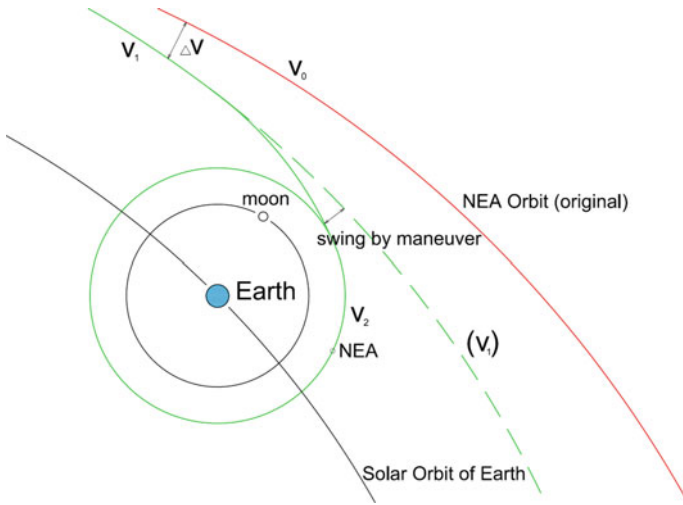


Fig. 22.10 Swing-by manoeuvre of the PHA into a stable Earth orbit beyond the Moon

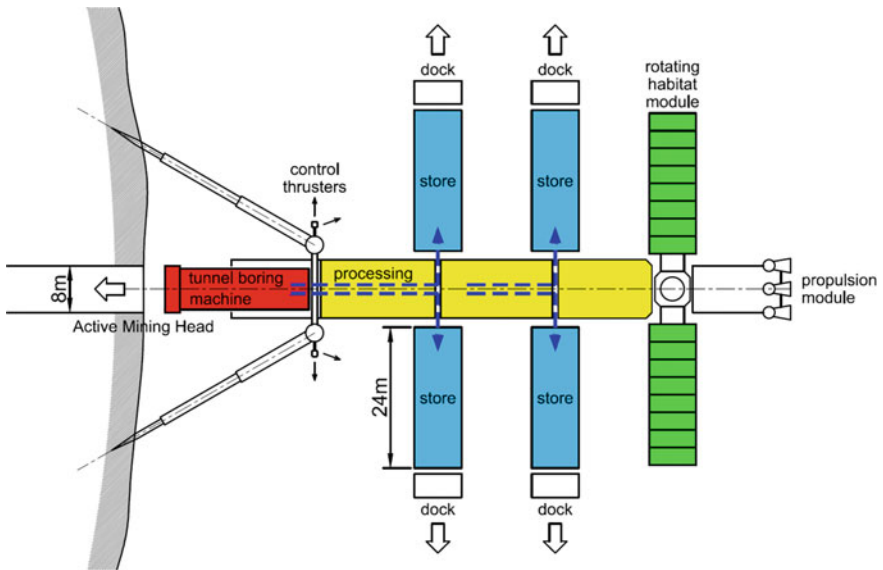


Fig. 22.11 Manned mining station, schematic

asteroids which we choose for mining should have a minimum bulk density of approx. 2.5 g/cm^3 . At lower density the drilling could destabilize the structural constitution of the soil and cause dangerous fractures in the asteroid.

The AMH is a flexible drilling, digging and tunnel-boring machine. The drill bits are furnished with cutting elements of sintered artificial diamonds. The AMH must

work more slowly, smoothly and precisely than in a terrestrial mine, so as not to disturb the structural stability of the asteroid rock. First it drills a *main central tunnel* of 8 m diameter to the center of the asteroid, and then it excavates step by step a spherical cave up to 50% of the NEA's volume. The excavated cavern is permanently filled with a pressurized gas. Thus the muck can be removed easily in a flexible vacuum conveyor tube. In the manned mining station the mined material is processed, stored and prepared for transport. Unmanned *cargo ships* transport the material to LEO or the Lagrange points of the Earth–Moon system for further industrial use, e.g. in metallurgical plants. The rate of mining advance, muck removal and storage must be kept equal to the rate of cargo shipping, which can be done more easily when the NEA is in an Earth orbit than in its original solar orbit.

The inner surface of the tunnel and the cave cannot be lined with concrete (shotcrete) as on Earth because of the lack of air and water. After the end of the mining process a smooth walling can be made by laser sintering of the rock surface. The thickness of the NEA's remaining crust depends on the bulk density of the asteroid rock and on the diameter of the celestial body and the cave. It should be at least approx. 30–50 m.

After completion of the mining process and the laser sintering of the inner surface, the remaining hollow asteroid can be used as a shelter for industrial facilities or for the storage of products like water, oxygen or other gases. The rock hull provides shelter against micrometeorites, cosmic rays, solar flares, and last but not least, thermal insulation. NEAs with more than 400 m diameter can be used for *human habitats* with artificial gravity.

22.5 Building a Habitat Inside an Asteroid

After exploiting up to 50% of an NEA's volume by mining, we can build a *rotating toroidal habitat* inside the remaining cave. The shell of the hollow asteroid provides a good shelter against meteorites, cosmic rays and solar flares. The radius of the rotating torus should be at least 100 m, to minimize the Coriolis acceleration (Puttkamer 1987). Thus we select an asteroid with a minimum diameter of 400 m, e.g., 2008 EV5, for a prototype asteroid habitat.

22.5.1 Cosmic Rays and Solar Flares

Any spacecraft or space station in orbit or on the surface of the Moon or Mars has to provide shelter against cosmic rays, solar flares, micrometeorites and huge temperature amplitudes. The temperature amplitude on the Moon is from approx. $-170\text{ }^{\circ}\text{C}$ to $+130\text{ }^{\circ}\text{C}$. For this reason, a lunar or Martian base has to be covered with huge masses of regolith. The longer a person stays on the Moon the higher the regolith layer should be, up to 3 m (Benaroya 2018). Therefore lava tubes on the

Moon and Mars may be used for the permanent presence of humans. It is evident that a hollow asteroid with a remaining shell of 30 m thickness is the ultimate solution for shielding.

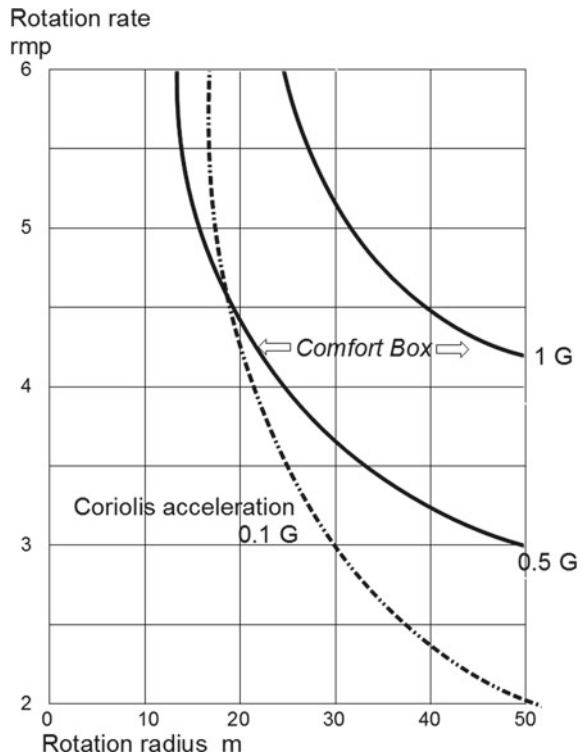
22.5.2 Artificial Gravity

The lack of gravitation in outer space and on celestial bodies much smaller than Earth implies many dangers to human health. The primary threats are bone demineralization, muscle atrophy and orthostatic intolerance. It is evident that the simulation of gravity by the use of centrifugal forces will provide comfortable environments for humans in space; the bigger the radius, the better the conditions.

At large radii the Coriolis acceleration, which may disturb the vestibular sense, can be neglected. In 1987 NASA engineer Jesco von Puttkamer published the “comfort box” concept (Fig. 22.12), which defines acceptable living conditions in a spinning space station (Puttkamer 1987).

According to Fig. 22.12, a rotating space station should have a minimum radius of 30 m and a spin rate of 4 rpm (rotations per minute) to simulate an artificial gravity of

Fig. 22.12 Comfort box defined by rotation rate and radius of a space station



0.9 g. To create 1 g we need 50 m radius and a spin rate of approx. 4.2 rpm. The goal should thus be to build future space stations as well as an interplanetary spaceship. A crucial problem to be solved by engineers will be the design of the joint between rotating and non-rotating parts of the structure.

22.5.3 A Prototype Habitat

The various materials produced by asteroid mining can be used for construction. Oxygen, hydrogen and carbon can be extracted from C-type and similar NEAs. Natural sunlight can be collected outside the asteroid's shell by parabolic mirrors and beamed into the cave through the central tunnel.

The torus can be built of prefabricated pneumatic elements. After assembling the elements, the whole torus is inflated. Once sheltered by the shell of the NEA we can use big glass panels for illumination. Natural sunlight can be beamed into the center of the cave by an array of parabolic mirrors. The focused sunlight is distributed by a central mirror cone into the torus. The parabolic mirror array is designed as a free-floating structure outside the asteroid, with independent rotation to collect the sunlight. The central mirror cone is furnished with small parabolic facets to distribute the light.

The radius of a rotating torus inside the asteroid shell should be at least 100 m, to minimize the Coriolis acceleration (Figs. 22.13 and 22.14).

As an alternative habitat design, one could imagine living on the inner surface of a rotating hollow asteroid instead of building an artificial torus. Given a bulk density about 3 g/cm³ the resulting tensions caused by centrifugal forces would limit the possible "artificial gravity" (Maindl et al. 2019a, b).

22.5.4 Living Conditions and Housing

The entire toroidal structure will be rotating approximately 3 to 4 rotations per minute by the use of circular electromagnetic bearings (magnetic levitation) and provide up to 90% of terrestrial gravity. In case of a failure the habitat will be supported by additional mechanical bearings. The Coriolis acceleration will be approx. 0.05 g, still comfortable for humans.

The non-rotating *spaceport* at the asteroid's surface provides docking and storage modules and is derived from the former manned mining station. All interior buildings and furniture are lightweight constructions, partially made by 3D plotting and similar methods. To mold a landscape we can use the slag remaining from the asteroid mining processes. Additionally lightweight materials like carbon fibers and foamed metals will be used.

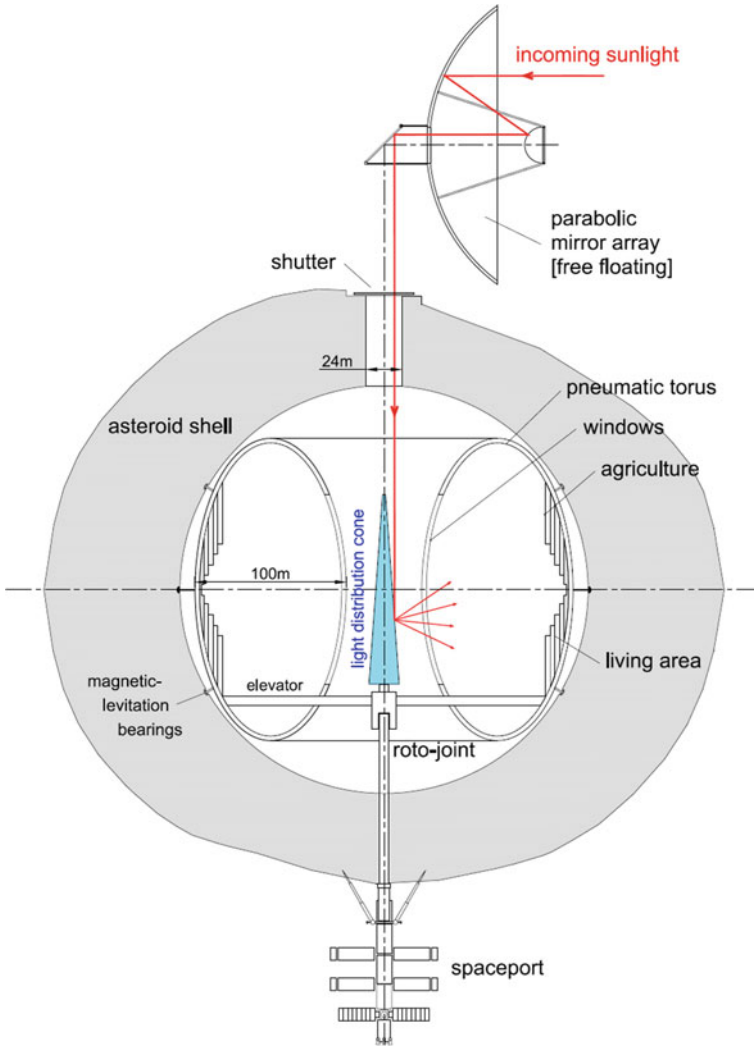
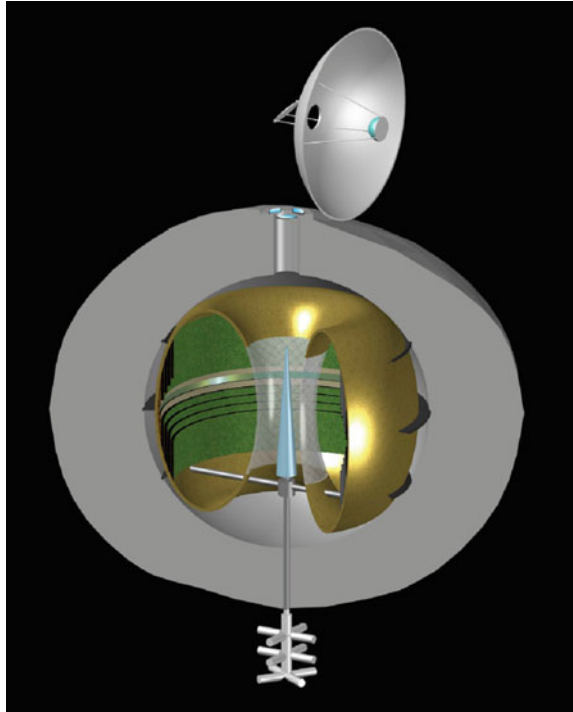


Fig. 22.13 Prototype asteroid habitat, cross-section (Grandl and Bazso 2013)

22.5.5 Climate and Agriculture

To avoid thermal overloading of the cave, a system of heat absorbers and heat exchangers will be installed inside and outside the asteroid shell. Thus the surplus heat can be absorbed, transformed into electrical power or radiated from the surface as microwaves. Oxygen and water are extracted from asteroid material during the mining process or in industrial plants in space. The goal is to generate a “closed water cycle” in a space colony that is almost self-sufficient. By adjusting the shutter

Fig. 22.14 Prototype asteroid habitat, cross-section/perspective view; The rotating torus is driven by electromagnetic bearings in the vacuum cavern of the asteroid. The free-floating mirror array in space beams the sunlight into the cave, where the light is finally distributed by a central mirror cone (Grandl and Bazso 2013)



elements at the top of the tunnel we can simulate a 24-h cycle of day and night. The simulation of “seasons” within limits is also possible.

The inner surface of the torus is used for housing, gardening, agriculture (permaculture) and a public center. Vertical farms and aquaculture provide nutrition for approx. 2,000 inhabitants.

22.6 Conclusions and Future Scope

Planet Earth is just the cradle of mankind. The possible expansion of human life into the solar system depends on our decisions. But can we afford the enormous costs for this giant leap? Yes, we can! The money for big space enterprises can easily be taken from military budgets all over the world, if we successfully avoid fighting wars. But this is politics.

In the coming centuries some hundreds of thousands of people could settle in the Earth–Moon system and constitute a new society beyond planet Earth. Solar-powered satellites and industrial facilities will orbit Earth and Moon, producing rare products for mankind’s increasing population. Space elevators will transport materials and products between Earth and low-Earth orbit (Pearson et al. 2005). Spaceships will

start from lunar orbit and the Lagrange points to Mars, the asteroid belt and the moons of Jupiter to search for scientific knowledge and to use the resources of the entire solar system for the benefit of mankind. Earth then can become a “green planet”, because most of the mining and heavy industry processes will be deployed in space. Arrays of space telescopes will survey the solar system to detect and help to deflect hazardous celestial bodies before they approach the Earth–Moon system. Finally human life in the solar system beyond planet Earth may become the next step of biological evolution.

Acknowledgements The authors wish to thank Rudolf Dvorak and Akos Bazso, Vienna Institute of Astronomy, for their support over many years.

We also sincerely thank Haym Benaroya, Department of Mechanical & Aerospace Engineering, Rutgers University, New Jersey.

References

- Alvarez, L.W., W. Alvarez, F. Asaro, and H.V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 208: 1095–1108.
- Annals of the New York Academy of Science, 822 (1977)
- Benaroya, H. 2018. *Building habitats on the moon*, 2018. Engineering Approaches to Lunar Settlements: Springer-Verlag.
- Binzel, R.P., T. Gehrels, and M.S. Matthews (eds). 1999. Asteroids II, pp. 316–335.
- Bussard, R.W. 1997. System technical and economic features of QED-engine drive space transportation. In 33rd AIAA/ASME/SAE/ASEE joint propulsion conference, AIAA 97–3071.
- Bussard, R.W. 2002. An advanced fusion energy system for outer-planet space propulsion. *Space Technology and Applications International Forum* 608.
- Chobotov, V.A. 1991. *Orbital mechanics*. Washington, DC: AIAA Education Series. AIAA, Inc.
- Dvorak, R. 1998. The dynamic evolution of the Atens (asteroids). *Celestial Mechanics and Dynamical Astronomy* 69 (1–2): 103–118.
- Eggl, S., A. Ivantsov, D. Hestroffer, D. Perna., D. Bancelin, and W. Thuillot. 2013. High precision astrometry in asteroid mitigation-the NEO shield perspective, French Society of Astronomy & Astrophysics: SF2A-2013. In *Proceedings of the annual meeting of the French society of astronomy and astrophysics* ed. Cambresy, L., Martins, F., Nuss, E., Palacios, A., pp.169–176.
- Fraser, S. 2013. Electric power system options for robotic miners. In Badescu, V. ed. *Asteroids-prospective energy and material resources*, pp. 247–270. Springer.
- Gatland, K. 1981. *Space technology*. London: Salamander Books Ltd.
- Germano, A. and W. Grandl. 1993. Astropolis-space colonization in the 21st century. In Faughnan, B. (ed.) *Proceedings of the eleventh SSI-Princeton conference on space manufacturing 9: the high frontier, accession, development and utilization*, May 12–15, pp. 252–268. American Institute of Aeronautics and Astronautics.
- Grandl, W. 2017. Human life in the solar system. *REACH-Reviews in Human Space Exploration* 5C: 9–21.
- Grandl, W. 2012. Building the first lunar base—construction, transport, assembly. In *Moon-prospective energy and material resources*, ed. V. Badescu, 633–640. Springer.
- Grandl, W. and A. Bazso. 2013. Near Earth Asteroids-prospection, orbit modification, mining and habitation. In Badescu, V. (ed.) *Asteroids-prospective energy and material resources*, pp.415–438. Springer.

- Maindl, T, R. Miksch, and B. Loibnegger (2019a) Stability of a rotating asteroid housing a space station. *Frontiers in Astronomy and Space Sciences*, 6: 37.
- Maindl, T., C. Schäfer, B. Loibnegger, and R. Miksch. 2019b. Tensile loads in porous rotating asteroids with artificial caverns. In *50th lunar and planetary conference*, held 18–22 March 2019b at The Woodlands, Texas. LPI Contribution No. 2132, id.1094.
- Mielke, H. 1986. *Lexikon der Raumfahrt und Weltraumforschung*. Berlin.
- Milani, A., M. Carpino, G. Hahn, and A.M. Nobili. 1989. Dynamics of planet-crossing asteroids-classes of orbital behavior. *Icarus* 78: 212–269.
- Mittlefehldt, D.W. 2003. Achondrites. *Treatise on Geochemistry* 1: 291–324.
- Morrison, D. 2004. NASA Planetary Defense 2, 27 Feb 2004, NEO News.
- O’Neill, G.K. 1976. *The high frontier-human colonies in space*. Princeton, New Jersey: Space Studies Institute Press (1989).
- Pearson, J., E. Levin, J. Oldson, and H. Wykes. 2005. Lunar space elevators for cislunar space development-phase I final technical report. Star Technology and Research Inc., 3213 Carmel Bay Drive, Suite 200, Mount Pleasant, SC, 29466-8513.
- Plato (347 BCE). 1578. *Platonis opera quae extant omnia. Ex nova Ioannis Serrani interpretatione. Excudebat Henricus Stephanus*; Gent 1578.
- Puttkamer, J.V. 1987. *Der Mensch im Weltraum-eine Notwendigkeit*. Frankfurt am Main: Umschau-Verlag.
- Roy, A.E. 1988. *Orbital motion*. Bristol Philadelphia: Institute of Physics Publishing.
- Shoemaker, E.M., J.G. Williams, E.F. Helin, and R.F. Wolfe. 1979 Earth-crossing asteroids-orbital classes, collision rates with earth, and origin. In Gehrels, T. (ed.) *Asteroids*, pp. 253–282.
- Taylor, T., A.C. Zuppero, A. Germano, and W. Grandl. 1995: IAA-95–IAA.1.3.03 commercial asteroid resource development and utilization. In *46th international astronomical congress*, Oslo, Norway

Chapter 23

Resources from Asteroids and Comets



Daniel Britt and Kevin Cannon

Abstract Key to understanding the nature and scope of asteroid and comet resources are the fundamental differences between terrestrial economic geology and the driving processes creating extraterrestrial resources. What creates resources on asteroids and comets are processes that occurred during or immediately after solar system accretion and have been dormant for the vast majority of the solar system's 4.5-billion-year history. The major resource potential from asteroids for the foreseeable future will be volatiles and iron. The sources of these resources will be hydrated carbonaceous chondrite parent bodies and core fragments of differentiated asteroids. We can make informed estimates of the mineralogy of near-Earth space small asteroids based on the over 70,000 meteorite samples that have been collected, along with decades of telescopic observations, and the results of several spacecraft flyby and rendezvous missions. Small asteroids are overwhelmingly single mineralogical assemblages and overwhelmingly structural rubble piles. Because of their simpler mineralogy, most asteroid "prospecting" can be done remotely via ground-based telescopes.

23.1 Introduction

Before starting on the detailed discussion of resources from asteroids and comets, the following are key takeaways about the economic geology of asteroids and comets.

- We know a great deal about the mineralogy of asteroids through the study of over 70,000 recovered meteorites.
- The heating, differentiation, and active tectonics that drive mineralization on Earth have been absent from asteroids since the first few million years of solar

D. Britt (✉)

Department of Physics, University of Central Florida, 4111 Libra Drive, PSB 430, Orlando, FL 32816-2385, USA
e-mail: dbritt@ucf.edu

K. Cannon

Colorado School of Mines, 1500 Illinois St, Golden, CO 80401, USA
e-mail: cannon@mines.edu

system history. Mineralization on asteroids has none of the characteristics found in terrestrial ore bodies.

- The major resource potential from asteroids for the foreseeable future will be volatiles and iron. The sources of these resources will be hydrated carbonaceous chondrite parent bodies and core fragments of differentiated asteroids.
- Small asteroids are overwhelmingly single mineralogical assemblages. Any small mineralogical variation in asteroids is the result of the accreted debris from the impact of meteoroids. Searching for ore bodies and “high grading” are meaningless concepts on small asteroids.
- Because small asteroids are single mineralogical assemblages, the bulk of asteroid prospecting can be done remotely via telescope.
- Small asteroids are overwhelmingly rubble piles.
- The major early asteroidal resources are going to be largely confined to volatile-rich (resource water) and metal-rich (resource iron and nickel) asteroids.
- Comets, while a major potential source of organic carbon and volatiles, will not be an economic resource for the foreseeable future. Cometary orbits tend to be highly inclined to the ecliptic and highly eccentric, so the energy cost of rendezvous and return for these objects will be prohibitively high.

23.2 Accretion and Asteroid/comet Source Regions

Comets and asteroids are leftovers from the original accretion in the solar system. One of the key concepts for understanding the nature and scope of asteroid resources is that minerals evolve. What creates ore bodies and economic resources on Earth are a complex interplay between tectonics, abundant volatiles (i.e., water), and the slow release of the Earth’s heat. On comets and asteroids these processes are almost entirely absent. What creates resources on asteroids and comets are processes that occurred during or immediately after solar system accretion and have been dormant for the vast majority of the solar system’s 4.5-billion-year history.

The solar system started as a cold molecular cloud composed of roughly 98% hydrogen and helium, and 2% of all the other elements. This molecular cloud was affected by a nearby supernova which seeded what became the protoplanetary nebula with an abundance of short-lived radioactive elements such as aluminum-26 and iron-60. These isotopes have relative short half-lives of 717,000 and 2.6 million years respectively, so their effect on accreting asteroids is over in a few million years. Because of the abundance and short half-life of ^{26}Al , the heat generation potential of this element is about one million times that of uranium. These abundant “hot” isotopes had the effect of heating early accreting planetesimals to melting temperatures. The other effect of the supernova was a shockwave that compressed the molecular cloud and started a gravitational collapse that produced our solar system. The physics of the collapse in the nebula disk and the accretion of the solar system are complex and beyond the scope of this chapter, but the key points for comet and asteroid resources are the following.

- (1) There was a radial gradient in density and temperature, both being high toward the center of the collapsing nebula and falling off with distance.
- (2) High-temperature and low-temperature materials segregated with radial distance. The inner regions tended to be poor in volatiles and ices because of the high temperatures, but rich in silicate minerals. The outer regions (distances greater than 3 astronomical units) tended to preserve ices, organics, and low-temperature minerals.
- (3) While temperature affected what types of materials could form and survive in different zones of the solar system, the solar nebula was still turbulent and mixed high-temperature minerals into all zones.
- (4) Even though there was mixing, zones in the solar system formed with distinct chemical traits driven by differences in oxidation state and the availability of volatiles.
- (5) Accretion of the planets and asteroids was very rapid with the bulk of the asteroids accreting within roughly 4 million years (e.g., Desch et al. 2018). In the solar nebula what minerals were formed depended critically on the abundances of available elements. Hydrogen and helium are by far the most abundant elements in the universe, but for all the other elements their abundances are driven by stellar nucleosynthesis. Again, the physics of this process are beyond the scope of this chapter, but the bottom line is that a relatively few elements dominate the formation of most minerals. For example, 93% of the mass of the whole Earth is composed of just four elements, iron, oxygen, silicon, and magnesium in that order (Hazen et al. 2008). Not surprisingly the composition of terrestrial minerals and the silicate portion of asteroids are dominated by these elements. For low-temperature accretion we add in the hugely abundant element hydrogen (75% of the normal matter in the universe), usually in the form of water ice bound chemically with oxygen.

The timing of accretion from dust into planets was largely driven by the density of available material. In the inner regions of the solar system where nebular material densities and temperatures were highest, accretion occurred early and was dominated by high-temperature minerals such as silicates and metal, while being depleted in volatiles. Early accreting asteroids also incorporated high levels of short-lived radioactive elements. The heat from the decay of the short-lived radioactive elements turned the early accreting asteroids into melted balls of liquid rock and metal. This is the source of the iron-rich asteroids (meteorite types and their expected asteroid parents are detailed in Table 23.1). Asteroids that accreted early melted and differentiated, with the heavier iron and siderophile elements sinking to the center of the asteroid to a metal-rich core. The lighter silicate minerals formed the mantle and crust. Some of these differentiated or “igneous” (i.e., volcanic) asteroids survived the age of the solar system, such as 4 Vesta, which is the largest example of a differentiated asteroid (or minor planet). For most differentiated asteroids, subsequent collisional impacts have shattered and stripped off the silicate mantle and crust, leaving the much stronger bare metallic cores exposed. The cores are composed largely of iron and nickel with varying minor amounts of siderophile elements (those that dissolve

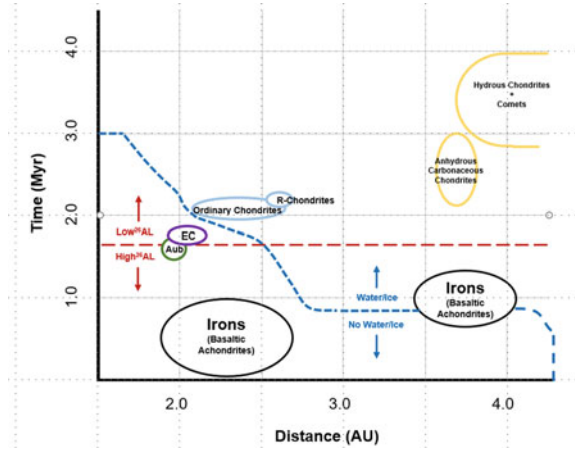
readily in iron) and chalcophile elements (those that combine readily with sulfur). Studies of the more than 1200 iron-nickel meteorites in the collections suggest that we have samples of 40–60 iron asteroid parent bodies (Hutchison 2004).

The primary resource potential for iron-nickel asteroids, of course, is iron. There has been discussion of platinum-group elements as resources and even of solid platinum asteroids. Let us be clear, however, solid platinum asteroids do not exist and cannot exist. What we have learned from studies of over 1200 iron-nickel meteorites is that siderophile elements (which include the platinum-group elements) are disseminated at the micron level within the bulk of iron-nickel. Because these objects were subjected to a single heating event which produced their differentiation and then slow cooling within the core of an asteroid, there is no mechanism for subsequent processing and concentration of elements into higher-grade ores of any sort (Hazen et al. 2008). What is available in iron-rich asteroids are disseminated siderophile elements such as 2–12 parts per million platinum (i.e., 2–12 grams per ton), 1–2 parts per million gold, 1–20 ppm iridium, up to 400 ppm germanium, and up to 100 ppm gallium (Papike 1998; Hoashi 1990). Again, these elements are not concentrated into veins or enriched ores but disseminated at the micron level throughout the iron body.

As accretion continued, later-forming asteroids incorporated progressively smaller amounts of short-lived radioactive elements. As a result, asteroids that accreted after a few million years did not have sufficient heat to fully melt. For asteroids accreting in the inner asteroid belt where volatiles were depleted and silicates dominated, there is a transition between differentiated asteroids and asteroids that experienced varying degrees of dry metamorphism or dehydration followed by metamorphism. Accreting during this time and in this zone are the chondritic asteroids, the parent bodies of enstatite and ordinary chondrites. In the meteorite collection we can identify multiple grades of thermal metamorphism ranging from meteorites that are only slightly metamorphosed and still retain some volatiles (for example, H3, L3, and LL3 ordinary chondrites), higher degrees of metamorphism (for example, the grade 4 to 6 ordinary chondrites), to meteorites that have undergone varying degrees of partial melting such as the achondrites. These meteorites show solar system abundances of high-temperature minerals and are dominated by silicates and varying amounts of iron-nickel, but are depleted in volatiles. The variations in degree of metamorphism probably reflect variations in the size of the asteroid, the timing of accretion, and the location (Hutchison 2004) (Fig. 23.1).

Further out in the solar system and later in the accretional chronology are the carbonaceous chondrites. These bodies accreted with progressively lower amounts of short-lived radioactive elements and in the presence of progressively larger amounts of water ice. Relatively low volatile-bearing carbonaceous chondrites like the CVs and the COs have bulk mineralogies similar to ordinary chondrites, but also accreted some volatiles and carbonaceous organics along with enough radioactive elements to produce modest thermal metamorphism. Accreting still later and probably farther out in the solar system are the volatile-rich carbonaceous chondrites, the CM, CR and CI carbonaceous chondrites. The original mineralogy of these asteroids included silicates such as abundant olivine and pyroxene, carbonaceous organics, and water

Fig. 23.1 Accretional zones for the major resource asteroid groups in time and distance



ice. They also incorporated modest amounts of radioactive elements, but in this case the radioactive heat input was only enough to melt the accreted ice and drive aqueous metamorphism. The melted ice interacted with the abundant olivine, converting the olivine into the hydrated phyllosilicate serpentine along with a number of other hydrated clay minerals. This had the effect of locking some of the original water into the serpentine crystal structure in the form of hydroxyls. The hydroxyl molecule OH in serpentine is the major resource in the volatile-rich carbonaceous chondrites. The amount of aqueous alteration, the degree of hydration, the degree of metamorphism, and the local chemical conditions varied hugely in these meteorites and thus the parent bodies. This indicates that accretional conditions varied in both time and space for the bodies (Hutchison 2004; Desch et al. 2018).

Comets are the last stage in this accretion continuum. Comets accreted a similar mix of high-temperature silicates (olivine and pyroxene), ices, and carbonaceous organics as did carbonaceous chondrites, but with very low amounts of short-lived radioactive isotopes, since we are now several million years after the start of solar system accretion. As a result, they never underwent significant heating or metamorphism and retained their accretional ices and low-temperature organics. Comets accreted in very cold regions of the solar system since in many cases they retain some proportion of super-volatiles, ices such as CO₂, CO, and CH₄ that have very low sublimation temperatures and thus are not stable in the inner solar system. Comets are probably much richer in complex organics and of course retain a significant portion of volatile ices, mostly water ice. The organics and ices are a significant resource, but cometary orbital characteristics, highly inclined to the ecliptic and eccentric, make rendezvous with comets very expensive in terms of fuel and thus this resource is almost entirely inaccessible for the foreseeable future. Comets that evolve into orbits in the inner solar system devolatize and become extinct, depleted of volatiles and unable to exhibit the characteristic cometary coma. It is likely that some D- or P-type asteroids are actually extinct comets.

The processes that we have described so far—accretion, heating, and subsequent differentiation or metamorphism—occurred within a few million or tens of millions of years at the birth of the solar system. This is what fundamentally determined the mineralogy and basic structure of asteroid resources. Minerals, thus resources, evolve and respond to changing chemical conditions usually driven by heat sources. Mineral evolution and mineral concentration are fundamental factors of economic geology on Earth because of the release of heat from the core and the resulting complex interplay of mantle plumes and plate tectonics, along with the activity of water and oxygen, and the effects of life produce a complex and evolving mineralogy on Earth. There are approximately 5,400 currently identified terrestrial minerals. The corresponding number of minerals on asteroids tops out at about 250 (Hazen et al. 2008). This huge difference in mineralization reflects a huge difference in the nature of terrestrial versus extraterrestrial resources. On Earth ore mineralization is fundamentally about active processes that concentrate economically important minerals into ore bodies. These processes tend to be associated with igneous and aqueous activity that operates over long periods of time; essentially, they scavenge elements and minerals from huge bodies of rock and concentrate them into economically viable ores. Most of these mineralization processes have never been active on asteroids and the few that were lasted only for the first few million years of solar system history (Hazen et al. 2008). As a result, asteroids do not have ore bodies in the same sense as those found on the Earth. From a resource point of view, what are not found on asteroids are “ore bodies”, i.e., concentrations of economically important minerals that are significantly higher than the surrounding bedrock. This is because there are simply no concentration mechanisms beyond the original differentiation and metamorphism processes. In this sense, in small asteroids that are either volatile rich or iron rich, the entire asteroid is the ore body. We do not expect to find significant differences in the concentration of important minerals across the entire body (Hamilton and OSIRIS-Rex Team 2019; Hutchison, 2004).

Included in Table 23.1 are the resource, the meteorite type from which the resource is derived, the likely asteroid class, diagnostic characteristics of that asteroid type that can be observed remotely, and the major resource mineralogy.

To summarize, asteroid compositional groups as detailed in Table 23.1 (thus asteroid resources), accreted and evolved separated in time, location, heliocentric distance, and chemical environment. Seemingly small differences in accretion times resulted in significant differences in available energy from the decay of short-lived radioisotopes, available accretional materials and thus chemistry, and accretional location. Early accreting large asteroids melted and differentiated, producing silicate crusts and mantles while concentrating siderophile elements in an iron-nickel core. Later accreting groups had fewer radioactive elements and more available water-ice and carbon compounds. The accretional locations will have large implications for the next stage of asteroid evolution, planetary migration, and asteroid collisional evolution.

Table 23.1 Asteroid Resources

Resource	Meteorite type	Asteroid class (Bus DeMilio Taxonomy)	Dignostic characteristics	Major mineralogy
<i>Major Early Resources</i>				
Water, organics, silicates	CI, CM, CR, C2 carbonaceous chondrites	Hydrated C Complex, D, Xc	Low albedo	Serpentine, Olivine, Organics
Iron-nickel	Magmatic irons	X, Xk	High radar albedo	Iron-nickel, Troilite
<i>Other Asteroid Resources</i>				
Silicates, iron-nickel	H, LL, L ordinary chondrites,	S, Sq, Sr, O, Q	Olivine and pyroxene absorptions	Olivine, Pyroxene, Iron-nickel
Silicates	Howardites, eucrites, diogenites	V, Sv	Strong pyroxene absorptions	Pryoxene, Basalt
Silicates	CD, CV, CK carbonaceous chondrites	K, L, Anhydrous C Complex	Weak olivine and pyroxene absorptions	Olivine, Pyroxene
Silicates	Enstatite chondrites and aubrites	Xe, Xc	High albedo	Enstatite
Silicates	Brachinites, lodranites	A, Sa	Strong olivine absorptions	Olivine

23.3 Early Asteroid Collisional Evolution

At the same time as asteroids were accreting, planets were forming. The cores of what would become the terrestrial and gas giant planets accreted smaller planetesimals directly, but some of these interactions did not result in accretion. In some cases, the orbit of a planetesimal crossing into the gravity well of the planet would be perturbed by that planet's gravity, either losing or gaining orbital energy and transferring that loss or gain to the perturbing planet. The net effect is generally to radically change the orbit of the small body and slightly change the orbit of the planet. Huge numbers of these interactions can add up, causing significant changes in the orbits of the accreting planets. Numerous studies have suggested that the gas giants migrated significant heliocentric distances during the early solar system (e.g. Morbidelli et al. 2010; Bottke et al. 2012). For example, it is thought that Neptune migrated from about 8 AU (astronomical units, 1 AU = the Earth's average distance from the Sun) out to its current 30 AU orbit.

The effect of migrating planets on asteroid resources was dramatic. The migrations primarily of Jupiter and Saturn had the effect of sweeping the asteroid belt with gravitational interactions, increasing the eccentricity and inclination of objects in the

belt which greatly increased the relative velocities of asteroids. This orbital stirring and perturbations from the gas giants had two major effects that we can see in the asteroid belt today. The first was to greatly deplete the asteroid belt. About 90–95% of the original mass of the asteroid belt was lost due to the perturbations from the gas giants (Bottke et al. 2012). This material was either accreted onto the other planets was thrown out of the solar system and into the Oort cloud. The second effect was to make asteroid collisions much more destructive and much less likely to result in accretion. During this period all but the largest asteroids were reduced to rubble piles. This is the era when many of the early accreting differentiated asteroids lost their silicate mantles and crusts, exposing their iron-rich cores. Given the geochemistry we see in iron meteorites, this probably happened to at least 40–60 parent bodies (Hutchison 2004).

23.4 The Asteroid Belt Today

The structure of today's asteroid belt is plotted in Fig. 23.2 by orbital inclination (the orbit's tilt relative to the plane of the ecliptic) and the orbital semi-major axis (roughly the average heliocentric distance) in AU. Each dot represents the orbital characteristics of one of the first 100,000 numbered asteroids. Several characteristics can be immediately identified from this figure. First the area within 1 astronomical unit of Jupiter has been almost completely cleared of all asteroids. Second, asteroids tend to exist in orbital clumps. For example, there are the Hungarias near 2 AU and between 20° and 30° inclination and the Hildas at 4 AU. These are dynamical groups whose orbits are essentially trapped in a relatively stable dynamical space against perturbations from planets. In the case of the Hungarias the dynamical stability is relative to perturbations from both Jupiter and Mars.

Another feature of the asteroid belt are the Kirkwood gaps which lie at whole number orbital resonances with Jupiter. These are gaps in orbital element space (there are no physical gaps in the belt) where small gravitational interactions with Jupiter result in the asteroids being perturbed out of the gaps. Typically, these gravitational nudges from Jupiter increase the eccentricity of an asteroid's orbit, sending them into planet-crossing orbits with very short lifetimes. Asteroids in the asteroid belt have stable orbits with lifetimes on the order of the age of the solar system, but asteroids moving to planet-crossing orbits will collide with other asteroids, planets, the Sun, or Jupiter in a few tens of millions of years. This process of perturbing asteroidal orbits provides a conveyor belt of material that populates the inventory of small asteroids in near-Earth space. Once perturbed into the inner solar system, an asteroid will orbitally evolve by gravitational interactions with the terrestrial planets. These interactions result in orbital changes that can produce a wide range of outcomes including collision with the Sun, ejection into the Oort cloud at the edge of the solar system, collisions with other small asteroids, collisions with planets, or orbital change that produces a near-circular, low-inclination orbit in near-Earth space. This outcome is very rare but given the astronomically long time periods and

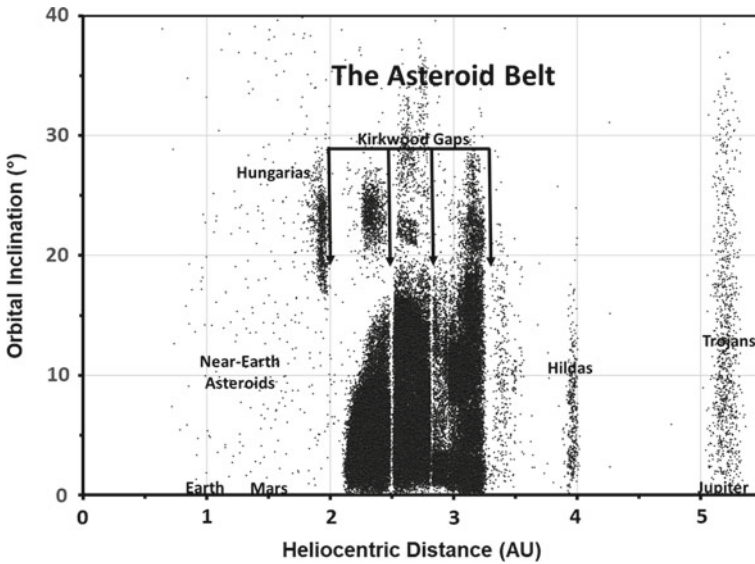


Fig. 23.2 The asteroid belt today

large number of events, near-Earth space is populated with a number of small, but potentially resource-rich asteroids.

For near-term resources, the key population are those small asteroids in Earth-like orbits with long synodic periods relative to the Earth. A synodic period is the time it takes for an object to return to the same location vis-à-vis Earth. Asteroids with orbits near 1 AU and a long synodic period are moving slowly relative to Earth, either slowly catching up or dropping behind the Earth in their orbits. While they are near Earth these objects are in position for relatively low-energy rendezvous for as much as several years. These asteroids are accessible for resource exploitation with long windows of orbital accessibility. Surveys suggest that there are perhaps several hundred small (>100-m diameter) asteroids populating this resource-rich band. It is useful to remember that, as shown in Table 23.1, only a small minority of asteroids, the water-rich hydrated asteroids and the metal-rich fragments asteroid cores, have near-term resource potential. Table 23.2 gives examples of the resource potential for likely resource asteroids from hydrated asteroids and metal-rich asteroids. The actual population of these resource-rich asteroids in near-Earth space has been modeled, but not directly confirmed by observations. Section 23.5 discusses resource assessment strategies for these objects.

What are the physical characteristics of small near-Earth asteroids? Because of the collisional evolution from larger parent bodies in the asteroid belt, small asteroids are overwhelmingly piles of gravitationally bound rubble. Their surfaces are covered with blocky cobbles and boulders, and are very rough on the scale of a few meters. The dusty regolith that is so common on the Moon is almost entirely absent in small asteroids because of their low gravity. Any sort of disturbance in the dust, either from

Table 23.2 Examples of resource potential in asteroids

Meteorite type	Major resource mineralogy	Resource	Possible resource amount
<i>Major Early Resources</i>			
C1, C2 carbonaceous chondrites	Mg-serpentine, organics	Structural OH in the serpentine, H in the organics	18–35 wt. % H ₂ O
CM carbonaceous chondrites	Fe-serpentine, organics	Structural OH in the Serpentine, H in the organics	10–20 wt. % H ₂ O
CR carbonaceous chondrites	Mg-serpentine, organics	Structural OH in the Serpentine, H in the organics	4–8 wt. % H ₂ O
Irons (Major Mineralogy)	Iron, Nickel	Relic asteroidal iron core	70–95 wt. % Fe 5–12 wt. % Ni
Irons (Selected Siderophile Elements)	Disseminated siderophile elements	Relic asteroidal iron core	2–12 ppm Pt 0.15–2 ppm Au 0.05–400 ppm Ge

micrometeorite impact or electrostatic repulsion, will lift the dust off the surface to be swept away by the solar wind. The net effect is to strongly deplete fine-scale dust from the surfaces of small asteroids. The surfaces are also subject to strong solar radiation, extreme heating and cooling with rotation, cosmic rays, solar wind, and micrometeorite bombardment. These energy inputs produce a wide array of damage to asteroid surface material, ranging from thermal fracturing and impact fracturing to radiation-induced spallation and electrostatic charging. Fundamentally the surfaces of small asteroids are going to be heavily modified by their exposure to the space environment.

Exposure to the space environment “space weathers” asteroids. Just as exposure on the surface of the Earth will break rock down into fragments and soil, exposure to space induces a variety of changes to the surface material including changes in albedo, reflectance spectra, and particle size. What does not change is the fundamental mineralogy and geochemistry of the surface, so from the resource point of view space weathering essentially has no effect. It does, however, affect our ability to remotely detect asteroid resources. Weathering-induced changes in albedo, spectral slope, and spectral absorption band strength make it harder to identify the fundamental mineralogy of the object. While this is a challenge for remote identification, knowledge of how materials space weather will allow us to interpret weathered remote data and identify what is the baseline mineralogy of these objects.

We have primarily discussed asteroids as potential extraterrestrial resources, but this focus is because of the relative ease of access for asteroids. Inherently comets have substantial resource potential as they are typically much more volatile rich than asteroids. They contain a substantial portion of frozen volatiles, mostly in the form

of water ice. It is hard to generalize about the volatile content of comets, since observations show that the ice fraction may vary anywhere between 12 and 80%. Comets are composed of very weak and very low-density materials, making them physically attractive targets for mining. However, there are a number of cometary characteristics that make them unlikely to be near-term resource prospects. First, comet orbits tend to be highly inclined and very eccentric, which makes any rendezvous very expensive in terms of energy. Just like asteroids, cometary orbits can evolve from gravitational close encounters with planets and some will eventually achieve lower eccentricity and inclination orbits. However, these sorts of interactions take a great deal of time and comets that have evolved into asteroid-like orbits have usually depleted their volatiles by sublimation and become extinct. Second, active comets in the inner solar system are inherently unstable because their ices will be sublimating vigorously in the relatively warm conditions. This is what produces the characteristic cometary tail but also means that comets in near-Earth space will be surrounded by a cloud of debris. Finally, comets are simply not anywhere near as abundant in near-Earth space as small asteroids.

23.5 Resource Assessment for Near-Earth Asteroids

The final topic is how to identify, assess, and grade resources in near-Earth space. As discussed in Sect. 23.4, asteroid ores are fundamentally different from terrestrial ore bodies in that asteroids do not have the long-running, thermal-driven, aqueous and metamorphic processes that separate and concentrate terrestrial ores. There simply are no separation and concentration mechanisms active on asteroids beyond the initial heating that occurred immediately after accretion 4.5 billion years ago. For the purposes of resources, a small asteroid should be regarded a single ore body. Viable resources, at least in the near term, are volatiles and metal (Table 23.1). These resources will be found in different types of asteroids with different accretional, collisional and evolutionary histories. Metal-rich asteroids will be fragments of the cores of disrupted differentiated asteroids. These come from asteroid parent bodies that formed in the earliest stages of solar nebula accretion and incorporated enough radioactive elements to heat and melt the entire asteroid. Mineral evolution for these bodies ended after the metal-rich cores were formed and cooled. Subsequent collisional evolution disrupted these asteroid parent bodies, stripped off their silicate crusts and mantles, and exposed the metallic cores. The exposed cores were further shattered by collisions producing smaller and smaller fragments. Orbital evolution migrated some of these fragments into the inner solar system where gravitational encounters with the inner planets circularized their orbits and put them in a position to eventually become attractive resource prospects.

Volatile-rich asteroids accreted at least several million years later than the metal-rich asteroids, in different locations and under different chemical conditions in the evolving planetary nebula. They incorporated many times fewer radioactive elements and much more water ice. The low radioactive element content means that they never

heated enough to do more than melt the accreted ice and cause widespread aqueous alteration of silicate minerals. As a result, these objects are rich in alteration minerals that have hydroxyl groups locked into their crystal structure. They also followed a similar collisional and orbital evolution path as the metal-rich asteroids to deliver them into accessible orbits for resource exploitation.

Detecting and characterizing these resource-bearing asteroids in near-Earth space is relatively straightforward with remote sensing, but not without its challenges. Both asteroid groups show relatively featureless reflectance spectra in the visible and near infrared wavelengths, which do not provide diagnostic mineral identification features. This lack of diagnostic features sets them apart from the silicate-rich anhydrous rocky asteroids which are a lower resource priority. The primary diagnostic remote observations are albedo, both visible light albedo and radar albedo. A characteristic of the hydrated carbonaceous chondrite meteorites is their extremely low visible albedo, lower than any other type of meteorite. This is driven by their much higher content of opaque carbon-rich materials including organics which are strongly associated with hydration. The fundamental diagnostic feature of these asteroids is that they are much darker than any other asteroids. In surveying potential targets, asteroids with albedos of 4% or less are very likely hydrated resources. We base this conclusion on the analysis of approximately 70,000 meteorites that have been collected and studied on Earth (Hutchison 2004).

Another potential remote sensing tool are observations of the 3 μm hydroxyl and water absorption bands. These are challenging observations at the best of times since the solar flux at 3 μm is very small and the thermal flux at this wavelength in the inner solar system is significant. The thermal flux needs to be modeled and removed in order to assess the 3 μm bands, but at 1 AU thermal flux absolutely dominates in this wavelength range and makes removal very difficult. Another factor is the small size of the target asteroids which make the reflected light signal coming off the body very weak and hard to detect.

For the metal-rich asteroids, the diagnostic observation is radar albedo. Metallic surfaces tend to have featureless but red sloped reflectance spectra in the visible and near IR along with a moderate visible albedo. What sets this type of resource apart from other asteroids is a strong radar return from the metallic surface. Planetary radar observations will be critical for identifying and characterizing this resource.

The challenges of these observations primarily come from the nature of the resource-bearing asteroids. The targets are very small, often less than 100 m in diameter, and in long synodic periods relative to the Earth. Objects this small are effectively not observable, either by telescopes or by radar, unless they are in close approach to Earth. Small asteroids, particularly small dark asteroids, do not reflect much light, so are often beyond the detection sensitivity of even the largest telescopes unless close to Earth. The strength of radar signals falls off with the inverse fourth power distance, so a small target that does not reflect much radar energy is effectively not observable unless it is within about 0.1 AU of the Earth. Long synodic orbital periods mean that for most of their orbits these potential resource asteroids are too distant for observation and can only be characterized during the few years when their orbits take them close to Earth (Fig. 23.3).

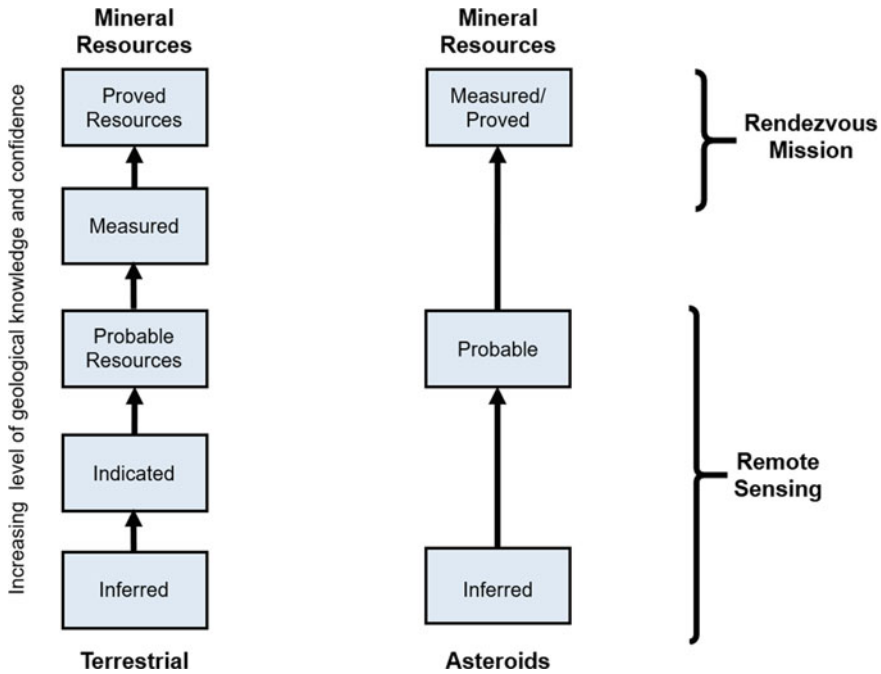


Fig. 23.3 The terrestrial and asteroidal resource assessment process

In terrestrial mining, moving an ore body from an inferred resource to a Proven Reserve is a process of increasing knowledge of the extent, grade, and detailed location of the ore body. This is driven by the fact that terrestrial ores are concentrated in discrete locations by the Earth’s hydrothermal and metamorphic processes. As discussed in Sects. 23.1–23.3, these processes are largely absent from asteroids, so that for small asteroids, asteroidal ore bodies are typically the entire asteroid. Remote sensing can characterize the general mineralogy of the resource asteroid and, with radar remote sensing, provide a very accurate estimation of asteroidal size. For a water-rich asteroid with albedo data and radar imaging we rapidly move through the Indicated Resource and Probable Reserve levels since we know the size of the reserve and its likely mineralogy. The uncertainty that remains is the level of hydration of the asteroid, which can only be measured in situ. Once that measurement has taken place, probably during the early stages of a resource exploitation mission, then the asteroid moves into a Proven Reserve where we have a high confidence estimate of the available resource. In terrestrial resource exploration, a process of geological characterization and analysis produces a steadily increasing knowledge of the location and grade of the ore body represented by a smooth function of increasing probability of the amount of the resource. This is illustrated in Fig. 23.4, where the red line represents the increasing localization of the ore bodies in terrestrial mineral exploration. For asteroid resources this increasing knowledge is more of

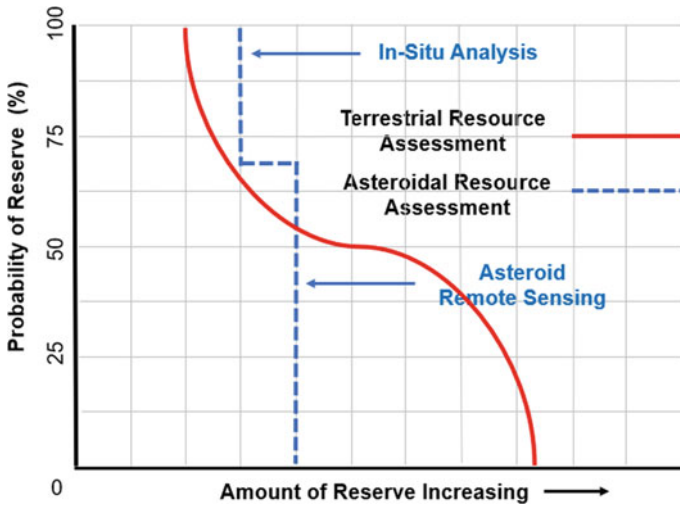


Fig. 23.4 Resource assessment and reserve proving on asteroids. Terrestrial resource assessment is represented by the red line as increasing exploration localizes the ore body. The decreasing estimate of total resource in the terrestrial case is a typical consequence of increasing knowledge. On asteroids the process is a step function where remote sensing identifies water- or metal-rich asteroids and in-situ analysis “proves” the resource

a step function, where the initial remote sensing characterization of resource-rich asteroid puts it into a fairly narrow range of possible resource content. For example, with low-albedo CI and CM carbonaceous chondrites, water content can range from 10 to 30 weight percent (Hutchison 2004). The exact amount of water is indeterminate remotely but provides a baseline range that can be rapidly characterized in situ. The discovery of a low-albedo near-Earth asteroid, along with its size, puts limits on the total amount of the available resource. The first step function moves rapidly from Inferred to Indicated to Probable Reserves through remote identification of the mineralogy and size of the asteroid. The next step function will be a direct measurement of the water content of the asteroid. Since it is very unlikely that the water content will vary significantly over the whole asteroid, the initial measurement is very likely indicative of the resource potential for the whole asteroid. This moves the object from a Probable Reserve to a Proven Reserve, again in a step function.

23.6 Summary and Guidance for Resource Development

What drives economic geology and ore mineralization on Earth is fundamentally different from the processes we find on asteroids. The heating, differentiation, active tectonics, and ore concentration that creates resources on Earth have been absent from asteroids since the first few million years of solar system history. Mineralization

on asteroids has none of the characteristics found in terrestrial ore bodies. Small asteroids are overwhelmingly single mineralogical assemblages and from a resource point of view, an entire small asteroid is the ore body. We know a great deal about the mineralogy of asteroids through the study of over 70,000 recovered meteorites. Any small mineralogical variations in asteroids are generally the result of the accreted debris from the impact of meteoroids. Searching for ore bodies and “high grading” are meaningless concepts on small asteroids since there is likely to be little or no variation in mineralogy across the small asteroid. The other factor to remember is that because of collisional evolution, small asteroids are going to be fragments of much larger bodies that have been repeatedly impacted over the age of the solar system. We should expect that small asteroids are going to be, overwhelmingly, rubble piles.

Comets, while a major potential source of organic carbon and volatiles, will not be an economic resource for the foreseeable future because of the nature of cometary orbits. These tend to be highly inclined to the ecliptic and highly eccentric, so the energy cost of rendezvous and return for a comet will be prohibitively high. The one exception is if the comet has spent enough time in the inner solar system to have had multiple gravitational close encounters with planets that could reduce inclination and circularize the orbit. In that case, it is very likely that the comet has been significantly devolatilized and processed because of the high temperatures in the inner solar system and may, as a result, have lost much of its resource potential.

Because small asteroids are overwhelmingly single mineralogical assemblages, the bulk of asteroid prospecting can be done remotely via telescope and by radar. The major resource potential from asteroids for the foreseeable future will be volatiles and iron. The sources will be hydrated carbonaceous chondrite parent bodies and core fragments of differentiated asteroids. Moving these objects from inferred resources to proven reserves is likely to be a step function since the type and size of the resource can be identified with visible light and radar remote sensing using existing telescopes and facilities. The next major step function is the in-situ measurement of the resource content of the asteroid.

Acknowledgements This work is supported by NASA, the Solar System Exploration Research Virtual Institute (SSERVI), and the Center for Lunar and Asteroid Surface Science (CLASS) under Cooperative Agreement 80NSSC19M0214.

References

- Bottke, W.F., D. Vokrouhlický, D. Minton, D. Nesvorný, A. Morbidelli, R. Brasser, B. Simonson, and H.F. Levison. 2012. An Archaean heavy bombardment from a destabilized extension of the asteroid belt. *Nature* 485 (7396): 78–81.
- Desch, S.J., A. Kalyaan, and C.M.O.’D. Kalyaan. 2018. The effect of Jupiter’s formation on the distribution of refractory elements and inclusions in meteorites. *The Astrophysical Journal* 238: 11. <https://doi.org/10.3847/1538-4365/aad95f>.
- Hamilton V.E. and the OSIRIS-Rex Team. 2019. Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nature Astronomy* 3: 332–340.

- Hazen, R.M., D. Papineau, W. Bleeker, R.T. Downs, J.M. Ferry, T.J. McCoy, D.A. Sverjensky, and H. Yang. 2008. Mineral evolution. *American Mineralogist* 95: 1720.
- Hoashi, M. 1990. The platinum group metals in iron meteorites. PhD Thesis, Massey University.
- Hutchison, R. 2004. *Meteorites: a petrological, chemical, and isotopic synthesis*. Cambridge University Press.
- Morbidelli, A., R. Brasser, R. Gomes, H.F. Levison, and K. Tsiganis. 2010. Evidence from the asteroid belt for a violent past evolution of Jupiter's orbit. *The Astronomical Journal* 140 (5): 1391–1501.
- Papike, J.J. (ed.). 1998. Planetary materials. *Reviews in Mineralogy* 36, Mineralogical Society of America.

Chapter 24

Asteroids: Small Bodies, Big Potential



Akbar D. Whizin

Abstract Harboring both the solar system's ancient past and its bountiful resources, asteroids are perhaps the best-suited targets of opportunity for resource extraction and a source of usable propellants for the build-out of a near-Earth space economy. The use of meteorites as analogs for the abundances the asteroids contain makes a strong case for the potential rewards that deep space exploration and in-situ resource utilization (ISRU) can provide. Until more rendezvous and sampling missions are flown and ground truth can be acquired, prospectors rely on remote sensing and meteorites for knowledge related to asteroid properties and potential reserves. The wide variety of asteroids and their theoretical representation in meteorites indicates a diverse set of potentially utilizable resources and consumables like water, sulfur, metal-rich minerals and alloys. Linkages between the meteorite catalogues and their parent bodies are few and far between, but the sample return of potentially hydrated C-complex asteroid material by Hayabusa2 and OSIRIS-REx could reduce the investor hesitancy that peaked at the end of the 2010s and curtailed commercial mining of asteroids. The framework required to lower mission cost and risk using asteroid resources as fuels and raw materials has become clearer in the past decade, but roadblocks and economic headwinds have led to fluctuations in the pace of progress towards mining asteroids. NASA is currently growing its technology development and maturation programs for the Moon, which could serve as a useful model for asteroid ISRU going forward. NASA will likely need to once again be a leader and trailblazer in this area so that smaller companies do not have to bear the financial burden of a capability build-out that could cost \$10 s–100 s of millions. The synergy between government and private interests continues to increase, benefitting future commercial players by lowering launch costs and providing new vehicles for the support spacecraft demonstration missions. There are many considerations to be discussed as the people of this planet look to expand into deep space.

A. D. Whizin (✉)

Space Science and Engineering Division, Southwest Research Institute, 6220 Culebra Rd, San Antonio, TX 78238, USA

e-mail: awhizin@swri.edu; awhizin@gmail.com

24.1 Asteroids

24.1.1 *Building Blocks of Planets*

From the primordial soup of gas and dust that encircled our young Sun some 4.6 Gya, condensing out from the cloud, micrometer-to-millimeter-sized dust particles accreted into centimeter-to-meter-sized aggregates known as protoplanetals (Goldreich and Ward 1973; Weidenschilling and Cuzzi 1993; Blum and Wurm 2008; Johansen et al. 2008; Morbidelli et al. 2009; Wada et al. 2009). These growing agglomerates formed into larger m-sized to km-sized sized bodies, known as planetesimals, that later went on through collisional accretion to form the planets. The asteroids as we know them are the scattered fragments of that material that never accreted into planets, and as such contain a valuable record of the of the state of materials and the formation regions. Much of the proto-nebula materials to this day reside in the larger asteroids in the inner solar system, Main Belt (MB), spread amongst the Trojans, Centaurs, and the Kuiper Belt or locked deep away in the Oort cloud. Decades of asteroid observations and analyses of meteorites followed by in-depth visits by spacecraft have allowed us to discover a great deal about how and when our solar system formed. The precious metals, minerals, and volatiles that can be found on these bodies will be an extremely valuable resource for humanity as we spread out throughout the solar system and explore the cosmos. Chapter 23, in this publication ‘Resources from Asteroids and Comets’, provides for a more detailed overview of asteroid and comet formation and the source regions for their source materials and resources.

24.1.2 *Orbital Properties and Dynamical Evolution*

Although somewhat radially mixed, orbital zones in the asteroid belt contain a wide variety of different kinds of asteroids, called taxonomic types. Dynamical processes and collisions have scattered groups of asteroids and their source bodies. The fragments of broken-up parent bodies (families) tend to stay grouped into orbits with similar orbital elements (a , e , i , ω , Ω). External forcing moves these fragments into locations in the MB known as gravitational resonances. The 5:2, 7:3, 2:1, and 3:1 mean motion resonances from Jupiter, and the ν_6 secular resonance from Saturn, supply “kicks” from gravitational encounters that amplify the perturbation to the object’s orbit. Successive kicks increase an object’s orbital eccentricity and semi-major axis. Systematic application of these periodic perturbations leads to a clearing of these zones in the MB, called the Kirkwood Gaps.

One way an asteroid in the MB can have its semi-major axis altered, resulting in its movement into a region with a resonance, is known as the “Yarkovsky effect” (Öpik 1951). Diurnal daytime heating of the subsolar side of a rotating asteroid leads to asymmetric re-radiation of the surface in the thermal infrared at night. Photons

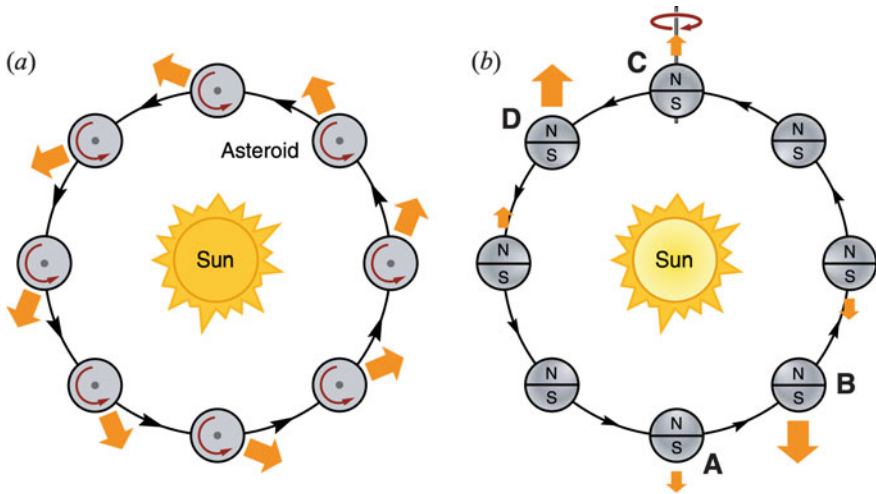


Fig. 24.1 A graphical representation of the diurnal thermal heating and re-radiation of an asteroid and the resulting recoil force that slowly changes its orbit (from Fig. 1, Bottke et al. 2006)

re-radiating off the cooling night-time surface carry momentum away, providing the asteroid with an opposing reaction force (Fig. 24.1). Over time this leads to a change in the asteroid's semi-major axis as it saps (or adds depending on the direction of the asteroid's spin-pole axis) orbital energy and moves inward (or outward) by the conservation of angular momentum. This effect moves asteroids into the resonances, sending them into the inner solar system, and once there can move them continually inward. This is the theorized primary mechanism for delivery of asteroid material into the near-Earth space by the aptly named near-Earth asteroids (NEAs) (Bottke et al. 2000; Nesvorny and Bottke 2004).

These dynamical “pathways” from the MB to the inner solar system are responsible for the steady supply of NEAs to our region of space. They are both a potential bounty and threat to life on Earth. The impact energy of even smaller NEAs in a collision with the Earth is enough to wipe out cities or countries and is quite a threat based on the energies and frequencies of Tunguska-level events ($\sim 10\text{kt}$) (Shoemaker 1983; Brown et al. 2002). The larger $\sim 1+$ km-sized bodies are a global threat to mankind (Alvarez et al. 1980; Vasilyev 1998). While the probability of a collision with the Earth is remote, the consequences of such an impact would be disastrous. In 2005, Congress tasked NASA with detecting 90% of NEAs with a diameter of 140 m or larger (NASA Authorization Act 2005), and directed the funding of research into the deflection of a potentially hazardous asteroid (PHA). This will be demonstrated for the first time with the upcoming Double Asteroid Redirection Test (DART) mission to NEA 65803 Didymos (Chang et al. 2015).

The vast majority of NEAs are not on Earth-crossing orbits, and are of little threat to the Earth; however, they need to be characterized in detail to aid in the understanding of the effectiveness of deflection concepts and technologies. Not only

do these bodies offer us an opportunity to study up close some of the fragments of primordial worlds (as we learn more about our own solar system and how it formed), but we can also more easily assay them for their resource potential as many have opportunistic orbits with respect to the Earth (e.g., 99942 Apophis, 101955 Benu).

24.1.3 Taxonomies

24.1.3.1 Overview

In general, asteroids come in a variety of flavors, or “taxonomies.” Observational properties such as colors and slopes of their spectra are used to classify their taxonomic type, with subtle differences distinguishing one from another (Tholen 1984; Bus and Binzel 2002; DeMeo et al. 2009). The three primary groupings are S, X, and C (originally standing for siliceous or stony, metal, and carbonaceous, but as we shall see, these monikers are overly generalized). These identifiers (from the Tholen classification system, and somewhat of a misnomer) broadly describe the three primary composition groups: metallic/iron-nickel-rich X-types, stony/rocky S-types, and the more pristine carbonaceous and aqueously altered C-types (most C-complexes contain <5% carbon, hence part of the misnomer, but they have very low albedo and do resemble charcoal). There are several sub-groups for each type characterized by distinct trends and differences in their spectral slopes as delineated in recent work by DeMeo et al. (2009) (see Fig. 24.2).

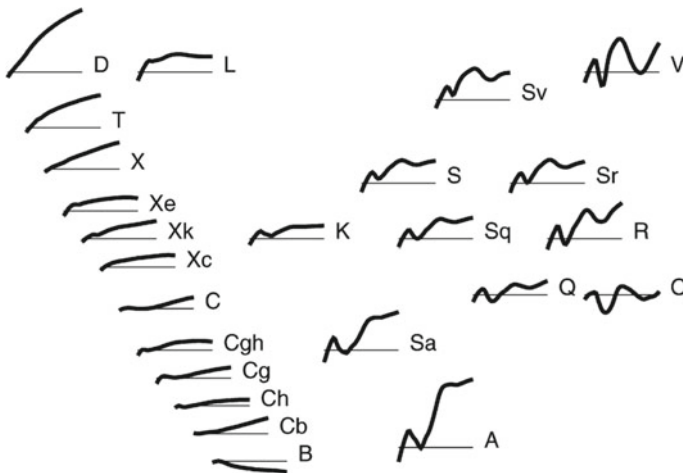


Fig. 24.2 The DeMeo asteroid taxonomy in the infrared (DeMeo et al. 2009, Fig. 15). The average spectra of all 24 classes are included showing the S-group, C-, and X-complexes

The three types are cited in different ways over time depending on the publication date. More refined taxonomic identification has been made and the X-, S-, and C-complex (Bus and Binzel 2002; DeMeo et al. 2009) have now replaced the Tholen classification S-, M-, and C-type grouping naming convention. The X-types, referred to now as the X-complex, refers to a group of asteroid types that includes the Xm-types, associated with metal-rich components. “X-complex” is now used in place of the former grouping moniker “M-type” (but not all X-complex are metallic). Similarly, the C-complex or C-group expands the “C-type” moniker into more refined additional subgroups (not all C-complex are primitive, hence the misnomer and need for the delineation in modern classification systems).

24.1.3.2 The Inner Solar System

Although there are orders of magnitude more MB objects than near-Earth objects (NEOs), the NEOs are much more easily reached by spacecraft and observed up close (NEOs include NEAs and near-Earth comets). We have rendezvoused with 12 small bodies by spacecraft so far (this includes comets 21P/Giacobini-Zinner, 26P/Grigg-Skjellerup, 103P/Hartley, 67P/Churyumov-Gerasimenko, 19P/Borrelly, 9P/Tempel 1, Halley’s Comet, and asteroids 433 Eros, 4179 Toutatis, 25,143 Itokawa, 162,173 Ryugu, and 101955 Benu). An example of a recent rubble-pile NEA imaged up close is Ryugu (Fig. 24.3).

The S-group and C-group asteroids are the most numerous Main Belt asteroids (MBAs), with the X-group asteroids making up around 8–10%. Due to the dynamical processes described in Sect. 24.1.2 acting on MBAs, the taxonomic statistics of NEAs are not 1:1 representative of MBAs. As a result of their source regions in the MB, S-group NEAs are more prevalent than C- or X-complex. This is represented by the meteorite catalogue of witnessed (falls) or found (finds) meteorites on Earth.

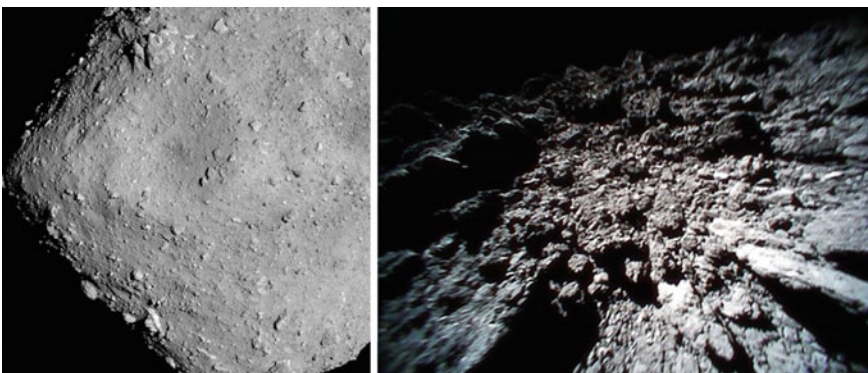


Fig. 24.3 Images of NEA Ryugu’s surface taken by the Hayabusa-2 spacecraft’s ONC-T camera (top), and the surface as seen from the MASCOT Rover-1B (bottom) (JAXA, Universities of Tokyo, Kochi, Rikkyo, Nahoya, Meijijim, Arizona, Chiba Inst. of Tech., and AIST)

Although a selection effect alters the meteorite catalogue that we can recover, the breadth of these samples reveals that many S-group and C-complex asteroids make their way to Earth. This selection effect can be described by two factors: (1) fragments of NEAs that end up striking Earth's atmosphere are generally vaporized prior to reaching the surface so only the strongest/largest types make it through, and (2) the NEAs that make it into the inner solar system to begin with are not a direct representation of the make-up of the MB due to the compositions and type of the families present near the resonances (Bottke et al. 2000). Despite this skew in the catalogue, we can conclude, based on the compositional data from the meteorites we have, that there are numerous types of asteroids that contain plentiful amounts of metals, rare elements, and volatiles such as water, which will be discussed in more detail later.

NEOs litter the near-Earth space, with more than 22,000 discovered so far, nearly all being asteroids (a tiny fraction of them are extinct comet nuclei). The more general term, NEOs, is used where references to objects of any form reside near the Earth. As NEAs represent nearly all of the NEOs, for the sake of simplicity we will use NEA and ignore the cometary component in this chapter (see Chap. 23 for more information regarding their resources). The discovery statistics show that according to their taxonomic classification, the NEAs are predominately X-, S-, and C-complex (Luu and Jewitt 1989; Stuart and Binzel 2004). As the most populous types, these are the primary bodies materially relevant for utilization in space mining or ISRU.

24.1.3.3 The S-Group

This subset contains the stony and stony iron asteroids, predominately metamorphosed silicate rock characterized by a moderate spectral slope with an absorption feature around 1 μm from the silicate components. This grouping includes the A-, K-, L-, Q-, R-, and S-types, as well as the S- "assemblages" Sa-, Sk-, Sl-, Sq-, and Sr-type transition objects. With high amounts of olivine and pyroxene minerals, these bodies appear spectrally redder, due to the alteration effects of space weathering. This group, representing the second-most populous class object in the MB (17%), and the most populous class in the near-Earth region, has been linked as a result of decades of work to ordinary chondrite (OC) meteorites. More specifically, the HED meteorites (howardite, eucrites, and diogenites) were found to be originally from the asteroid/dwarf planet 4 Vesta (McSween et al. 2013). The NASA DAWN mission helped to link the HEDs to Vesta, in much the same way that the Hayabusa-1 mission to asteroid Itokawa, which brought back samples, established the link to the LL-ordinary chondrites (Nakamura et al. 2011).

Each asteroid type is roughly thought to have a corresponding meteorite class. The linking of these classes to asteroid types is an active ongoing area of planetary science. To prove the links, samples are usually required, as the remotely sensed surface spectra are altered from space weathering and impacts. As we can study meteorite samples in laboratories on Earth, and with the ever-improving field of microscopy, more and more information exists in the catalogue to help connect the

terrestrial samples with their asteroid counterparts. The source or parent bodies can then be ascertained. This aids in the ongoing efforts to trace the origins of asteroids and their material source regions. The associated meteorite classes (correctly linked) are vital to better understanding the resources and material properties present on their parent bodies. So far, the LL-ordinary chondrites and HEDs are the only proven links, the OSIRIS-REx and Hayabusa2 missions both successfully obtained samples from two C-complex NEAs (the former is currently en route to Earth to deliver its payload, the latter returned samples in 2020) with hopes of confirming links to carbonaceous chondrites (CCs).

24.1.3.4 The C-Complex

The C-complex consists of the “carbonaceous”-type asteroids, the C-, B-, F-, G-, and D-types. The more pristine (less alteration) and largest grouping of asteroid types (~75%), this group hosts a broad range of members including the largest asteroid in the solar system: 1 Ceres (a G-type, visited by the DAWN spacecraft). These very low-albedo, high-porosity objects are difficult to observe, possibly making surveys and NEA discoveries skewed in favor of the S- and X-group objects. Other notable C-group asteroids includes 2 Pallas (B-type, second largest asteroid), 253 Mathilde (visited by the NEAR spacecraft), and 10 Hygeia (C-type, large family in the MB). These bodies are not only interesting owing to their potentially primitive nature and similarities to icy outer solar system or primordial objects, but also to the potentially large quantities of hydrated minerals that have been detected on their surfaces (Rivkin et al. 2002). C-complex bodies are thought to be connected to the carbonaceous chondrite meteorite class. This would indicate that these bodies started as primordial, and underwent thermal and aqueous alteration. The thermal alteration is shown in the high metal component in the CB and CH classes. The metamorphosis of some minerals due to the presence of water, indicated by the high percentages of phyllosilicates in CI and CMs, results in them being rich in volatiles like H₂O.

NEAs with low perihelion (closest point to the sun in their orbit) are called Atens (a < 1 AU and perihelion is inside Earth’s orbit), and C-groups are conspicuously rare this close to the Sun (Gravnik et al. 2018). This could be due to the enhanced thermal fracturing and weathering these bodies would endure, leading to an early dynamical end (Delbo et al. 2014). This is potentially another reason why we see fewer C-group objects as NEAs, and fewer as meteorites. Ongoing work in this area, including the recent sampling and surface activities by OSIRIS-REx and Hayabusa-2, are revealing the friability and weak strengths of smaller C-group NEAs. Many small asteroids (<1 km) have been found to be rubble-piles, that is, loosely bound bodies consisting of large and small fragments, possibly due to reaccumulating after a disruption event (Sanchez and Scheeres 2014; Walsh 2018). Bennu, Itokawa, and Ryugu are all examples of rubble-pile objects.

24.1.3.5 The X-Complex

The X-complex is an umbrella grouping describing what were previously called “metal” asteroids. It contains the M-, E-, and P-types as well as the associated subgroups Xc-, Xe-, and Xk-type transition objects corresponding to the C-, E- and K-type asteroids (Bus and Binzel et al. 2002). Transition objects more closely resemble the spectra of the corresponding type but contain indicators of the metallic asteroids in the X-group. Some, like the Xc- and Xk-types, match well to iron meteorites, while some match stony irons or enstatite chondrites with some overlap with certain metal-rich CCs (Neeley et al. 2014).

Although these asteroids make up ~8–10% of all asteroids, they are often the most prominently discussed in the context of space mining and resource extraction. This is due to the theorized links to the iron-nickel meteorites that contain, among other things, large weight percentages (wt. %) of iron, nickel, associated alloys, and trace metals that include gold and other platinum group metals. 16 Psyche is the target of the upcoming NASA Psyche mission. An M-type asteroid, 16 Psyche is thought to be a fragment of a shattered planetary core—an intriguing prospect, one that is not entirely supported by the more recent observational campaigns (Landsman et al. 2018; Becker et al. 2020). Headlines describing this body as a trillion-dollar “gold mine” are both outlandish and greatly exaggerated. Market supply and demand aside, there is no economic value in bringing ore to the Earth with today’s launch and mining costs. In the future with lower costs, these resources will be used in situ to build more in-space equipment and infrastructure. It will likely never be cheaper than simply mining the ore on Earth. The NASA Psyche mission will tell us more about this unique object, the first metallic asteroid to be visited by spacecraft, and will shed light on X-complex asteroids’ resource potential. That mission may help link M-type asteroids to certain iron meteorites, which we use as an analog for potentially minable metallic asteroids. By using the correct meteorite as a chemical and material properties analog, we can begin to frame some of the engineering requirements for the eventual extraction and hypothetical profitability of bodies linked to them.

24.1.4 Meteorites as Analogs

24.1.4.1 Overview

The meteorite collections of recovered or found stones are a treasure trove of information regarding our nearest neighbors. Although all but a few meteorites remain to be linked with their parent asteroids and source regions, there are proposed links between the X-, S-, and C-groups. Spectral signatures help link the two, and give us a way to potentially estimate the compositions and properties of asteroids from laboratory measurements taken on meteorite samples (Cloutis et al. 2014). These three groups are the hypothesized parent bodies of the three meteorite classes – ordinary chondrites, carbonaceous chondrites, and iron-nickels. This is, however, a generalization:

there are many sub-groups within each, and the matching is difficult, complicated in part by the incompleteness of the catalogue (Gaffey et al. 2002), and usually requiring direct samples due to the uniqueness and diversity in composition of each. It should be noted that the relative abundances of the minerals do not directly correlate with the apparent abundances obtained from spectral analysis (Adams and McCord 1970). Additionally, there may be missing spectral analogs as well for asteroids (Britt and Lebofsky 1992). By studying each meteorite's compositions and material properties, as well as those of specific minerals found to be present (e.g., olivine and pyroxene) (Reddy et al. 2015), we can make broader extrapolations towards the asteroids we discover or already know we might want to extract resources from.

24.1.4.2 The Ordinary Chondrites

The chondrites (ordinary, carbonaceous, and enstatite) are the large class of stones likely from lightly thermally altered primordial bodies owing to the significant percentage of chondrules in their matrix (chondrules are round mineral assemblages that condensed out of the gas and dust in the early stages of planet formation). Among these the ordinary chondrites (OCs), the dominant type to fall to Earth (80–90%), are relevant to space resources. The parent bodies of OCs contain potentially valuable resources, as they can have moderate free Fe–Ni metal and significant total iron wt. % (25–31% for H-chondrites, 20–25% for L-chondrites, and ~ 20% for LL-chondrites in the form of non-magnetic FeO), and contain useful elements such as Si, S, and Mg. This is owing to the fact that OCs are predominately made up of minerals forsterite (Mg-rich olivine) and bronzite, with some iron-nickel and troilite (Gaffey 1976; Gaffey et al. 1993).

Itokawa (an S-type) was shown to be linked with the LL-chondrites (representing ~2% of all OCs) through analysis of returned samples from the Hayabusa-1 spacecraft (Nakamura et al. 2011). There have been hypothesized links to the parent body of the L- and H-chondrites and others, but no direct link has yet been proved. The importance of linking meteorites to their parent bodies becomes clear, since available quantities of valuable resources can change dramatically even within one subgroup (3% free Ni–Fe instead of 25%, for instance). All this essentially implies the source regions for the statistically common OCs are NEAs and inner MBAs, meaning this group of ore bodies are plentiful and more easily accessible by spacecraft.

24.1.4.3 The Iron-Nickels

The iron alloy-rich iron-nickel meteorites are thought to be the fragments of differentiated planet cores and represent the predominantly metal meteorites. Due to their high metal content, they are hypothesized to originate from the X-group asteroids because spectrally they are the most similar to that group (Burbine et al. 2002; Clark et al. 2004). These are rare, only ~5% of falls (but the largest in size). This is proportionally less than the X-complex's fraction of asteroids as a whole, indicating they

probably disproportionately enter the near-Earth space to become NEAs or meteors and/or are simply found at higher rates due to their strengths. Metallic iron is an Fe–Ni alloy and is the dominant alloy in most, consisting mostly of kamacite and taenite. The parent bodies of these meteorites are likely to be the prime targets for metals mining and manufacturing interests.

The M-, Xc-, and Xk-type asteroids are a potential match for the iron meteorites for surveyed X-complex asteroids (Fornasier et al. 2010; Neely et al. 2014). The dominant chemical elements are Fe, Ni, and a tiny amount of Co. The Fe content makes up more than 70–95%, while the Ni content can range from 5 to 25%, (e.g., octahedrites <18%). They also contain trace elements, including Mo, Ru, Rh, W, Re, Os, Ir, Pt in grains, and likely formed due to the very high temperatures the parent bodies of the iron-nickels were exposed to (Clark et al. 2004; Campbell and Humayun 2005). But not all of the meteorite parent bodies underwent heating and thermal alteration.

24.1.4.4 The Carbonaceous Chondrites

A sub-class of chondrites, and the closest in composition to the solar nebula itself, the carbonaceous chondrites include eight groups that contain more unaltered primitive compositions. In ascending order from 1–6 most aqueously altered to most thermally altered they are: C2/ungrouped, CI1, CM2/3, CR2, CV3, CO3, C3/4-ungrouped, CK3/6, CB6, and CH6. Some CCs did experience unusual formation conditions or heating events (Grimm and McSween 1989) and have large percentages of iron, for example, CB and CHs have been found to have more than 50% Fe (Jarosewich 1990). These attributes also correlate to their porosities, being as high as 35% and as low as 5% (Macke thesis 2010; Macke et al. 2011; Bland et al. 2004; Ralchenko et al. 2014). As a result of little to no thermal processing, the CCs can be hard as stone or crumbly clods, and contain organics, poly-cyclic aromatic hydrocarbons (PAHs), and hydrated serpentine clays (>70% for some CCs). The phyllosilicate matrix-rich CCs indicate the presence of source bodies rich in OH, H₂O, and other volatiles. The minerals trap volatiles in between layers of the crystalline lattice, i.e., absorbed water.

The CCs are hypothesized to originate from the C-complex asteroids, such as C-, B-, or Cg/Ch-types (Rivkin et al. 2015). They characteristically share many similar spectral traits and mineral compositions. CMs, the moderately aqueously altered volatile- and chondrule-rich meteorites are very homogeneous, implying their parent bodies would also be. Therefore, we can use CM chondrites to infer parent body properties with low uncertainties. Samples are needed in order to verify the link between CMs and the hypothesized parent bodies, the Ch-types (Gaffey et al. 2002; Fieber-Beyer et al. 2012; Fornasier et al. 2014), and positively link the other CCs with their potential C-, and B-type parent bodies, in addition to the possible link between D-type asteroids and the C2-ungrouped Tagish lake (Hiroi et al. 2001; Brown et al. 2000). Recent samples successfully acquired from the Hayabusa-2 mission have been returned to Earth (December 2020). Samples from OSIRIS-REx are en route

home and will arrive in 2023. As was the case with Hayabusa-1 samples from NEA Itokawa, both samples will be analyzed to determine the best compositional and spectral match to link to the asteroid type along with ground-truth material and mechanical properties. The end result of this will be terrestrial samples for which we can readily infer most of the intrinsic mechanical and compositional properties of the parent asteroids (porosity, mineralogy, chemical composition, grain density, homogeneity, but not certain bulk properties).

24.1.5 *Physical Characteristics and Surface Environment*

A majority of asteroids are predicted to be rubble-pile structures i.e., composed of gravitationally bound size distribution of smaller components with few, including the largest, as monolithic or cohesive bodies (e.g., Britt et al. 2002; Scheeres et al. 2010; Walsh 2018; Watanabe et al. 2019), with recent spacecraft data continuing to support this (e.g., Michel et al. 2020). Due to this, NEAs will often resemble a peanut or a spinning top (Fig. 24.4). This is a probable result of the break-up and re-accumulation event that body likely endured or of YORP spin-up, the results of which make extremely high porosities and weak surface strengths possible. Recently, Benu's porosity was determined to be ~50% (Hamilton et al. 2019), and Ryugu's 25–50% (Jaumann et al. 2019), compared to larger 4 Vesta's (S-type) <5% porosity (Neumann et al. 2014), strongly suggestive of considerable void fraction and rubble-pile structures. Most asteroids likely have a high degree of macroporosity (Flynn et al. 1999; Britt et al. 2002). On the surface, thermal breakdown and impacts weaken the top layer of regolith and boulders (Delbo et al. 2014; Molaro et al. 2020).

As a result of this overall rubble-pile structure, no cohesive surface truly exists. Instead, NEA surfaces are covered with highly porous, weak rubble, boulders, and regolith sized from 1 mm to tens of meters (Popova et al. 2001; Michikami et al. 2008; Sanchez and Scheeres 2014; Brown et al. 2016). While some meteorites have

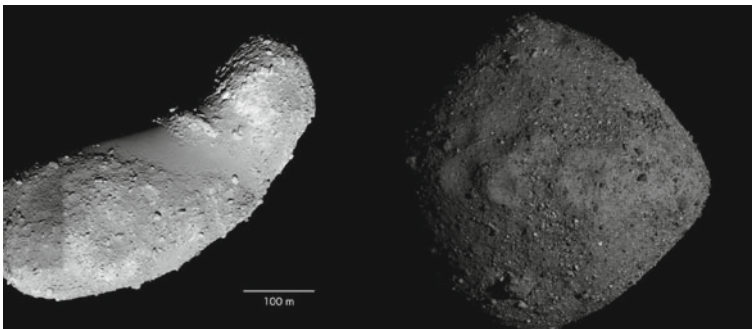


Fig. 24.4 Images of the rubble-pile S-type Itokawa taken by the Hayabusa-1 spacecraft (JAXA 2005), and the B-type Benu by OSIRIS-REx (NASA 2020)

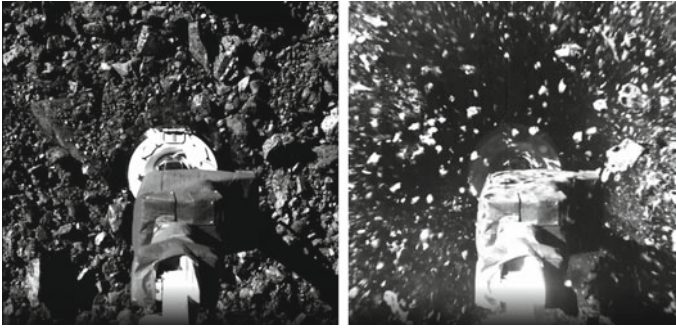


Fig. 24.5 SamCam images of the TAG event during the sample collection at Bennu by OSIRIS-REx (NASA, Oct 2020)

significant compressive strengths, like OCs with strengths ranging from $\sim 6\text{--}374$ MPa (Kimberly and Ramesh 2011), the inferred low strengths of the surface boulders from bolides (meteors that do not reach the surface) can be typically on the order 0.1 Mpa (Brown et al. 2016). This coupled with the low gravity could make surface operations very difficult, as evidenced by the touch-and-go (TAG) sample collection event performed by OSIRIS-REx plunging into the cobble-strewn Nightingale Crater on Bennu (Fig. 24.5). The measured low strengths of hydrated CCs, coupled with thermal and collisional degradation of the surface, implies weakened compressive and shear strengths of the boulders and bulk cobble. Any interaction with the surface likely penetrates to some degree, lofting large quantities of material, complicating anchoring, landing, and excavation operations. All this, coupled with the low gravity ($\mu\text{-g}$), and dusty potentially ejecta-filled near-surface region (Bottke et al. 2020), presents a challenge for con ops and mining operations alike.

24.2 Asteroid Resource Potential

24.2.1 Overview

Grouped as a whole, asteroids possess tremendous (perhaps the greatest of all) resource potential for humanity's ambitions to colonize and explore the solar system. Asteroids represent the elemental, the primordial, bodies nearly untouched since the dawn of time. They contain countless viable consumables and industrial manufacturing materials left over from their formation, while the majority have not been processed under the extreme temperatures and pressures of the terrestrial planets (some asteroids are planetary cores or mantles and thus have been to varying extents), and therefore do not contain certain precious gems and minerals found on the Earth. Even so, asteroids themselves contain a great deal of the necessary materials to build complex spacecraft parts and the resources and volatiles to produce propellants and

other important by-products such as organic compounds, binders, or human consumables like water and oxygen. In general, the types of resources could be grouped for convenience into metals, non-metals, volatiles, and in-situ applications.

The primary usable resources in the near and even medium term are water and metals, probably just Fe and Ni. The OCs and CCs contain mafic minerals that are a majority Si, O, Fe, Mg, Ca, Al, and Na in addition to a long list of trace elements like Cr, Ti, Zn, and Co (Dunn thesis 2008). While this sounds like a wide array of resources, the elements are bound in the metal phases or crystal lattice of the grains. Extracting and separating most of these resources through conventional comminution and ore-processing methods would consume tremendous energy and would prove to be a technical hurdle in low gravity. Despite the long time horizon until they are feasibly mined, it is still useful to discuss them in the same context.

24.2.2 *That's so Metal*

Within the three classes of asteroid analogs (i.e., OCs, CCs, and iron-nickel meteorites), metal alloys, mafic minerals, feldspars, iron-rich magnetite, oxides, and phyllosilicates dominate. Fe and Ni, Fe-oxides, Mg- and Fe-rich mafic minerals like olivine and pyroxene share significant percentages in OCs, and are found to some degree across all OCs and CCs (Burbine et al. 2002). Some of the CCs are heavy with phyllosilicate matrix but contain metals such as Mg, Al, Ca, Ti, and Zn. The CB and CH CCs contain majority iron and metals. The iron meteorites are predominately Ni-Fe alloys, with some small portions of chondritic materials (for a range in iron wt. % see Fig. 24.6). Due to the large amounts of iron present in nearly every meteorite type, nearly every asteroid can be a source of either free Fe, metallic iron alloys, or iron oxides. Over 90% of the following are contained within the metal phase: Ni, Co, Pt, Au, W, Ge, Pd and others. Trace amounts of Au (1.2 ppm), and Pt (8.5 ppm levels) are likely at too low levels to be mined in massive quantities (Nicheporuk and Brown 1965; Mason and Graham 1970).

Much has been made in the media over the presence of gold asteroids worth trillions of dollars, but their existence is no truer than is the claim it could be

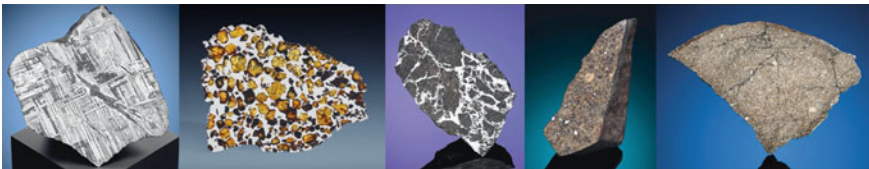


Fig. 24.6 Iron and stony iron meteorites showing the descending quantities of Ni/Fe alloys and iron content. From left to right: Pallasite with significant Widmanstätten pattern, pallasite with significant olivine gems, a silicated octahedrite, L5/6, and L6-ordinary chondrites (meteorite images from Christie's Auctions)

used a refueling stop with a perihelion of 2.5 AU (Carter 2020 “A Bizarre Trillion Dollar Asteroid Worth More than Our Planet is Now Aligned with the Earth and Sun”, Forbes.com). Additionally, metals other than the Ni–Fe alloys do not exist as contiguous deposits within an asteroid (the iron meteorites suggest that the M-type and alloy-rich asteroids, while rare, do exist). Asteroids do not contain veins of metal like they do on the Earth formed from tectonic and igneous processes, so the distribution is different. Nevertheless, they contain plentiful quantities worth mining and potentially containing ore deposits (even platinum group metals), as defined terrestrially, if they could one day be profitably mined. Even asteroids with low iron likely still contain ~ 15% or more by wt. % due to the siderophile elements present. A more useful realistic breakdown of metals in asteroids, coming from iron meteorites (could represent possible compositions of certain X-complex asteroids) would be as follows: Fe 75–95%, Ni 5–25%, Co 0.22%, with <1000 ppm of Cr and Mn.

24.2.3 *That’s Not so Metal*

Primarily in the OC and CCs, mineral assemblages of phyllosilicate, olivine, pyroxene, plagioclase, diopside, sulfate, carbonate (found in CIs as Calcite) exist within the matrix or in clasts, inclusions and chondrules (see OCs in Fig. 24.7). Both contain minerals such as troilite and pyrrhotite (FeS), that contain sulfur, a versatile but corrosive element that is touched on in more detail in Sect. 24.3.5. The most thermally unprocessed CCs contain as much as 5% hydrocarbons and organics, implying the presence of carbon (Buseck and Hua 1993). The hydrocarbons come in poly-aromatic form and include phenanthrene, anthracene, naphthalene, pyrene, and fluoranthene (Zenobi et al. 1989; Wing and Bada 1991). Most of the non-metal components present in meteorites are part of the rock matrix, inclusions, or incorporated as fundamental structures of the minerals themselves, e.g., Si in olivine as $\text{Fe}_2(\text{Mg}_2)\text{SiO}_4$. This makes these potential resources difficult to obtain in pure forms.

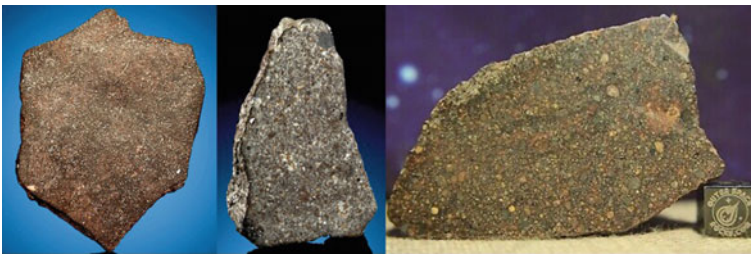


Fig. 24.7 Examples of ordinary chondrites H, L and LL (differing in amount of total iron, High, Low, and Low with Low metals) (images from Heritage Auctions and Outer Space Rock Store)



Fig. 24.8 C2-ungrouped Tagish Lake and CI1 Orgueil, two of the most aqueously altered and hydrated carbonaceous chondrite meteorites known (water wt. % ~15–22%; Jarosewich 1990; Baker et al. 2002) (images from Museum of National History of Paris)

24.2.4 Valuable Volatiles

The CCs are known to be water rich and are the primary target for hydrated mineral mining. CIs and CMs (and C2-ungrouped) contain a large percentage H_2O/OH by weight (see CCs in Fig. 24.8). Across the CCs, correlated with increased aqueous alteration state, peg the average water content at 20.1% for type 1, 13.4% for type 2, and 0.99% for type 3 (Wiik 1956). Alteration states, that is, the degree of alteration, indicate the past history of metamorphosis (petrological type). For CIs this is 1, for CMs this is 2 or 3, for CRs this is 2 or 3 and so on. The CI, CM, and CR (and C2-ungrouped) are the most aqueously altered, and thus contain the most hydrated mineral phases in their matrices. The phyllosilicates (serpentes, smectite, mica, and chlorite) contain hydroxyl- and water-bearing minerals, and are plentiful in CCs (Rubin 1997). The amount of matrix present combined with petrological type also works as a correlative indicator for the amount of hydrated minerals. Other volatiles known to be released during heating experiments are: OH, H, CO_2 , CO, N_2 , O_2 , S, H_2S , SO_2 and trace amounts of other gases (Springmann et al. 2019). The known quantities in case of CCs and their possible C-group asteroid parent bodies (also the most common group among MBAs), make their parent bodies prime targets for exploration and resource scouting due to the sheer total wt. % of volatiles and hydrated minerals.

24.2.5 In-Situ Resourcefulness

There are many uses for the materials present on other worlds, and asteroids are no different. Alternative forms of in-situ resource utilization (ISRU) need to be included

in the discussion of potential resources because these are exploitable and beneficial. There are specific use cases for non-consumables that could aid in the maintenance and protection of long-duration deep-space habitats. Examples of these include shielding, for instance the regolith itself can be used as a radiation and micrometeoroid impact shield with little or no pre-processing. Excavating and covering a structure under a couple of meters of material can protect human occupants from dangerous solar energetic particles (SEPs) and galactic cosmic rays (GCRs). Before any discussion on the application of the potential resources or their abundances can occur, it is useful to discuss the search for, confirmation, and extraction of these resources as the next section attempts to do. See Chap. 22 for a more in-depth discussion on non-consumable uses of asteroids, including as habitats.

24.3 Prospecting and Extraction

24.3.1 Overview

The principal ISRU categories of asteroid materials can be described as: scouting (remote sensing), prospecting (sampling), excavation (removal), extraction (separation), synthesis (processing), and storage (preservation). Generally, these progress in series and have many similarities to mining and prospecting activities on Earth. Care must be taken to not confuse the two, however. Terrestrial mining practices and methodologies do not translate to the asteroid regime. There is a natural tug-of-war between engineers and industry mining players in the ISRU/space resources community regarding the best excavation practices and ways to assess the “ore” body. While these traditional methods can be adapted (Keszthelyi et al. 2017 USGS Asteroid Resources Assessment), asteroids do not contain ore of the kind we know on Earth (see Chap. 34 for more on this). As scientists have been studying asteroids for decades in great detail, including sending spacecraft to visit and sample them, we have a unique insight into the composition and mechanical state of asteroids, and the physics of the low-gravity regime in which they lie. This is an entirely different engineering problem than anything we face on Earth (Gertsch et al. 1997), and will require new approaches, which terrestrial miner financiers struggle to justify due to costs and development times. It is therefore prudent to examine the problem from this point of view.

For decades the spectroscopic surveys of MBAs and NEAs have allowed for detailed taxonomic classifications, which coupled with the study of meteorite properties lets us assess the resource potential of entire types of asteroids and certain source regions. Asteroid surface reflectance spectroscopy and other remote sensing techniques reveal much of a body’s properties and composition. Space weathering (irradiation and impact shock darkening of the surface) and low observational magnitudes complicate these assessments, but much of what is needed to assess the potential

can be done by remote sensing. Chapter 23 in this publication ('Resources from Asteroids and Comets', by Britt and Cannon) covers well the remote sensing (scouting) and assessment of a body's resources, and the reader is encouraged to review that chapter for a useful insight in the context of this section as it sets the stage well for this discussion. This chapter primarily focuses on the act of sampling the asteroid, the extraction of the raw materials, and their processing into more usable forms. The storage and the unmentioned final step of the economy (a critical step involving regulations, laws, commercial entities and their business cases) are outside the scope of this discussion, but are discussed in Chaps. 30–34.

24.3.2 *Sampling and Ground Truth*

In order to verify the presence of and assess the precise physical state of an asteroid's resources, direct sampling and material characterization are needed. Until then, remote sensing can determine which asteroids are worth targeting. Scientifically, this is sufficient for a great many precursor surveys and resource assessments; however it has not been enough to financially incentivize investors so far, and likely will not be in the near future. For now, we can use meteorites as analogs, but they have been weathered by Earth's atmosphere, as have asteroid surfaces by space weathering. A meteorite likely represents the core of some boulder or other piece liberated from an asteroid at some point in the distant past. The asteroid's surface cannot be expected to exist in the exact same state due to ongoing surface alteration processes.

Meteorites only offer us a piece of the puzzle: less than 10% of meteorites even make it to the surface before being vaporized. The strengths implied by analysis of this (Brown et al. 2016) can be used to represent certain sub-surface properties, or those of large constituents (outcrops or monolithic components). As most of the small NEAs that have been observed or visited are rubble piles, care must be taken in applying meteorite properties to the possibly weak, porous, thermally degraded, and cobble-strewn surface. Recently, the Hayabusa2 and OSIRIS-Rex missions that sampled material from Ryugu and Bennu highlighted this fact. The projectile sampler and the TAG sampling arm interacted violently or pushed slightly further into the surface than anticipated and caused large ejections of cobble and rock (Fig. 24.5).

The rubble-pile asteroids' formation mechanisms, together with our knowledge of meteorite properties, imply that resources are distributed more homogeneously. This implies deep drilling or digging operations will likely not be as necessary on asteroids. While volatiles may have been removed from surface layers due to thermal cycles (can penetrate some centimeters) or other processes, it likely will not be necessary to go much further than the sub-surface (~meters) for the resources. Added to this, the weakened surface and fractured boulders may make it easier to excavate the desired material. The flip side, however, is that the weakened and boulder field-covered surface will be low in compressive and shear strengths (100 s–1000 s Pa) (Scheeres et al. 2010; Sanchez and Scheeres 2014), and thus make landing/anchoring much more challenging.

One method for obtaining surface and sub-surface properties and resource assessments is a penetrator, a simple concept proposed often, but more rarely flown due to its risky and technically challenging nature. Successful examples of this have been Viking (Mars), Huygens (Titan), and Philae(67P). Since landing on most small asteroids is too risky, a fired penetrator is an attractive concept as it could implant itself meters into the surface and take measurements of volatile or metal contents using mass spectrometers or residual gas analyzers. This type of measurement would provide the definitive confirmation of the resource, and simultaneously information about its state and the surrounding rock properties needed for excavation. A penetrator would give a great deal of information about the subsurface, but may not be a requirement. Even simple surface sampling and landed devices could suffice, such as the Clockwork Starfish experiment (Fig. 24.9) in development at Southwest Research Institute (Parker et al. 2021) under a Flight Opportunities Program grant (Carver 2020, ‘Asteroid Sampling Technology Tested on Blue Origin’s Suborbital Rocket’). While it can be argued that remote sensing is sufficient for resource assessment, the ground truth of the asteroid’s resources is required for commercial and industrial players to reduce the financial risk enough to invest in a mining expedition. There is no precedent for this type of mining, and the typical norms and rules for prospecting and establishment of mines here on Earth have translated to a lack of investor confidence to take the (literal) leap to mine asteroids. The samples from the potentially water-rich asteroids Ryugu and Bennu returned in the early 2020s may help bridge this divide and provide interested parties with a sufficient incentive to begin a prospecting expedition and develop the necessary spacecraft and technology.

24.3.3 Characterization of Resources

Direct ground-truth sampling will allow for a characterization of the quantities and state of the resource. We have only our analogs (and of course a few small samples from sample return missions) that describe the water or iron wt. %; with a direct measurement we can estimate the quantities of resources present and their state. This will enable the excavation and processing at potential mining sites to be optimized, and the best possible and lowest risk site to be selected, depending on the distribution of the resources. If homogeneous, this distribution will be less important than the surface character and the asteroid’s properties which will limit the ability of a craft to interact with the surface. Engineering an excavator around the strengths of meteorites may prove to be incorrect since the bulk properties like porosity and strength differ significantly. The quality and quantity of the resource are also required data. While we may already have a good understanding of the water content or potential metal by wt. %, verification of the surrounding rock composition is essential. This will come into play more as we develop the chemical and minerals models surrounding the extraction and processing of the raw materials. Liberating the volatiles through thermal processing and release, for instance, will mean the simultaneous release of corrosive and potentially problematic gases. Hydrogen sulfide and sulfur dioxide will



Fig. 24.9 The Clockwork Starfish asteroid surface sampling technology demonstration before (a), during (b–e) and after (f) a microgravity flight aboard Blue Origin’s New Shepard (NS-13) suborbital flight (Images courtesy project PI, Dr Daniel Durda)

be present, according to analog studies on meteorites, as will mercury and arsenic (see Sect. 24.3.5).

It is not known whether surrounding rock would contain exogenic material, perhaps from a collision during the asteroid’s past. Exogenic material has been hypothesized for the dark material on 4 Vesta, and as an explanation for some material seen on the surface of Ryugu; however these materials comprise a very small wt. %. Understanding the properties of the surface and sub-surface, i.e., rock and bulk properties, is essential to development of mining tools and processes. The bulk porosity of the asteroid, which was found to be ~50% for Bennu, will place constraints on drilling and probing of the sub-surface, as the body will react commensurately and unpredictably to removal and movement of the asteroid’s material. The strengths of the rock and its electromagnetic properties will affect the design of drill bits or excavation tools. The penetrability of the potentially loose cobble and bulk surface material will affect the landing, or anchoring to the surface. All of this is complicated further by the gravitational environment.

24.3.4 Excavation in Microgravity

24.3.4.1 Low-Gravity Interactions

Ryugu and Bennu, targeted by Hayabusa2 and OSIRIS-REx, respectively, were excellent choices to demonstrate low-gravity operations and sampling maneuvers during the rendezvous in 2018 around small rubble-pile NEAs. The low surface strength seen by these spacecraft during their sampling and surface operations in 2019 and 2020 highlights the difficulties faced by a mining and processing craft at an NEA. Asteroids like these possess such a low gravitational field (10^{-6} m/s^2), with an escape velocity in cm s^{-1} ; it is very hard to land on the asteroid in the conventional sense. Even the smallest forces cause bouncing and the ejection of cobble from the surface (Fig. 24.5). The navigation and con ops teams were extremely successful planning and executing the hover and touch-n-go maneuvers in an environment with so many hazards and unknowns. This can act as a blueprint for future con ops and mission planning. The projectile event of Hayabusa2's sampling maneuver caused the violent release of tens of thousands of particles and rocks, whilst even the more gentle approach of OSIRIS-REx caused very similar behavior.

24.3.4.2 Surface Interactions and Anchoring

The asteroid's spin rate can make interacting with the surface difficult, but the two sample return missions were able to do this quite well. However, they did not land or have prolonged contact with the surface. That will require new techniques, anchoring, or special lander design. Anchoring to the surface will not be trivial, since the loose agglomerated cobble and boulders are thought to be held by cohesion and low gravity. Anchoring to a boulder, for instance, may only work up to the limit of its cohesive force with the surface and not provide a suitable anchor point. Harpoon-based anchors will likely not be a viable solution. This is evidenced by two events: Philae's failed touchdown at comet 67P/Rosetta mission, and the touch-n-go at Bennu that shows how loose asteroid boulder and cobble covered surfaces would react very poorly to being disturbed by an anchor and the outcomes might be little different. This is why techniques such as magnetic anchoring and larger nets or attachment devices, or bagging techniques, are being explored. The presence of magnetite and free iron in asteroids makes the material ferromagnetic. When a magnetic device is placed near the regolith, the induced magnetization in the ferromagnetic material entrains large amounts of surrounding material with it, which is a highly efficient way of picking up, collecting, moving, and/or otherwise manipulating asteroidal material (Fig. 24.9). Attaching a spacecraft or excavation probe to the surface is necessary because of the low gravity. Due to the laws of physics, the almost negligible gravity means a probe or craft will simply push itself away from the surface, since without any normal force on the surface there is no leverage.

24.3.4.3 Drilling and Excavating

In conventional scooping or drilling, the surface or rock strength in addition to gravity provides adequate normal forcing on the scoop or drilling apparatus. In the case of a drill bit, this downward force required is known as the weight on bit. The drill bit “bites” and makes it way deeper through mechanical leverage on each turn. On an asteroid, specifically smaller ones, this is just not possible. One solution to this is the use of “low weight on bit” drills that require little to no external reactive forces, such as the reciprocating WASP drill. This drill operates by reciprocating opposing jackets of the drill bit’s sides, while moving the interior bit deeper with each stroke (Vincent and King 1995).

Another solution called the Apis™ Flight System, proposed and developed by TransAstra, an asteroid mining company, could be a solution to many of the engineering problems surrounding asteroid resource extraction (e.g., some that have been discussed here). In their design, a spacecraft “bags” an entire asteroid (likely only feasible for exceptionally small asteroids only tens of meters across). They have pioneered an approach called Optical Mining™, where a collimated beam of sunlight spalls and vaporizes a carbonaceous asteroid’s surface, as is shown in Fig. 24.10 (Sercel et al. 2016; Dreyer et al. 2016; Sercel et al. 2018). The novel bag approach allows for zero physical interaction with the asteroid while an intense beam of collimated sunlight heats and vaporizes the asteroid’s surface, liberating the volatiles onto collector plates inside the bag (Sercel et al. 2018; TransAstra Corp. <https://www.transastracorp.com/apis-spacecraft>).

Mining on NEAs will not be like it is on Earth. Novel concepts are required, as it will be neither financially nor physically effective to send the massive high-mass augers and other large-scale terrestrial mining equipment to an asteroid. There are also secondary considerations, such as what happens to the waste, or slag deposits. What effect do any ejected particles have on the mining craft or mother ship? Do these operations alter the asteroid’s orbit, and will the same techniques for a C-group asteroid be appropriate for a potentially stronger X-group metal ore body? Likely it will not, and this range of possible asteroid types, configurations, or the vast distances between objects in space will require, at least at first, a much more case-by-case approach. The chemical and mineralogical difference will present a host of other problems for the next phase of ISRU – the liberation and processing of the desired resources from the target body.

24.3.5 Separation of Volatiles and Contamination

The extraction of the metals and volatile elements of an asteroid can be as basic as moving preprocessed material into a bake-out chamber and thermally forcing the labile elemental release of the volatiles (for CCs). Laboratory studies conducted on CCs, for example, have shown that when subjected to heating and pyrolysis, meteorites released a wide range of gases, mostly volatiles with some trace elements

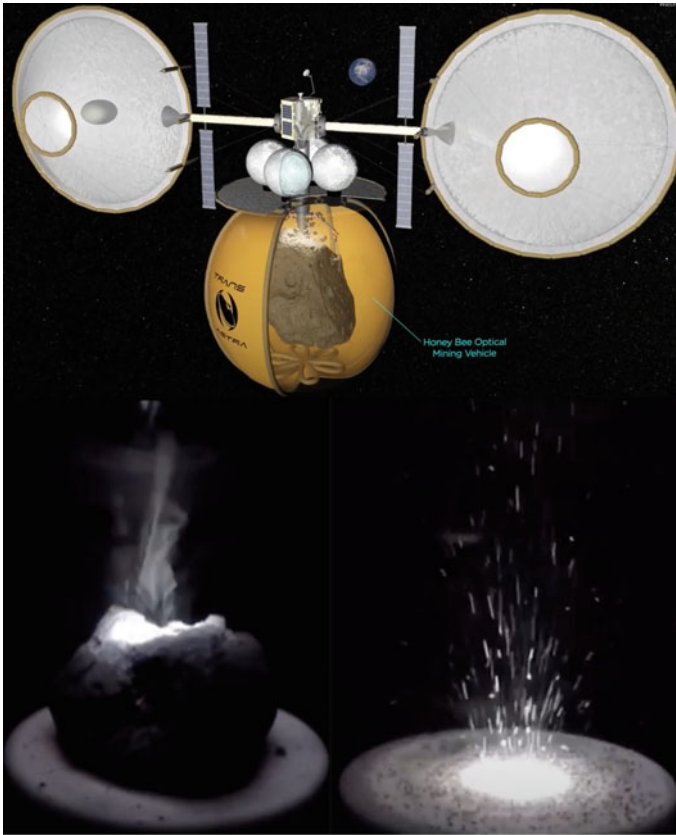


Fig. 24.10 TransAstra Corporation’s HoneyBee™ asteroid mining concept. Optical Mining™ vaporizes volatiles while disintegrating the rock and collecting water in the storage pods (Dreyer et al. 2016; Sercel et al. 2016; <https://www.transastracorp.com/apis-spacecraft>)

(Springmann et al. 2019). The primary release is $\text{H}_2\text{O}/\text{OH}$ starting at $250\text{ }^\circ\text{C}$, with the release curve extending to a peak at $\sim 700\text{ }^\circ\text{C}$. The secondary gases are SO_2 , CO_2 , chain hydrocarbons, aromatic hydrocarbons with the main trace elements S, Zn, Pb, and Hg. Surprisingly, liberated sulfur was not the only difficult element released during heating experiments (Fig. 24.11). Toxic and contaminating elements such as Hg, As, Se, Cd, Sb, and Te were also released in trace quantities.

The release of sulfur, which rapidly acquires two hydrogens post release, becoming hydrogen sulfide (H_2S), is a big problem for the separation and containment/processing of volatile resources. As a highly corrosive substance, eating away at metals, tank seals, and valves, H_2S and other corrosives will require advanced filtration and separation methods. The fossil-fuel industry has grappled with this by using electrochemical membrane separators, a molten process where gas is passed by the cathode in a cell (Alexander and Winnick 1994). The sulfide ions act as a

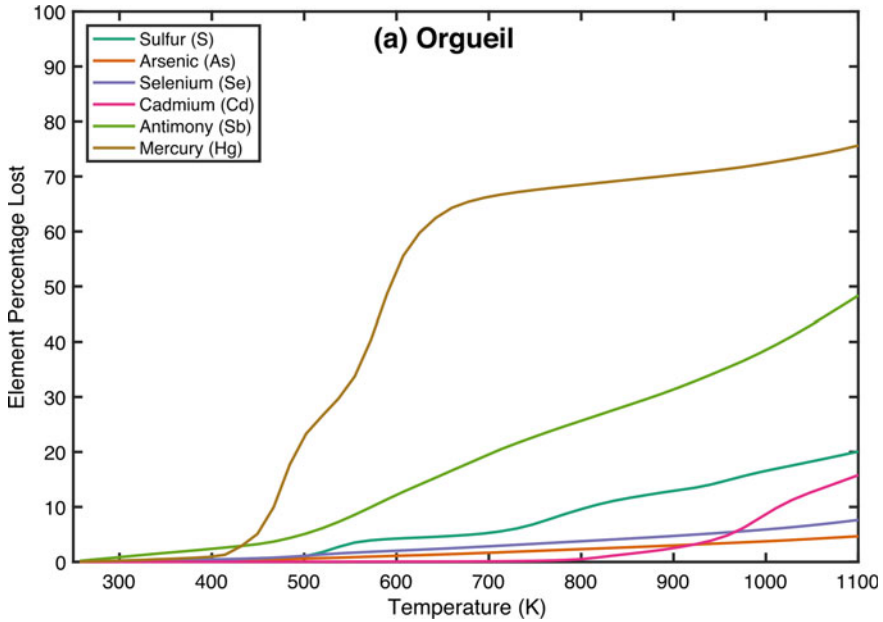
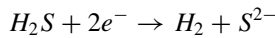


Fig. 24.11 Elemental volatile loss from heating experiments performed on CC Orgueil showing the percentage released compared to the measured total abundances of each (Fig. 2a from Springmann et al. 2019)

membrane and transport sulfide across the anode and the result is oxidized vaporous S₂ and H₂ in the reaction:



In the case of electrochemical membranes optimized for asteroid compositions, pre-filtering of the extracted resources for asteroids prior to storage or electrolysis could be done. Other more exotic catalyzing or purifying techniques will need to be explored for contaminant removal in ISRU processes. Post contaminant removal, the purity of an extracted water product will be constrained by the need for electrolyzing processes to create O and H. The purity required for electrolysis in a cell can be as high as 99%, but the contamination problem is compounded by the fact that some of these volatiles mixed in will corrode the catalyst electrodes in the cell, degrading efficiency and cell lifetime as well as affecting the quantity and state of the final products. Contaminant removal is an often neglected area of R&D in ISRU. The spawning of innovations through R&D, the use of heritage industrial and mining filtration methods, and integration of the phases of ISRU into a more complete end-to-end demonstration will all be needed to refine the contaminant removal processes.

24.4 Resource Utilization and Bootstrapping

24.4.1 Overview

The use of water to create propellant in space is regarded as the primary space resource, often cited as a gateway to the exploration and colonization of near-Earth space. As a starting point, the establishment of near-space refueling stations and propellant depots will enable the expansion of space-based assets in Earth or lunar orbit, into NEA space, Martian space, and beyond (Metzger et al. 2013). This occurs because refueling in orbit lowers the mass at launch, lowering the necessary launch propellant mass exponentially. Owing to the deep gravity well of the Earth, ~96% of launch mass has to be fuel and extra structure to accommodate the propellants (the rocket equation problem). For example, SpaceX's Falcon 9 Block 5 reusable rocket can ferry 22,800 kg to orbit with a base rocket mass of 549,054 kg (i.e., 4% is payload mass in a simplistic calculation). For SpaceX Falcon 9 vehicle overview, see <https://www.spacex.com/vehicles/falcon-9/>. In-space refueling leads to lower costs and opens up commercial and government markets. But it all starts with processed H and O, i.e., water. That is in part why so much emphasis is placed, not just in this chapter, but generally, on carbonaceous chondrite meteorites and C-complex asteroids. Besides lunar water at the poles, the CC meteorite's NEA parent bodies are the next nearest source of usable water, but with an even lower delta-v required than the Moon. The high levels of hydrated minerals present as bound water, when liberated and filtered, would serve as rocket fuels by electrolysis, splitting the oxygen and hydrogen into liquid Hydrolox propellant. Of course, cryogenic storage would be required, which adds another layer of complexity in the supply chain. The linking of the volatile and resource-rich CC meteorites to their parent bodies is therefore one of the next critical science goals for enabling the selection of specific NEA targets for extraction and processing.

24.4.2 *The Space Surrounding ISRU*

The fundamental resource is not gold or iron, but water. Perhaps the most versatile resource as well, capable of being used as a fuel, electricity source, lubricant, coolant, radiation shield, for life support systems and crop consumable, as well as for human consumption. While metals are critical for spacecraft and settlement construction, the kind of viable techniques for manufacturing and processing/smelting centers required for this are still presently nowhere near reality. Propellant production is the far more realistic near-term goal, and a solution to some of the roadblocks currently facing the expansion of efforts in space.

Why should we mine water in space? There is a myriad of problems for which this would create solutions. We as humans are stuck on this planet, bound in a deep gravitational well. Without exponentially adding more mass to rockets and spacecraft,

our exploration of deep space will be limited. The additional mass of more propellant to reach distant bodies, or to bring mining equipment, is limited by the Tsiolkovsky rocket equation:

$$\Delta v = v_e \ln m_0/m_f$$

The change in velocity of the rocket Δv is equal to the exponential decay from the initial (wet) mass m_0 to the final (dry) mass m_f through the use of the propellant expelled at the rate of v_e . This essentially means that at an engine's specific impulse, the rocket will require exponentially more fuel to reach orbit the larger m_0 is. The added burden of larger rockets with more engines and more fuel tanks just increases launch mass even more and, in the end, is limited by the chemical energy density of the fuel and number of engines. Not only is this a tremendous burden on launch costs, but it also places constraints on the payloads themselves. A large-scale asteroid mining operation would require specialized spacecraft loaded with enough fuel to reach its target (an NEA), leading to a large launch vehicle. The solution for accessible near-earth space is to simply launch tanks and spacecraft empty, refueling them in orbit with propellant made from water mined from asteroids (or potentially the Moon). Once such a system is in place, it would lower the cost and mission risk considerably by reducing the size of the rocket required and the amount of fuel.

In the New Space era, the number of technology and aerospace companies entering the fray grows each year, with over 50 companies currently. Most are low-revenue/low-profit startup companies and cannot afford the exorbitant launch costs through traditional providers like United Launch Alliance (the joint Boeing–Lockheed Martin venture). Companies are now able to use ride-sharing agreements for shared launch costs, generally for smaller-sized spacecraft and satellites. This is helping offset some of the cost burden, while the focus still remains heavily on technology development and creating a marketplace for the technology and the expertise driving the New Space era. Enter SpaceX, which has been rapidly lowering launch costs by many factors throughout the 2010s and early 2020s with reusable self-landing Falcon rockets and the new Starship concept being rapidly developed in Boca Chica, TX. Heralding a new era of modern rocketry, their aim is to reduce launch costs by a factor of 10. This will couple brilliantly with the push for larger refuellable in-space vehicles and the build-out of a near-Earth space ecosystem. Interestingly, the CEO of SpaceX, Elon Musk, has stated his desire to see SpaceX land on Mars and refuel the Starship vehicle using ISRU of harvested Martian water and Sabatier production of methane from H_2 (electrolyzed water) and atmospheric CO_2 (Musk and New Space 2017).

The government has weighed in heavily in recent years with encouragement and funding with new frameworks for commercial space. This manifested in a huge way in 2020 with the first ever commercial crew mission conducted by SpaceX and NASA in 2020. It is becoming apparent that the government is more rapidly moving towards allowing the commercial sector to provide the launch capabilities for space, while focusing (mostly through NASA) on the engineering and science as well as on the

astronauts themselves. The ever-changing US political landscape complicates larger-scale ambitions such as the Artemis Moon missions, missions to asteroids (canceled Asteroid Retrieval Mission), and the implantation of ISRU into the current road map for NASA's Moon-to-Mars directive.

NASA and the commercial space environment will benefit greatly if NASA can operate as it once did, leading the way and providing the technological gap-filling progress that is too risky or expensive for smaller commercial entities to attempt. With NASA leading and taking the initial risk posture, the private sector can benefit immensely as they did during the Apollo, Shuttle, and ISS eras. Taxpayers have much deeper wallets than smaller innovative new space companies and shouldering this burden will aid rapid progress. NASA has started towards maturing ISRU technology by first demonstrating it on the Mars Perseverance rover via the MOXIE experiment, already successfully capturing atmospheric CO₂ and converting it to oxygen (Hinterman and Hoffman 2020). On the Moon, this is taking the form of the VIPER lunar south pole mission currently funded and slated for the 2020s, a small first step in this arena to prospect for usable water at the lunar south pole. Ideally these will lead to a dedicated mission to establish a pilot ISRU processing plant on the Moon later in the decade.

The NASA Commercial Lunar Payload Services (CLPS) partnerships to land science payloads at several landing sites in the early 2020s intends to offer a useful vehicle for establishing low-cost commercial partnerships for the development and establishment of lunar ISRU technology. Critically, no asteroid analog to VIPER exists yet (an asteroid harvester concept is being matured by TransAstra under a NICA Phase III grant), and as NASA is one of the only players with experience rendezvousing with asteroids (this now includes the Japanese Space Agency JAXA and the European Space Agency ESA), the commercial sector needs a similar framework built for lunar-focused science and ISRU in order to reduce the risk threshold enough to encourage active asteroid exploration.

To sustain long-term deep-space missions and off-world outposts, ISRU is at the core of human presence sustainability and lowering mission risk. The refueling and resupply of water and propellants in space will lead to lower life-cycle costs for equipment and mission costs. The initial use of mined propellants as a solution to the problems facing deep-space exploration and launch costs will allow for an expansion of the mining of resources to metals and other usable materials. Asteroids can play a key role in shaping a space-based propellant resupply architecture and are on the pathway towards humankind pushing outward to Mars and beyond.

24.4.3 Asteroids as a Foundation: Stepping Stones and Expansion

24.4.3.1 An Enduring Resolve

With the focus at NASA and with federal funding having recently shifted away from big asteroid missions towards the Moon in the latter half of the 2010s, some uncertainty clouds the near-term vision and the broader direction, with constant partisan wrangling. There does seem to be a consensus, however, regarding the eventual exploration and possible utilization of asteroids. Those in the space resources community have a much more bullish stance, as is evident even from the many conference abstracts at meetings like The Space Resources Roundtable or ASCE Earth and Space Engineering Conferences, or by the several asteroid mining companies still in existence today.

The entire pathway for the ISRU of asteroid resources encompasses far more than is possible to discuss in this chapter. Still under consideration are issues such as the selection of optimum resource-rich NEAs from the observable spectra, the development of harvesting robotic spacecraft, the engineering and financial trade studies that will constrain selection and market timing further, the processing and filtration of the captured resources, and the development of in-orbit refueling or establishment of propellant depots to name just a few (to say nothing of the legal frameworks). Despite early set-backs in this field, resolute companies and researchers are tackling every aspect of the asteroid resource supply chain problem. By staying lean, only handling a limited number of the steps to mining, and acquiring renewable funding streams, many are lowering their risks. However, while the risks may be lower, the chances of success are commensurately lower as well, as many of the smaller entities will never reach scale or launch at all. Sometimes, fortune favors the bold.

24.4.3.2 Early Players

Probably the first notable enterprise was the groundbreaking Space Development Corporation (SpaceDev) founded by Jim Benson in 1996, with their well-conceived Near-Earth Asteroid Prospector (NEAP) mission (Benson 1998). SpaceDev represented the first serious effort towards a commercial exploration mission. The mixture of players including the likes of NASA's Dan Goldin, *Mining the Sky's* Jon Lewis (Lewis 1996), and a host of talented scientists and engineers turned a privately backed start-up into a publicly traded company through a reverse merger during the dot-com era. A combination of solicited instrument payloads, in theory to be funded by NASA, and private backing would enable NEAP to stake a claim on one of the hundreds of potentially resource-rich NEAs. Benson's insistence that the missions to NEAs be funded through the more traditional equity-financed route rather than government funded led to a greater reliance on private funding. There was also a greater than

realized reliance on the potential selection of NASA-funded instrument payloads that would have brought revenues of \$10–12 million each. The lack of a selection of any of the proposed instruments dealt a serious blow to SpaceDev's ambitions. Crafty venture maneuverings, including mergers and acquisitions, and the additional revenue from Department of Defense and NASA contracts kept SpaceDev afloat for several more years but the deep-space efforts were tabled. Although he had stepped down as SpaceDev's CEO to pivot to the *Dream Chaser* utilizing hybrid engines SpaceDev had pioneered, Benson's untimely death stymied some of the early asteroid mining endeavours. SpaceDev was acquired by Sierra Nevada Corporation weeks later (Ridenoure, 'NEAP: 15 years later', 2013).

In 2009 Planetary Resources, Inc (PRI; formerly Arkyd Astronautics), one of the most well-known entities, was formed. Its mission was to become the primary provider of asteroid resources for the future near-Earth space ecosystem. Big names backed the start-up run by former NASA Mars Exploration Rover and Phoenix Mars Lander flight director Chris Lewicki. In early investor rounds, the likes of Google's Larry Page and Eric Schmidt, Virgin Galactic's Sir Richard Branson, X-Prize Foundation's Peter Diamandis, Space Adventures Ltd.'s Eric Anderson, filmmaker James Cameron, as well as funding from the country of Luxembourg's Space Resources initiative powered the company. JPL engineers and Seattle-area tech players, among others, rounded out a capable team. Asteroid penetrator systems for subsurface resource verification, novel distributed systems design, spacecraft orbit and target optimization, and advanced camera technology were under rapid development during PRI's heyday. The company was developing a line of small satellites called "Arkyd" that possessed specialized hyperspectral imaging instruments that could survey NEAs up close. After successful development and launch of the Arkyd-3 satellite, the company was well underway in proving its technological prowess and began a large hiring spree to gear up for a suite of exploration missions (eventually exceeding 70 employees). Plans that would have seen the development of advanced prospecting spacecraft that would fly to six NEAs to verify the presence and abundance of hydrated minerals were halted when the company ran out of funding and sadly shuttered its doors in 2018. Subsequently its assets were sold to ConsenSys Space.

Another asteroid mining company of note was Deep Space Industries (DSI), founded in 2013 by Space Frontier Foundation's Rick Tumlinson, Orbit Fab's Daniel Faber, and several others. DSI endeavored to democratize space and lower costs, and to this end were developing a water-based propulsion system for in-orbit boosters and tugs as well as its own prospecting spacecraft called Xplorer. The company also ran out of venture capital funding and closed its doors in 2019 after being sold to Bradford Space. DSI had begun efforts to create asteroid simulants for sale to the community, but sold this warehouse and capability to Exolith Labs, run by Dr Daniel Britt (University of Central Florida), who is now a leading supplier of planetary surface regolith simulants.

24.4.3.3 The Current State of the Field

The efforts by these and other early players have helped shape the industry and reminded us about the difficulties faced by smaller companies having to shoulder the burden of high R&D and launch costs while faced with an uncertain or nonexistent market. The upfront costs to harvest resources are substantial (\$10–100s of millions), and companies trying to survive in this field have mostly centered around aerospace, defense, launch, or related industries. ULA has declared an offer for \$10,000 for every kg of water delivered in lunar orbit and laid a framework for a roadmap for usable resources in cis-lunar space (Kornuta et al., *Commercial Lunar Propellant Architecture* 2018), but even they are not ready to receive that product in the early 2020s. NASA, or governments more generally, are regarded as the primary customer for the immediate time being (Gateway, Artemis, Orion), with the primary near-term focus being the Moon (Artemis, Moon-to-Mars roadmap).

Many in the field share a common vision and are invested in the success of asteroid ISRU. The departure of some of the early players raises questions about the realistic timeline for asteroid mining (Abrahamian, ‘How the asteroid-mining bubble burst’ 2019). Given the complexities and a vast array of engineering and legal frameworks that still have to be worked out, the financial backing for and the extraction of resources from an asteroid will probably not come to pass for some years, presumably the late 2020s or later. There are companies still in operation today with grant-driven or non-revenue-based business models. Companies like Trans Astronautica Corporation have devoted themselves to pure R&D (won Phase II and III NASA Innovative Advance Concepts (NIAC) grants). Their development of Optical Mining™ technologies and ways to extract the liberated gases and ices offers a unique and differentiated approach. Others in the game exist, but with limited funding have focused more on developing a roadmap and/or initial concepts with few technological breakthroughs, e.g., Aten Engineering, Asteroid Mining Corporation, or the Planetoid Mining Company. More Earth- or lunar-focused entities like Deltion Engineering, Off World industries, or Moon Express may be able to pivot towards asteroid mining in the future if the market for it forms. In some ways the market and technological readiness for asteroid mining is a chicken-and-egg problem. The initial attempts may be limited to demonstration missions with no real customers, or funded entirely by tax payers. We may end up relying on the long timeframe of Orion and SLS (NASA) unless costs are dramatically reduced by Falcon 9 and Starship (SpaceX).

24.4.3.4 Asteroids as a Stepping Stone

The process of building out the near-Earth space, or “bootstrapping,” is a broad concept meant to convey a technological and infrastructure build-out of the key pieces of ISRU in cis-lunar space. It is possible due to advances in manufacturing techniques and robotics to establish a semi-self-sustaining (leading to full automation in a couple of decades) industrial manufacturing chain on the Moon, and perhaps

in near-earth space (Metzger et al. 2013). In the larger vision, asteroids are utilized for resources in a parallel and symbiotic manner as the Earth–Moon system. As a stepping stone, NEAs are energetically favorable with often low delta-vs required to reach them, with surveys of potential targets already taking place (e.g., Ieva et al. 2014). NEAs cost relatively little in fuel compared with the Moon’s gravity well. For reference, it would take roughly 96% of a rocket’s mass in fuel to leave Earth’s gravity well, but even with 1/6-g it is still ~50% for the Moon. Asteroids would be substantially lower than this. The exploitation of NEAs enables orbital refueling and manufacturing serving near-Earth space. Refueling depots will eventually allow us to expand the markets for a new space economy. Asteroid resources may provide a key aspect of the potential future growth and ultimately an economic driver, but that will take time.

To get there, we need to hone in on the major hurdles yet to be solved. Many gaps still exist in our understanding of asteroids, their resources, and the frameworks necessary for their eventual utilization. ASIME’s 2016 workshop, *Asteroid Science Intersections with In-Space Mine Engineering*, sought to bridge the divide between engineers, miners, and scientists with fruitful discussions that resulted in the forming of Science Knowledge Gaps (SKGs) (Graps et al. 2016). This led to the outlining of roadmaps toward economic opportunities and mining feasibility by interested parties (see Sects. 24.4.3.2 and 24.4.3.3). Some of the current key needs of asteroid ISRU can be combined and summarized in the following primary categories: (1) greater access to and ‘mining’ of NEA data, (2) constraining asteroid resources and geotechnical properties, (3) proximity and dust hazards and mitigation solutions, (4) advancement of extraction technologies, (5) higher-fidelity analogs for testing, and (6) a better understanding of the impact of activities and generated spacecraft or asteroid debris in or near cis-lunar space (Lewicki et al. 2020). This wide array of themes will likely not be solved by scientific pursuits alone. It seems likely that some sort of synergies between science, technology advances, and economic incentivization will need to play out in this new decade for further advances to occur.

24.5 Discussion and Conclusions

Grounded in Reality. For asteroid resource utilization and the build-out of a new space-faring era to become a reality, there need to be larger catalysts than currently exist. Several things are needed before this can occur. Ground truth may not be necessary to determine an asteroid’s volatile reserves, but it is likely needed to lower government or investors’ risks, a roadblock to taking that leap. In the past, commercial startups such as PRI and DSI did not succeed in their missions because financial backers did not have the prospecting knowledge comparable to terrestrial resource exploration and the assurances that come with that. Bound by the constraints of terrestrial mining, and the typical types of geological or chemical assessments required prior to funding a mining operation, investors saw high levels of risk coupled with high costs, compounded by a long time-horizon for their return on investment (likely

twice or more than that of a typical ROI for a venture-backed company). With no precedent, only a promise of profits, finding those willing to invest tens of millions of dollars has been difficult. Terrestrial industries and mining players simply do not have enough incentive to break their risk-to-reward posture. The return of samples from Ryugu and Bennu may help slowly change this. If water is revealed in significant proportions through analysis of these samples, and the extrapolations from meteorite analogs are verified, it could be the spark that is needed to spurn real movement in this area.

Steady Progress but Nowhere Fast. The current rate of progress in developing the spacecraft that will be required would have us touching down to mine an asteroid in the 2030s at the earliest. NASA, arguably the only real main customer currently, has for decades funded many ISRU R&D projects at various NASA centers. Innovations like the Regolith Advanced Surface Systems Operations Robot (RASSOR) excavator mining robot and landing pad construction at Kenney Swamp Works (Mueller et al. 2016; Mueller et al. 2017), in-space additive manufacturing (Grossman et al. 2019) and metals and oxygen extraction using molten regolith electrolysis occurring at Kennedy Space Center's Gaseous Lunar Oxygen from Regolith Electrolysis (GaLORE) project, and at Johnson Space Center (Schreiner et al. 2015), show the promise and potential, but also highlight part of the problem. Current funding is mostly internal, with very slow technology development. The Lunar Resource Prospector (RP) that was slated to be the first to demonstrate ISRU technology, and may have proven out this model, was canceled in 2018. Thankfully, NASA is developing the Volatiles Investigating Polar Exploration Rover (VIPER) mission, which essentially replaces RP and will launch this decade. The results of this mission could verify large amounts of water at the lunar south pole. This could lead to cis-lunar propellant refueling and lower long-term surface habitation risk by providing consumables in situ.

The problem is not a lack of brilliant minds working to solve the current slate of technological hurdles, it is a lack of funding to mature technology that is still in its infancy and larger-scale directives from the US government and NASA. Until these happen, private companies will be unable to take the risk. NASA needs to provide more than bite-sized attempts at spurring technological innovation. The Artemis program is slated to send a suite of new demonstrations, with ISRU listed on their planned roadmap. Those invested deeply hope that Artemis will be fully funded by Congress to develop and see these lunar demonstrations through.

On a basic level we need more R&D-level research funding, like the offerings in the Science Mission Directorate (SMD) omnibus solicitation Research Opportunities in Space and Earth Sciences (ROSES) and Science and Technology Mission Directorate (STMD) Space Technology Research, Development, Demonstration, and Infusion (REDDI). The Next Space Technologies for Exploration Partnerships (NextSTEP), NASA Small Business Innovation Research (SBIR), Small Business Technology Transfer (STTR), Centennial Challenge, and NASA Innovative Advanced Concepts (NIAC) all fall under the STMD directorate and are great examples of programs working to benefit advances in ISRU. Even more critical, however, are commercial

partnerships and dedicated programs at larger scales with longer time horizons to avoid the political cycling problem. New administrations and Congresses come in and totally derail previous plans or keep funding levels too low to have a large impact. ISRU is too multifaceted and the lead time for technology development is longer than political appointments. This is an unsustainable paradigm that greatly limits the impact ISRU could have on long-term space exploration and the eventual habitation of the Moon and Mars. This has been one of the obstacles preventing its development and integration into existing mission architectures.

With so many obstacles, ISRU remains a slow ramp, but has the potential to greatly reduce the cost and risk of space exploration. Many concepts remain purely theoretical or partial working demonstrations. If successfully developed and deployed, ISRU technology and infrastructure can improve life cycles, lower costs for equipment and delivery to bases, and achieve better mission performance and capabilities, better science, all right at the source. The primary reason is sustainability: the ability to rely, not on costly risky resupply missions, but on local in-situ resources. This paradigm would support more and more destinations ever expanding outward, but we need to solve all these parts to make asteroid mining and fuel depots a reality, otherwise it will always seem years in the future.

Final Takeaways. When eventual extraction begins, there are serious questions regarding the potential loss of priceless science knowledge due to surface activities. Information regarding impact flux, geological history, compositions, and other areas needs to be preserved for posterity, as does knowledge that is vital to understanding our solar system's history and resolving many unanswered science questions. If extraction spacecraft damage the surfaces and remove precious material, this knowledge could be lost forever. Will industrial and commercial companies care about erasing this knowledge? From the legal and regulatory framework perspective, we need to legislate, create charters, and international treaties (replacing the ancient Outer Space Treaty of 1967) that will direct players to document everything, to put in place certain preservation protocols. This way scientific progress and the development of commercial interests can coincide.

From where we currently stand, full-scale asteroid mining is still more than ten years away, despite our knowing about the huge resources they contain. Sadly, this estimate has not changed much since 2018 when some of the main commercial players were forced to close. Missions that will attempt extraction via small satellites or technology demonstrations might be able to perform initial assays and extraction in this decade if funding is made available and launch costs drastically reduced. After the successful Hayabusa2 and OSIRIS-REx sample return missions interacted with small potentially resource-rich NEAs, government and commercial interests might respond encouragingly. The focus will need to be on technological development and initial demonstration missions in order to put us on the path towards progress in the late 2020s and early 2030s to break the chicken-and-egg stalemate. There will never be a perfect time to take the risk, or rather, we can never know that time without hindsight. The initial costs will be high but may come down due to the coinciding of reusable rockets and further still from refueling in orbit. The long-term

path still seems clear. The risks of not pursuing near-Earth resources, of not pushing the boundaries of humanity deeper into space, will be far greater than profits lost.

Acknowledgements Acknowledgments are due to some of the people in my orbit for encouraging and supporting my interests and career in science and space resources, including my former Planetary Resources, Inc. teammates, professors and connections made at the University of Central Florida, and to my colleagues at Southwest Research Institute. Reviews by Dr Christopher Dreyer and Dr Daniel Britt helped improve the manuscript.

References

- Alvarez, L., W. Alvarez, F. Asaro, and H. Michel. 1980. Extraterrestrial cause for the Cretaceous-tertiary extinction. *Science* 208: 1095.
- Alexander, S.R., and J. Winnick. 1994. Removal of hydrogen sulfide from natural gas through and electrochemical membrane separator. *AIChE Journal* 40 (4): 613.
- Baker, L., I.A. Franchi, I.P. Wright, and C.T. Pillinger. 2002. The oxygen isotopic composition of water from Tagish Lake: Its relationship to low-temperature phases and other carbonaceous chondrites. *Meteoritics & Planetary Science* 37: 977.
- Becker, T.M., N. Cunningham, M. Philippa, L. Roth, L.M. Feaga, K.D. Retherford, Z.A. Landsman, E. Peavler, L.T. Elkins-Tanton, and J. Walhund. 2020. HST UV observations of asteroid (16) Psyche. *Planetary Science Journal* 1 (53): 1.
- Benson, J. W. 1998. Near earth asteroid prospector. *1998 IEEE Aerospace Conference Proceedings*, 5, pp. 457.
- Binzel, R.P., D. Lupishko, M. di Martino, R.J. Whiteley, and G.J. Hahn. 2002. "Physical properties of near-Earth objects. In *Asteroids III*, ed. W.F. Bottke, A. Cellino, P. Paollicchi, and R. Binzel, 255–271. Tucson: Univ. of Arizona Press.
- Bland, P.A., M.E. Zolensky, G.K. Benedix and M.A. Sephton. 2004. Weathering of chondritic meteorites. In D.S. Lauretta, and Jr H.Y. McSween (Eds.), *Meteorites and the Early Solar System II*, Univ. of Arizona, Tucson, pp. 853.
- Blum, J., and G. Wurm. 2008. The growth mechanisms of macroscopic bodies in protoplanetary disks. *Annual Review of Astronomy and Astrophysics* 46: 21.
- Bottke, W.F., D.P. Rubincam, and J.A. Burns. 2000. Dynamical evolution of main belt meteoroids: Numerical simulations incorporating planetary perturbations and Yarkovsky thermal forces. *Icarus* 145: 301.
- Bottke, W.F., A.V. Moorhead, H.C. Connolly Jr., C.W. Hergenrother, J.L. Molaro, P. Michel, M.C. Nolan, S.R. Schwartz, D. Vokrouhlicky, K.J. Walsh, and D.S. Lauretta. 2020. Meteoroid impacts as a source of Bennu's particle ejection events. *JGR Planets* 125 (8): 1.
- Bottke, W., D. Vokrounky, D. Rubincam and D. Nesvorný. 2006. The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annual Review of Earth and Planet Science* 34: 157.
- Britt, D.T., and L.A. Lebofsky. 1992. Spectral variations within asteroid classes. *Lunar and Planetary Science* 23: 161.
- Britt, D.T., D. Yeomans, K. Housen, and G. Consolmagno. 2002. Asteroid density, porosity, and structure. In *Asteroids III*, ed. W.F. Bottke, A. Cellino, P. Paollicchi, and R. Binzel, 485. Tucson: Univ. of Arizona Press.
- Brown, P., A.R. Hildebrand, M.E. Zolensky, M. Grady, R.N. Clayton, T.K. Mayeda, and 16 co-authors. 2000. The fall, recovery, orbit, and sand composition of the Tagish lake meteorite: A new type of carbonaceous chondrite. *Science* 290, 320.
- Brown, P., R.E. Spaldin, D.O. ReVelle, E. Tagliaferri, and S.P. Worden. 2002. The flux of small near-Earth objects colliding with the Earth. *Nature Letters* 420: 294.

- Brown, P., P. Wiegert, D. Clark, and E. Tagliaferri. 2016. Orbital and physical characteristics of meter-scale impactors from airburst observations. *Icarus* 266: 96.
- Burbine, T. 2002. Asteroids: Their composition and impact threat. *Bulletin of the Czech Geology Surveys* 77 (4): 243.
- Burbine, T.H., T.J. McCoy, A. Meibom, B. Gladman, and K. Keil. 2002. Meteoritic Parent Bodies: Their Number and Identification. In *Asteroids III*, ed. W.F. Bottke, A. Cellino, P. Paolicchi, and R. Binzel, 653. Tucson: Univ. of Arizona Press.
- Bus, S.J., and R.P. Binzel. 2002. Phase II of the small main-belt asteroid spectroscopic survey. *Icarus* 158: 146.
- Buseck, P.R., and X. Hua. 1993. Matrices of carbonaceous chondrite meteorites. *Annual Review of Earth and Planetary Sciences* 21: 255.
- Campbell, A.J., and M. Humayun. 2005. Composition of Group IVB Iron meteorites and their parent melt. *Geochimica Et Cosmochimica Acta* 69 (19): 4733.
- Cheng, A.F., J. Atchison, B. Kantsiper, A.S. Rivkin, A. Stickler, C. Reed, A. Galvez, I. Carnelli, P. Michel, and S. Ulamec. 2015. Asteroid impact and deflection assessment mission. *Acta Astronautica* 115: 262.
- Clark, B.E., S.J. Bus, A.S. Rivkin, M.K. Shepard, and S. Shah. 2004. Spectroscopy of X-type asteroids. *Astrophysical Journal* 128: 3070.
- Cloutis, E., R.P. Binzel, and M.J. Gaffey. 2014. Establishing asteroid-meteorite links. *Elements* 10 (1): 25.
- Delbo, M., G. Libourel, J. Wilkerson, N. Murdoch, P. Michel, K.T. Ramash, C. Ganino, C. Verati, and S. Marchi. 2014. Thermal fatigue as the origin of regolith on small asteroids. *Nature Letters* 508: 233.
- DeMeo, F.E., R.P. Binzel, S.M. Slivan, and S.J. Bus. 2009. An extension of the Bus asteroid taxonomy into the near-infrared. *Icarus* 202: 160.
- Dreyer, C.B., J. Sercel, L. Gertsch, A. Lampe, T.J. Canney and A. Abbud-Madrid. 2016. Optical mining subscale testing. In *Earth and Space 2016: Engineering for Extreme Environments* (pp. 493–506). Reston, VA: American Society of Civil Engineers, 2016.
- Dunn, T.L. 2008. *Determination of mineral abundances in ordinary chondrites using powder X-ray diffraction: Applications to parent body processes and asteroid spectroscopy*. PhD diss.: University of Tennessee.
- Fieber-Beyer, S.K., M.J. Gaffey, P.S. Hardersen, and V. Reddy. 2012. Near-infrared spectroscopy of 3:1 Kirkwood Gap asteroids: Mineralogical diversity and plausible meteorite parent bodies. *Icarus* 221: 593.
- Flynn, G.J., G.J. Consolmagno, P. Brown, and R.J. Macke. 2018. Physical properties of the stone meteorites: Implications for the properties of their parent bodies. *Chemie Der Erde* 78: 269.
- Flynn, G.J., L.B. Moore, and W. Klöck. 1999. Density and porosity of stone meteorites: Implications for the density, porosity, cratering, and collisional disruption of asteroids. *Icarus* 142: 97.
- Fornasier, S., B.E. Clark, E. Dotto, A. Migliorini, A. Ockert-Bell, and M.A. Barucci. 2010. Spectroscopic survey of M-type asteroids. *Icarus* 201 (2): 655.
- Fornasier, S., C. Lantz, M.A. Barucci, and M. Lazzarin. 2014. Aqueous alteration on main belt primitive asteroids: Results from visible spectroscopy. *Icarus* 233: 163.
- Fujiwara, A., J. Kawaguchi, D.K. Yeomans, M. Abe, T. Mukai, T. Okada, J. Saito, H. Yano, M. Yoshikawa, D.J. Scheeres, O. Barnouin-Jha, A.F. Cheng, H. Demura, R.W. Gaskell, N. Hirata, H. Ikeda, T. Kominato, H. Miyamoto, A.M. Nakamura, R. Nakamura, S. Sasaki, and K. Uesugi. 2006. The rubble-pile asteroid itokawa, as observed by Hayabusa. *Science* 312, 1330.
- Gaffey, M.J. 1976. Spectral reflectance characteristics of the meteorite classes. *Journal of Geophysical Research* 81 (5): 905.
- Gaffey, M.J., E.A. Cloutis, M.S. Kelley and K.L. Reed. 2002. Mineralogy of asteroids. In W.F. Bottke, A. Cellino, P. Paolicchi, and R.P. Binzel (Eds.), *Asteroids III*. University of Arizona Press, Tucson, pp. 183–204.
- Gaffey, M.J., J.F. Bell, R.H. Brown, T.H. Burbine, J.L. Piatek, K.L. Reed, and D.A. Chaky. 1993. Mineralogical variations within the S-type asteroid class. *Icarus* 106: 573.

- Gertsch, R., J.L. Remo and L.S. Gertsch. 1997. Near-earth resources. *Annals of the N.Y. Acad. Of Sci.*, 822, pp. 468.
- Goldreich, P., and W.R. Ward. 1973. The formation of planetesimals. *ApJ* 183: 1051.
- Graps, A., P. Blondel, G. Bonin, D. Britt and 27 co-authors. 2016. In-Space Utilisation of Asteroids: Answers to questions from the asteroid miners. *Europlanet*, pp. 85.
- Gravnik, M., A. Morbidelli, R. Jedicke, B. Bolin, W.F. Bottke, E. Beshore, D. Vokriuhilicky, D. Nesvorny, and P. Michel. 2018. Debaised orbit and absolute-magnitude distributions for near-Earth objects. *Icarus* 312: 181.
- Grimm, R.E., and H. Y. McSween. 1989. Water and the thermal evolution of carbonaceous chondrite parent bodies. *Icarus* 82: 244–280.
- Grossman, K.D., T.S. Skthivel, L. Sibille, J.G. Mantovani, and S. Seal. 2019. Regolith-derived ferrosilicon as a potential feedstock material for wire-based additive manufacturing. *Advances in Space Research* 63: 2212.
- Hamilton, V.E., A.A. Simon, P.R. Christensen, D.C. Reuter, B.E. Clark, M.A. Barucci, N.E. Bowles, W.V. Boynton, J.R. Brucato, E.A. Cloutis, Jr H.C. Connolly, K.L. Donaldson Hanna, J.P. Emery, H.L. Enos, S. Fornasier, C.W. Haberle, R.D. Hanna, E.S. Howell., H.H. Kaplan, L.P. Keller, C. Lantz, J.-Y. Li, L.F. Lim, T.J. McCoy, F. Merlin, M.C. Nolan, A. Praet, B. Rozitis, S.A. Sandford, D.L. Schrader, C.A. Thomas, X.-D. Zou, D.S. Lauretta and the OSIRIS-RExTeam. 2019. Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nature Astronomy*, 3, pp. 332.
- Hinterman, E., and J.A. Hoffman. 2020. Simulating oxygen production on mars for the mars oxygen in-situ resource utilization experiment. *Acta Astronautica* 170: 678.
- Hiroi, T., M.E. Zolensky, and C.M. Pieters. 2001. The Tagish Lake meteorite: A possible sample from a D-type asteroid. *Science* 293: 2234.
- Ieva, S., E. Dotto, D. Perma, M.A. Barucci, F. Bernardi, S. Fornasier, F. De Luise, E. Perozzi, A. Rossi, and J.R. Brucato. 2014. Low delta-V near-Earth asteroids: A survey of suitable targets for space missions. *A&a* 569 (A59): 1.
- Jarosewich, E. 1990. Chemical analysis of meteorites: A compilation of stony and iron meteorite analysis. *Meteoritics* 25: 323.
- Jaumann, R., N. Schmitz, T.-M. Ho, W.E. Schröder, K.A. Otto, K. Stephan, S. Elgner, K. Krohn, F. Preusker, F. Scholten, J. Biele, S. Ulamec, C. Krause, S. Sugita, K.-D. Matz, T. Roatsch, R. Parekh, S. Mottola, M. Grott, P. Michel, F. Trauthan, A. Koncz, H. Michaelis, C. Lange, J.T. Grundmann, M. Maibaum, K. Sasaki, F. Wolff, J. Reill, A. Moussi-Soffys, L. Lorda, W. Neumann, J.-B. Vincent, R. Wagner, J.-P. Bibring, S. Kameda, H. Yano, S. Watanabe, M. Yoshikawa, Y. Tsuda, T. Okada, T. Yoshimitsu, Y. Mimasu, T. Saiki, H. Yabuta, H. Rauer, R. Honda, T. Morota, Y. Yokota and T. Kouyama. 2019. Images from the surface of asteroid Ryugu show rocks similar to carbonaceous chondrite meteorites. *Science*, 365: 817.
- Johansen, A., F. Brauer, C.P. Dullemond, H. Klahr, and T. Henning. 2008. A coagulation- fragmentation model for the turbulent growth and destruction of preplanetesimals. *A&A* 486: 597.
- Keszthelyi, L.P., J.J. Hagerty, A. Bowers, K.J. Ellefsen, I. Ridley, T. King, D. Trilling, N. Moskovitz and W. Grundy. 2017. Feasibility study for the quantitative assessment of mineral resources in asteroids. Astrogeology Science Center, USGS, Open-File Report 2017–1041.
- Kimberley, J., and K. Ramesh. 2011. The dynamic strength of an ordinary chondrite. *Meteoropition Planetary Science* 46 (11): 1653.
- Kornuta, D., A. Abbud-Madrid, J. Atkinson, J. Barr, G. Barnhard, D. Beinhoff, B. Blair, V. Clark, J. Cyrus, B. DeWitt, C. Dreyer, B. Finger, J. Goff, K. Ho, L. Kelsey, J. Kervalva, B. Kutter, P. Metzger, L. Montgomery, P. Morrison, C. Neal, E. Otto, G. Roesler, J. Schier, B. Seifert, G. Sowers, P. Spudis, M. Sundahl, K. Zacny, and G. Zhu. 2018. Commercial lunar propellant architecture. *REACH* 13: 100026.
- Landsman, Z.A., J.P. Emery, H. Campins, J. Hanus, L.F. Lim, and D.P. Cruikshank. 2018. Asteroid (16) Psyche: Evidence for a silicate regolith from Spitzer space telescope spectroscopy. *Icarus* 304: 58.

- Lewicki, C., A. Graps, M. Elvis, M. Metzger, and A. Rivkin. 2020. Furthering asteroid resource utilization in the next decade through technology leadership. *Planetent Science and Astrobiology Decadal Survey 2023–2032* White Paper, pp. 1.
- Lewis, J.S. 1996. *Mining the Sky: Untold riches from the asteroids, comets, and planets*, 274. New York: Basic Books.
- Luu, J.X., and D.C. Jewitt. 1989. On the relative numbers of C-Types and S-Types among near earth asteroids. *Astronomical Journal* 98: 1905.
- Macke, J. Robert. 2010. Survey of Meteorite Physical Properties Density, Porosity and Magnetic Susceptibility. PhD diss., University of Central Florida, pp. 1638.
- Macke, R.J., G.J. Consolmagno, and D.T. Britt. 2011. Density, porosity, and magnetic susceptibility of carbonaceous chondrites. *Meteoritics & Planetary Science* 46 (12): 1842.
- Mason, B., and A.L. Graham. 1970. Minor and trace elements in meteoritic minerals. *Smithsonian Contributions to the Earth Science* 3: 1.
- McCord, T.B., J.B. Adams, and T.V. Johnson. 1970. Asteroid vesta: Spectral reflectivity and compositional implications. *Science* 168 (3938): 1445.
- McSween, H.Y., R.P. Binzel, C.D. Sanctis, E. Ammannito, T.H. Prettyman, A.W. Beck, V. Reddy, L. Le Corre, M.J. Gaffey, T.B. McCord, C.A. Raymond, C.T. Russel and the Dawn Science Team. 2013. *Meteoritics & Planetary Science* 48(11): 2090.
- Metzger, P., A. Muscatello, R.P. Mueller, and J. Mantovani. 2013. Affordable, rapid bootstrapping of the space industry and solar system civilization. *Journal of Aerospace Engineering* 26 (1): 18.
- Michel, P., R.L. Ballouz, O.S. Barnouin, M. Jutzi, K.J. Walsh, B.H. May, C. Manzoni, D.C. Richardson, S.R. Schwartz, S. Sugita, S. Watanabe, H. Miyamoto, M. Hirabayashi, W.F. Bottke, H.C. Connoly, M. Yoshikawa, and D.S. Lauretta. 2020. Collisional formation of top-shaped asteroids and implications for the origins of Ryugu and Bennu. *Nature Communications* 11: 2655.
- Michikami, T., A.M. Nakamura, N. Hirata, R.W. Gaskell, R. Nakamura, T. Honda, Ct. Honda, K. Hiroka, J. Saito, H. Demura, M. Ishiguro and H. Miyamoto. 2008. Size-frequency statistics of boulders on global surface of asteroid 25143 Itokawa. *Earth and Planets Space*, 60, pp. 13.
- Molaro, J.L., C.W. Hergenrother, S.R. Chesley, K.J. Walsh, R.D. Hanna, C.W. Haberle, S.R. Schwartz, R.-L. Ballouz, W.F. Bottke, H.J. Campins, D.S. Lauretta. 2020. Thermal fatigue as a driving mechanism for activity on asteroid Bennu. *JGR* 125(8).
- Morbidelli, A., W. Bottke, D. Nesvorný, and H. Levison. 2009. Asteroids were born big. *Icarus* 204 (2): 558.
- Mueller, R.P., J.D. Smith, J.M. Schuler, A.J. Nick, N.J. Gelino, K.W. Leucht, I.I. Townsend, and A.G. Dokoset. 2016. Design of an excavation robot: Regolith advanced surface systems operations robot (RASSOR) 2.0. *ASCE Earth and Space*, 163.
- Mueller, R.P., J.C. Fikes, M.P. Case, B. Khoshnevis, M.R. Fiske, J.E. Edmunson, R. Kelso, R. Romo and C. Andersen. 2017. Additive construction with mobile emplacement (AMCE). 68th IAC, D3.2.1, pp. 1.
- Musk, E. 2017. Making humans a multi-planetary species. *New Space* 5 (2): 46.
- Nakamura, T., et al. 2011. Itokawa dust particles: A direct link between S-type asteroids and ordinary chondrites. *Science* 333 (6046): 1113.
- NASA Authorization Act of 2005. December 30, 2005, U.S. House. 109th Congress, 1st Session., Public Law 109–155. <https://www.gpo.gov/fdsys/pkg/PLAW-109publ155/pdf/PLAW-109publ155.pdf>.
- Neeley, J.R., B.E. Clark, M.E. Ockert-Bell, M.K. Shepard, J. Conklin, E.A. Cloutis, S. Fornasier, and S.J. Bus. 2014. The composition of M-type asteroids II: Synthesis of spectroscopic and radar observations. *Icarus* 238: 37.
- Nesvorný, D., and W. Bottke. 2004. Detection of the Yarkovsky effect for main-belt asteroids. *Icarus* 170: 324.
- Neumann, W., D. Breuer, and T. Spohn. 2014. Differentiation of Vesta: Implications for a shallow magma ocean. *Earth and Planetary Science Letters* 395: 267.

- Nicheporuk, W., and H. Brown. 1965. The distribution of platinum and palladium metals in iron meteorites and in metal phase of ordinary chondrites. *Journal of Geophysics Research* 70 (2): 459.
- Öpik, E.J. 1951. Collision probabilities with the planets and the distribution of interplanetary matter. *Proceedings. Royal Irish Academy* 54: 165.
- Parker, A.H., D.D. Durda, A.D. Whizin, K.J. Walsh, M. Shoffner and B. Pyke. 2021. BORE-II demonstration of microsatellite-scale magnetic asteroid regolith sample collection system. *52nd Lunar and Planetary Science Conference*, LPI Cont. No. 2548, pp. 2454.
- Popova, O., J. Borovicka, W.K. Hartmann, P. Spurny, E. Gnos, I. Nemtchinov, and J.N. Trigo-Rodríguez. 2001. Very low strengths of interplanetary meteoroids and small asteroids. *Meteoritics & Planetary Science* 46 (10): 1525.
- Ralchenko, M., D.T. Britt, C. Samson, C.D.K. Her and P.J.A. McCausland. 2014. bulk physical properties of the tagish lake meteorite frozen pristine fragments. *45th Lunar and Planetary Science Conference* LPI Cont. No. 1777, pp. 1021.
- Reddy, V., T.L. Dunn, C.A. Thomas, N.A. Moskovitz and T.H. Burbine. 2015. Mineralogy and surface composition of asteroids. In: *Asteroids IV*, Michel, P., DeMeo, F.E., Bottke, W.F. (Eds.), University of Arizona Press, Tucson, 895, pp. 43.
- Rivkin, A.S., H. Campins, J.P. Emery, E.S. Howell, J. Licandro, D. Takir and F. Vilas. 2015. Astronomical observations of volatiles on asteroids. In: *Asteroids IV*, Michel, P., DeMeo, F.E., Bottke, W.F. (eds.), University of Arizona Press, Tucson, 895, pp. 65.
- Rivkin, A.S., E.S. Howell, F. Vilas, and L.A. Lebofsky. 2002. Hydrated minerals of asteroids: The astronomical record. In *Asteroids III*, ed. W.F. Bottke, A. Cellino, P. Paolicchi, and R. Binzel, 235. Tucson: Univ. of Arizona Press.
- Rubin, A.E. 1997. Mineralogy of meteorite groups. *Meteoritics and Planetary Science* 32: 231–247.
- Sanchez, P., and D.J. Scheeres. 2014. The strength of regolith and rubble pile asteroids. *Meteoritics & Planetary Science* 49 (5): 788.
- Scheeres, D.J., C.M. Hartzell, P. Sanchez, and M. Swift. 2010. Scaling forces to asteroid surfaces: The role of cohesion. *Icarus* 210: 968.
- Schreiner, S., L. Sibille, J. Dominguez, A. Sirk, J. Hoffman and G. Sanders. 2015. Development of a molten regolith electrolysis reactor model for lunar in-situ resource utilization. *Aerospace Research Central, 8th Symposium on Space Resource Utilization*.
- Sercel, J.C., C.B. Dreyer, A. Abbud-Madrid, D. Britt, R. Jedicke, L. Gertsch and S.G. Love. 2016. A coordinated research program to develop the technology to optical mine asteroids. In *Earth and Space 2016: Engineering for Extreme Environments*, pp. 507–522. Reston, VA: American Society of Civil Engineers.
- Sercel, J.C., C.E. Peterson, J.R. French, A. Longman, S.G. Love and Shishko, R. 2018. Stepping stones: Economic analysis of space transportation supplied from NEO resources. In *2018 IEEE Aerospace Conference* (pp. 1–21). IEEE.
- Shoemaker, E.M. 1983. Asteroid and comet bombardment of the Earth. *Annual Review of Earth and Planetary Sciences* 11: 461–494.
- Springmann, A., D.S. Lauretta, B. Klaue, Y.S. Goreva, J.D. Blum, A. Andronikov, and J.K. Steckloff. 2019. Thermal alteration of labile elements in carbonaceous chondrites. *Icarus* 324: 104.
- Stuart, J., and R. Binzel. 2004. Bias-corrected population, size distribution, and impact hazard for the near-Earth objects. *Icarus* 170: 295.
- Tholen, D.J. 1984. Asteroid taxonomy from cluster analysis of photometry.” PhD thesis, University of Arizona, pp. 150.
- Vincent, J., and M. King. 1995. The mechanism of drilling by wood wasp ovipositors. *Biomimetics* 3 (4): 187.
- Vasilyev, N. 1998. The Tunguska meteorite problem today. *Planetary and Space Science* 46 (2): 129.
- Wada, K., H. Tanaka, T. Suyama, H. Kimura, and T. Yamamoto. 2009. Collisional growth conditions for dust aggregates. *ApJ* 702: 1490.
- Walsh, K. 2018. Rubble pile asteroids. *Annual Review of Astronomy and Astrophysics* 56 (1): 593.

- Watanabe, S., M. Hirabayashi, N. Hirata, Na. Hirata, R. Noguchi, and 83 co-authors. 2019. Hayabusa2 arrives at the carbonaceous asteroid 162173 Ryugu—A spinning top-shaped rubble pile. *Science* 364, 268.
- Weidenschilling, S.J. and J.N. Cuzzi. 1993. *Formation of planetesimals in the solar nebula*. PSPL-UA Press, pp. 1031.
- Wing, M.R., and J.L. Bada. 1991. Geochronology on the parent body of the carbonaceous chondrite Ivuna. *Geochimica Et Cosmochimica Acta* 55: 2937.
- Wiik, H.B. 1956. The chemical composition of some stony meteorites. *Geochim. Cosmochim. Acta* 9: 279–289.
- Zenobi, R., J.M. Philippoz, P.R. Buseck, and R.N. Zare. 1989. Spatially resolved organic analysis of the Allende meteorite. *Science* 246: 1026.

Internet Links

- Abrahamian, A.A. (06/26/2019). How the asteroid-mining bubble burst. *MIT Technology Review*. <https://www.technologyreview.com/2019/06/26/134510/asteroid-mining-bubble-burst-history/>.
- Boyle, A. (11/04/2019). One year after Planetary Resources faded into history, space mining retains its appeal. *GeekWire*. <https://www.geekwire.com/2019/one-year-planetary-resources-faded-history-space-mining-retains-appeal/>.
- Carter, J. (05/12/2020). A Bizarre Trillion Dollar Asteroid Worth More than Our Planet is Now Aligned with the Earth and Sun. *Forbes*. <https://www.forbes.com/sites/jamiecartereurope/2020/12/05/a-bizarre-trillion-dollar-asteroid-worth-more-than-our-planet-is-now-aligned-with-the-earth-and-sun/?sh=31ba1b1b31c9>.
- Carver, J. (10/13/2020). Asteroid Sampling Technology Tested on Blue Origin’s Suborbital Rocket.” *SpaceRef*. <http://spaceref.com/news/viewpr.html?pid=56426>.
- Gaseous Lunar Oxygen from Regolith Electrolysis (GaLORE): <https://www.nasa.gov/feature/nasa-kennedy-to-develop-tech-to-melt-moon-dust-extract-oxygen>.
- Regolith Advanced Surface Systems Operations Robot (RASSOR): <https://technology.nasa.gov/patent/KSC-TOPS-7>.
- Ridenoure, R. (06/17/2013). NEAP: 15 years later. *The Space Review*. <https://www.thespacereview.com/article/2315/1>.
- SpaceX Falcon 9 rocket vehicle specification and capabilities: <https://www.spacex.com/vehicles/falcon-9/>.
- TransAstra Apis™ Space Craft and HoneyBee™ Optical Mining™ Concept. <https://www.transastracorp.com/apis-spacecraft>.

Chapter 25

Exploration of Asteroids and Comets with Innovative Propulsion Systems



Lorenzo Niccolai, Alessandro A. Quarta, and Giovanni Mengali

Abstract In recent decades, growing interest has arisen within the scientific community in asteroids, in particular near-Earth objects (NEOs), which are asteroids or comets characterized by a semimajor axis close to that of the Earth. Interest in these objects is motivated not only by their capability of providing answers to important questions concerning the origins of the solar system, but also by the threat of a possible future collision with Earth. It is quite common for NEOs to have peculiar orbits, with non-negligible inclinations and/or eccentricities. This implies that a rendezvous mission with such objects requires a large propellant consumption and gravity-assist maneuvers that significantly increase flight time when chemical (and also electrical) thrusters are used. Therefore, a promising option for exploring NEOs is provided by a propellantless propulsive system as solar sails or electric solar wind sails (E-sails), which are capable of generating thrust without consuming any propellant. In this chapter, the optimal solar sail- or E-sail-based transfers towards some of the most relevant NEOs are generated by solving an optimal control problem through an indirect approach. The results are compared with the transfer times obtained assuming a chemical thruster and a bi-impulsive maneuver.

List of Acronyms

DART	Double Asteroid Redirection Test
E-sail	Electric solar wind sail
IKAROS	Interplanetary Kite-craft Accelerated by Radiation Of the Sun

L. Niccolai · A. A. Quarta (✉) · G. Mengali
Department of Civil and Industrial Engineering, University of Pisa, Via G. Caruso 8, 56122 Pisa, Italy

e-mail: a.quarta@ing.unipi.it

L. Niccolai

e-mail: lorenzo.niccolai@ing.unipi.it

G. Mengali

e-mail: g.mengali@ing.unipi.it

LEO	Low-Earth orbit
OKEANOS	Oversize Kite-craft for Exploration and Astronautics in the Outer Solar System
OSIRIS-Rex	Origins Spectral Interpretation Resource Identification Security Regolith Explorer
NEA	Near-Earth asteroid
NEO	Near-Earth object
NEAR	Near-Earth Asteroid Rendezvous
NHATS	Near-Earth Object Human Space Flight Accessible Targets Study

25.1 Introduction

In the last few decades, growing interest has arisen within the scientific community in asteroids, in particular near-Earth asteroids (NEAs). This interest is motivated by NEAs' ability to provide answers to very important questions concerning the origin of the solar system, but also by the threat of a possible future collision with Earth. In this regard, an interesting and extensive review of the importance of NEA exploration for the next space era has been provided by Perna et al. (2013). From an historical viewpoint, the first close encounter with an asteroid was performed in October 1991 by the spacecraft Galileo, which passed about 1600 km from the surface of Gaspra (Belton et al. 1992). Galileo was followed by other asteroid flyby missions, such as Near-Earth Asteroid Rendezvous (NEAR) and Deep Space 1, which encountered the asteroid Mathilde in June 1997 (Veverka et al. 1999) and asteroid Braille in July 1999 (Oberst et al. 2001), respectively. A further milestone in robotic asteroid exploration was achieved by the NEAR-Shoemaker spacecraft, which performed proximity scientific observations (Veverka et al. 2001) and even surface operations (McCoy et al. 2002) on the asteroid 433 Eros in 2000. In 2005, the Japanese mission Hayabusa landed on asteroid Itokawa (Baker 2006), and in 2010 some samples collected from the asteroid surface returned to the Earth (Barucci et al. 2011). NASA's mission Dawn, launched in 2007, orbited around the asteroid Vesta in 2011 (Russell et al. 2013) and then reached the dwarf planet Ceres in 2015, remaining operative until 2018 (Rayman 2020a, b).

Interest in asteroid exploration remains very high, as is confirmed by operating and planned missions like JAXA's Hayabusa 2, NASA's OSIRIS-REx and DART, and ESA's Hera. In particular, Hayabusa 2 recently placed three rovers and a lander on the surface of asteroid Ryugu (Tsuda et al. 2020), and collected (by means of a kinetic impactor) some underground samples (Saiki et al. 2020) that are returned to the Earth. Origins Spectral Interpretation Resource Identification Security Regolith Explorer (OSIRIS-REx) performed a rendezvous with asteroid Bennu in December 2018 (Enos et al. 2019), and is going to be inserted into a return trajectory towards the Earth in March 2021, with the aim of bringing back some samples taken from the asteroid surface (Lauretta et al. 2017). NASA's mission Double Asteroid Redirection

Test (DART) has been launched in 2021 and has performed a deflection of the binary asteroid Didymos (Cheng et al. 2020). Then, the spacecraft Hera will investigate the effects of DART impact by analyzing the crater (with the aid of two CubeSats that will be carried onboard) and the new Didymos heliocentric trajectory, in order to assess the feasibility of an asteroid deflection mission in case of possible threats against Earth (Adams et al. 2019; Carnelli et al. 2019).

Like asteroids, comets have captured a significant scientific interest for a long time now. Indeed, during the last passing of Halley's comet, some of the most important world space agencies launched a number of satellites aimed at studying the comet's properties. Soviet spacecraft Vega 1 and Vega 2 reached Halley's comet in March 1986 after a Venus gravity assist and obtained some relevant data on the comet's physical parameters and composition (Kissel et al. 1986b). Simultaneously, ESA's mission Giotto also encountered the Halley's comet, providing other data on its physical composition (Kissel et al. 1986a) and its interaction with interplanetary magnetic field (Neubeuer et al. 1986), and later moving to comet Grigg-Skjellerup, which was reached in 1993 (McKenna-Lawlor et al. 1993). JAXA's missions Sakigake and Suisai (Itoh et al. 1986) conducted scientific observations of the interplanetary space (Saito et al. 1986) and provided UV images of the comet (Kaneda et al. 1986), respectively. NASA did not send a mission to the Halley's comet, but managed to launch several missions in the following years, including the International Cometary Explorer (ICE) towards comet Giacobini-Zinner (Von Roseninge et al. 1986), the Deep Space 1, which visited comet Borrelly (Boice et al. 2007), and the Stardust probe, which collected some samples from the coma of comet Wild 2 (Brownlee 2014). However, the most relevant cometary exploration mission was certainly ESA's Rosetta mission, launched in 2004, which made a successful rendezvous with comet Churyumov-Gerasimenko in May 2014, mapped the cometary surface by remote sensing (Taylor et al. 2017), and performed in-situ observations of the cometary environment (Güttler et al. 2019). Probably the more important achievement of the Rosetta mission was the landing of Philae on the comet's surface in November 2014 (Boehnhardt et al. 2017).

Typically, NEAs and comets have peculiar orbital elements, with non-negligible inclinations and/or eccentricities (Bottke et al. 2000). This implies that the required ΔV for a rendezvous mission that exploits a high-thrust propulsion system is large. Therefore, many of the transfer trajectories of the previously cited missions (including Vega, Giotto and Rosetta) rely on gravity-assist maneuvers that allow the spacecraft to get a sort of "propellantless ΔV ", but significantly increase the transfer times and narrow the launch windows. In order to (partially) overcome these issues, the use of a low-thrust (and high-specific impulse) propulsion system, such as an electric thruster, constitutes a valid alternative to chemical engines, as has been confirmed by some of the aforementioned missions, such as Deep Space 1 and Hayabusa 1 and 2. However, even an electric thruster requires propellant to produce deep space thrust, and the possibility of modifying the orbital parameters is therefore limited by its storage capabilities and by the available onboard power. Thus, the use of propellantless propulsion systems, such as solar sails or electric solar wind sails, could represent a promising option for these advanced mission scenarios.

A solar sail generates propulsive acceleration by exchanging momentum with the impinging photons, thus exploiting the solar radiation pressure. An extensive review of solar sailing and its possible applications may be found in the works by Fu et al. (2016), McInnes (1999), Vulpetti et al. (2015), and Wright (1992). Although the existence of solar radiation pressure is a consequence of Maxwell's equation, the first possible application of a solar sail-based spacecraft was proposed only before the last passing of Halley's comet, when the use of a solar sail was suggested as a solution to rendezvous with the comet in 1986 (Friedman et al. 1978; Sauer 1977; Wright and Warmke 1976). A preliminary study was conducted by NASA, but the project was later dismissed because of the high associated failure risks. More recently, after the successes of solar sailing missions IKAROS (Mori et al. 2010; Tsuda et al. 2011), NanoSail-D2 (Johnson et al. 2011), LightSail-1 (Nye and Greeson 2016) and LightSail-2 (Betts et al. 2019), the utilization of a solar sail as a primary propulsive source in the cruise phase has been proposed for NASA's Near-Earth Asteroid Scout (NEA Scout) (McNutt et al. 2014; Russell Lockett et al. 2019). The renewed interest in solar sailing for the exploration of asteroids is also confirmed by JAXA's planned OKEANOS mission, which should exploit a "solar power sail", that is, a combination of a solar sail and an ion engine powered by solar cells placed on the sail membrane, to reach the Trojan asteroids (Mori et al. 2020). In this context Mengali et al. (2007a, b, c), analyzed the performance of that sort of hybrid propulsion system in a heliocentric mission scenario.

The electric solar wind sail (or E-sail), which is a less mature technology with respect to the solar sail, was originally proposed by Janhunen (2004). The first experimental in-situ data on the E-sail principle should be provided by the Finnish satellite Aalto-1 (Kestilä et al. 2013) during its deorbiting phase. An E-sail propulsion system basically consists of a spinning grid of charged tethers, which generate thrust from the electrostatic interaction with the incoming ions of the solar wind, thus exploiting the solar wind dynamic pressure. Similarly to a solar sail, this propulsive concept could be used to explore asteroids or comets, eliminating the need to perform flybys or store a large amount of propellant.

Even assuming that the spacecraft is propelled by a low-thrust system, matching the orbital parameters of an asteroid or a comet could be a very complex problem to solve, in particular for objects with peculiar orbital parameters, such as those of Halley's comet. A substantial simplification of the mission requirements could be obtained by relaxing the constraints on the terminal phase of the transfer, for example by performing a close encounter instead of a rendezvous. This strategy, denoted "nodal flyby", has been proposed by Perozzi et al. (2001), and applied by Hughes et al. (2004) to a solar sail directed towards a comet. However, the nodal flyby strategy significantly decreases the mission performance level, because it limits the proximity operations and prevents any possibility of harvesting some samples from the target body (Mengali and Quarta 2014). Therefore, the following analysis will only focus on rendezvous missions, without considering the nodal flyby case.

The aim of this chapter is to briefly describe the heliocentric transfer towards an NEA by considering both a classical two-impulse mission scenario, and a more

complex transfer in which a propellantless propulsion system (either a solar sail or an E-sail) is used to produce the required propulsive acceleration.

25.2 Minimum Delta-V to Rendezvous

The classical approach to determine the accessibility of an asteroid or, in general, of a small celestial body denoted a near-Earth object (NEO) consists of calculating of the total (impulsive) velocity change ΔV required to rendezvous with the target object. An accurate estimation of the effective value of the total velocity change can be performed only when the mission profile is well defined, but some strategies exist to provide an estimation that is sufficiently accurate for a preliminary mission design phase.

25.2.1 *Approximate Values with Shoemaker and Helin's Approach*

A simplified strategy to estimate the required ΔV for a rendezvous with an NEA is due to Shoemaker and Helin (1978). This strategy assumes a two-impulse transfer from a low-Earth orbit (LEO) to the asteroid orbit and uses the patched conic approximation, thus neglecting the Earth's orbital eccentricity and assuming that the rendezvous with the asteroid takes place at the apogee of the transfer orbit.

A sketch of the mission scenario is given in Fig. 25.1 One-half of the required inclination change is performed with the first impulse, and the remaining half with the final impulse. To simplify calculations, the impulses (and all the velocities involved in the analysis) are expressed in dimensionless term by normalizing them to the Earth's orbital speed, while distances are normalized to the Earth's orbital radius. We will adopt this same procedure, which amounts to expressing all the variables in heliocentric canonical units. The total required (dimensionless) Δv is used as a figure of merit of the whole transfer. According to Shoemaker and Helin (1978), the value of the dimensionless Δv is therefore

$$\Delta v = \Delta v_I + \Delta v_R \quad (25.1)$$

where Δv_I is the dimensionless velocity variation required to insert the spacecraft into a transfer trajectory from an LEO (using a tangential impulse at perigee of the hyperbolic escape orbit), and Δv_R is the dimensionless velocity variation required to rendezvous with the target asteroid (see Fig. 25.1).

Shoemaker and Helin (1978) specialized the characteristics of the transfer trajectories to the specific class of target asteroids that were known in the 1970s. In so doing, they distinguished three classes of NEAs: Amor (24% of NEAs), Apollo

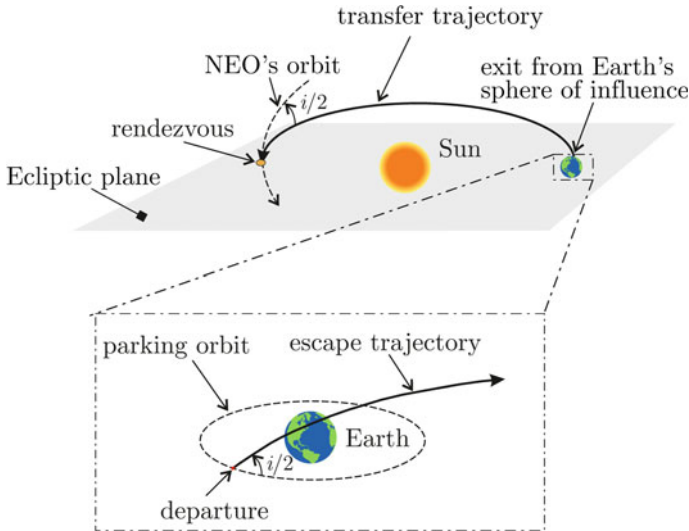


Fig. 25.1 Sketch of the Earth–NEO transfer mission

(54%) and Aten (22%) asteroids. Amor asteroids track heliocentric orbits that are constantly outside the Earth’s orbit, so that their perihelion distance is greater than 1.017au (but smaller than 1.3au). Apollo asteroids, instead, cross the Earth’s orbit, but spend most of their orbital period outside it. They are characterized by an orbital semimajor axis greater than 1au and a perihelion distance smaller than 1.017au. Finally, Aten asteroids may cross the Earth’s orbit, but spend most of their orbital period inside it, since they have an orbital semimajor axis smaller than 1au.

Based on the previously discussed assumptions, the dimensionless initial velocity of the transfer ellipse is given by

$$v_p = \sqrt{\frac{2r_a}{r_a + 1}} \tag{25.2}$$

where r_a denotes the aphelion distance of the asteroid (normalized to the Earth’s orbital radius). The relative velocity v_{dep} of the spacecraft with respect to the Earth at the departure of the transfer orbit is therefore, in the cases of Amor and Apollo asteroids

$$v_{dep} = \sqrt{3 - \frac{2}{r_a + 1} - 2\sqrt{\frac{2r_a}{r_a + 1}} \cos(i/2)} \tag{25.3}$$

where i is the inclination of the asteroid orbit with respect to the ecliptic. Note that the possible inclusion of a corrective term for Earth (nonzero) orbital eccentricity in Eq. (25.3) would only vary the result by less than 1%. When considering an

Aten asteroid, the optimal rendezvous in terms of minimum required Δv would occur at the perihelion, but Shoemaker and Helin (1978) choose to maintain the rendezvous location at aphelion. The authors also assume that the transfer ellipse has a semimajor axis equal to 1 au and is tangent to the asteroid's orbit at aphelion (that is, the rendezvous position). Under these assumptions, Shoemaker and Helin (1978) calculate the relative velocity at departure as

$$v_{\text{dep}} = \sqrt{2 - 2\sqrt{2r_a - r_a^2} \cos(i/2)} \tag{25.4}$$

Accordingly, Eq. (25.3) or (25.4) allow the initial impulse to be found as

$$\Delta v_I = \sqrt{v_{\text{dep}}^2 + v_{\text{esc}}^2} - v_0 \tag{25.5}$$

where v_{esc} is the dimensionless escape velocity from the (LEO) parking orbit and v_0 is the dimensionless spacecraft velocity on the parking orbit. The final velocity variation Δv_R required for a rendezvous maneuver is simply calculated by Carnot's theorem as

$$\Delta v_R = \sqrt{v_a^2 + v_t^2 - 2v_a v_t \cos(i/2)} \tag{25.6}$$

where v_a denotes the spacecraft orbital (dimensionless) velocity at the aphelion of the transfer trajectory, while v_t is the target body (dimensionless) velocity at the rendezvous heliocentric position.

Shoemaker and Helin (1978) also provide expressions of v_a and v_t by distinguishing the cases of Amor, Apollo and Aten asteroids. In the case of an Amor asteroid, the dimensionless velocities v_a and v_t are obtained as

$$v_a = \sqrt{\frac{3}{r_a} - \frac{2}{r_a + 1} - \frac{2}{r_a} \sqrt{\frac{2}{r_a + 1}} \cos(i/2)} \tag{25.7}$$

$$v_t = \sqrt{\frac{3}{r_a} - \frac{1}{a_t} - \frac{2}{r_a} \sqrt{\frac{a_t}{r_a} (1 - e_t^2)}} \tag{25.8}$$

where a_t and e_t denote the (dimensionless) semimajor axis and the eccentricity of the target body heliocentric orbit, respectively. On the other hand, for Apollo asteroids, the velocities are

$$v_a = \sqrt{\frac{3}{r_a} - \frac{2}{r_a + 1} - \frac{2}{r_a} \sqrt{\frac{2}{r_a + 1}}} \tag{25.9}$$

$$v_t = \sqrt{\frac{3}{r_a} - \frac{1}{a_t} - \frac{2}{r_a} \sqrt{\frac{a_t}{r_a} (1 - e_t^2)} \cos(i/2)} \quad (25.10)$$

while the expressions of v_a and v_t for Aten asteroids are

$$v_a = \sqrt{\frac{3}{r_a} - 1 - \frac{2}{r_a} \sqrt{2 - r_a}} \quad (25.11)$$

$$v_t = \sqrt{\frac{3}{r_a} - \frac{1}{a_t} - \frac{2}{r_a} \sqrt{\frac{a_t}{r_a} (1 - e_t^2)} \cos(i/2)} \quad (25.12)$$

Finally, Shoemaker and Helin (1978) propose a semi-empirical formula to calculate the actual, dimensional ΔV (in km/s) required to rendezvous with a NEA, viz.

$$\Delta V = (30\Delta v + 0.5)\text{km/s} \quad (25.13)$$

where Δv is the dimensionless value given by Eq. (25.1).

The approximate approach proposed by Shoemaker and Helin (1978) has been compared with the outputs of the Near-Earth Object Human Space Flight Accessible Targets Study (NHATS). The latter is an accurate study coordinated by NASA and aimed at estimating the effective ΔV required to perform a rendezvous with an NEA (Barbee et al. 2011). More details on NHATS are provided in Sect. 25.2. The results of the comparison between Shoemaker and Helin's method and the outputs from NHATS are discussed by Murphy (2015), who highlights that the results from Shoemaker–Helin equations show systematic errors and a significant scattering of the results with respect to the values provided by NHATS. We agree with the latter conclusion, since the approach proposed by Shoemaker and Helin (1978) may be improved, even just using most of their simplifying assumptions. In particular, in our following discussion the patched conic approach is maintained, and the plane change is again split into equal parts between the initial and the final impulse (see Fig. 25.1). Moreover, if the rendezvous is assumed always to take place at the asteroid's aphelion, there is no distinction among Amor asteroids, Apollo asteroids, and Aten asteroids with an aphelion distance greater than 1 au, since in all these cases the rendezvous occurs at the aphelion of the transfer orbit. Therefore, the only special case is constituted by Aten asteroids with an aphelion distance smaller than the Sun–Earth distance. In this case the rendezvous takes place at the perihelion of the transfer orbit, but the calculation of the total (dimensionless) Δv is not modified, since only the directions of the impulses are affected, but not their magnitude. To sum up, in this sort of modified Shoemaker and Helin's approach that we are going to discuss, the distinction among asteroid families does not affect the required Δv . Finally, an additional term of 0.5 km/s is added to the results to account for orbital perturbation and other uncertainty sources, in analogy with the original Shoemaker and Helin's analysis.

Under the previous assumptions, and maintaining the normalization convention described above, the dimensionless spacecraft velocity at the exit from Earth’s sphere of influence is still given by Eq. (25.2), and the dimensionless velocity at the perigee of the escape hyperbola becomes

$$v_{ph} = \sqrt{v_{dep}^2 + v_{esc}^2} \tag{25.14}$$

where the symbols have the same meaning as those in Eq. (25.5). The value of v_{dep} given by Eq. (25.3) (which is still valid) may be used to calculate the magnitude of the first dimensionless velocity variation Δv_I

$$\Delta v_I = \sqrt{v_{ph}^2 + v_0^2 - 2v_{ph}v_0 \cos(i/2)} \tag{25.15}$$

Note that Eq. (25.5) is valid only if the LEO orbit and the transfer orbit are coplanar (i.e., there is no plane change with the first impulse). The final velocity variation Δv_R is again obtained from Carnot’s theorem as

$$\Delta v_R = \sqrt{v_a^2 + v_t^2 - 2v_a v_t \cos(i/2)} \tag{25.16}$$

The (dimensionless) spacecraft velocity at apheleon v_a and that of asteroid v_t are given by

$$v_a = \sqrt{\frac{2}{r_a(r_a + 1)}} \tag{25.17}$$

$$v_t = \sqrt{\frac{2a_t - r_a}{r_a a_t}} \tag{25.18}$$

Note that Eqs. (25.17) and (25.18) are significantly different from those proposed by Shoemaker and Helin (1978), and do not distinguish the scenarios according to the family of the target asteroid. The total required (dimensional) ΔV for asteroid (or NEO) rendezvous is then estimated, similarly to the Shoemaker and Helin method, as

$$\Delta V = 30(\Delta v_I + \Delta v_R) + 0.5 \text{ km/s} \tag{25.19}$$

where Δv_I and Δv_R are given by Eqs. (25.5) and (25.16), respectively.

25.2.2 *Optimal Orbit-To-Orbit Two-Impulses Transfer*

The NHATS started in September 2010 with the aim of providing an estimation of the required ΔV for a rendezvous with a NEO (in particular, asteroids or comets

passing near the Earth) (Barbee et al. 2011). The aim of NHATS is to identify a set of NEOs that could constitute the target of future manned exploration missions. NHATS is coordinated by NASA, and the analyses are performed by Jet Propulsion Laboratory (JPL) and Goddard Space Flight Center (GSFC).

NHATS is based on some simplifying assumptions. First, similarly to the Shoemaker and Helin approach, only bi-impulsive transfers are considered, with no mid-course maneuvers, gravity assists, or low thrust-propelled arcs. The position of the Earth and of the target NEO are provided by the JPL full precision ephemeris. Then, potential transfer trajectories are identified using a Lambert-based approach and accounting for every gravitational perturbation source. The NEO accessibility is assessed by means of embedded trajectory grids (Barbee et al. 2010). The latter allows a number of candidate transfer trajectories to be generated by varying the departure date with a time-step of 8 h in the 2020–2045 range. NEOs are considered to be accessible only provided the following constraints are met: excess velocity at the exit of Earth's sphere of influence less than or equal to 7.746 km/s (i.e., $v_{\text{dep}} \leq 0.2582$); total mission ΔV less than or equal to 12 km/s; maximum total flight time of 450 days; minimum stay time at the NEO of 8 days; and maximum Earth re-entry velocity of 12 km/s at an altitude of 120 km.

The minimum ΔV trajectory and the minimum flight-time trajectories are then registered for each potential target NEO. For the purpose of this analysis, it is worth remarking that the ΔV associated with each part of the mission is registered during the NHATS analyses. In particular, the outbound ΔV is the total impulse required to escape from the LEO and to rendezvous with the target NEO, so that it essentially coincides with the ΔV evaluated through Eq. (25.13) or (25.19).

The NHATS project has led to the development of a useful and validated tool capable of providing an estimation of the minimum ΔV required to perform a rendezvous with NEOs included in the JPL Small Body Database (SBD), which is constantly updated. A target NEO can be selected by enforcing a number of constraints, such as the maximum flight duration, maximum ΔV , target magnitude, and minimum stay time after the rendezvous.

In order to provide a quick comparison between the Shoemaker and Helin approach, the variant proposed in this chapter, and the NHATS outputs, Table 25.1 shows the ΔV required to reach some NEAs. The results show that, although the refined model is in general more accurate in estimating the actual ΔV for rendezvous, as expected the NHATS analysis is far more reliable, since it considers as potential solutions also trajectories that are significantly different from the over-simplified solution by Shoemaker and Helin. Finally, it is worth remarking that the required total velocity changes are quite large, thus justifying the analysis performed in the next section focused on propellantless propulsion systems.

Table 25.1 Required ΔV (in km/s) to reach some relevant NEAs, estimated with Shoemaker and Helin’s original approach (SH), with the refined approach (R), and with NHATS

	SH	R	NHATS
1943 Anteros	5.405	6.207	8.737
3361 Orpheus	5.501	6.081	6.530
25,143 Itokawa	4.598	4.693	7.298
85,585 Mjolnir	5.578	5.711	7.247
99,942 Apophis	5.648	6.081	4.765
101,955 Bennu	5.058	5.490	5.462
162,173 Ryugu	4.623	5.301	5.850
367,943 Duende	7.255	7.528	6.471

25.3 Propellantless Propulsion Systems

The large values of ΔV required to rendezvous with NEAs and comets strongly support the use of propellantless propulsion systems to provide the required thrust during the cruise (heliocentric) phase. In order to analyze the corresponding mission scenarios, suitable models for expressing the acceleration provided by such advanced propulsion systems are required.

25.3.1 Solar Sail Thrust Model

Since modeling the thrust generated by a solar sail is in principle rather complex, several mathematical tools have been proposed in the literature.

The simpler model is the ideal force model, which assumes that the sail is perfectly flat and constantly lies on a plane (referred to as nominal plane), and that each photon impinging on the sail surface is specularly reflected.

Other more complex models take into account the optical properties of the sail surface (Heaton and Artusio-Glimpse 2015; McInnes 1999), the variations of such optical properties with the temperature (Ancona and Kezerashvili 2017; Kezerashvili 2008, 2014; Mengali et al. 2007a), more accurate reflection models (Vulpetti 2013; Zola et al. 2018), and the time fluctuations of solar radiation pressure (Caruso et al. 2020; Niccolai et al. 2019). Because the purpose of this study is a preliminary mission analysis, the optical force model, which is a good compromise between simplicity and accuracy, will be used for the numerical simulations.

Assuming an optical reflection model, the propulsive acceleration \mathbf{a} generated by a solar sail at a specific distance r from the Sun is given by (Mengali and Quarta 2005)

$$\mathbf{a} = \frac{\beta_{\sigma} \mu_{\odot}}{2r^2} (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) [b_1 \hat{\mathbf{r}} + (b_2 \hat{\mathbf{r}} \cdot \hat{\mathbf{n}} + b_3) \hat{\mathbf{n}}] \tag{25.20}$$

where $\beta_\sigma \triangleq \tilde{\sigma}/(m/S)$ is the dimensionless sail loading (Dachwald et al. 2006), with $e \tilde{\sigma} \triangleq 1.593 \text{ g/m}^2$, m and S denoting the total spacecraft mass and the sail surface area, respectively. Also, b_1, b_2 and b_3 are the dimensionless force coefficients related to the thermo-optical properties of the reflective film, $\hat{\mathbf{r}} \triangleq \mathbf{r}/r$ is the radial unit vector, and $\hat{\mathbf{n}}$ is the unit vector normal to the sail in the direction opposite to the Sun.

A typical performance parameter for a solar sail is the characteristic acceleration a_c , that is, the acceleration experienced by a Sun-facing solar sail. In this case, the sail reference plane is oriented perpendicular to the Sun–spacecraft line (that is, $\hat{\mathbf{r}} \equiv \hat{\mathbf{n}}$) at a distance $r_0 \triangleq 1 \text{ au}$. From Eq. (25.20) it is found that

$$a_c = \frac{(b_1 + b_2 + b_3)\beta_\sigma \mu_\odot}{2r_0^2} \quad (25.21)$$

The optical force model for a sail with a highly reflective aluminum-coated front side and a highly emissive chromium-coated back side is, instead, characterized by the coefficients (Mengali and Quarta 2005) $b_1 = 0.1728$, $b_2 = 1.6544$ and $b_3 = -0.0109$. In an ideal force model, instead, the coefficients are $b_1 = b_3 = 0$ and $b_2 = 2$.

25.3.2 Solar Sail Optimal Transfers

The equations of motion for a spacecraft propelled by a solar sail are written in a Cartesian heliocentric inertial frame $\mathcal{T}_\odot(\hat{\mathbf{i}}_x, \hat{\mathbf{i}}_y, \hat{\mathbf{i}}_z)$ as

$$\dot{\mathbf{r}} = \mathbf{v} \quad (25.22)$$

$$\dot{\mathbf{v}} = -\frac{\mu_\odot}{r^3}\mathbf{r} + \mathbf{a} \quad (25.23)$$

where \mathbf{r} and \mathbf{v} are the spacecraft position and velocity, while \mathbf{a} is the propulsive acceleration generated by the solar sail.

For numerical simulation purposes, it is useful to introduce the state vector \mathbf{x} of modified equinoctial elements (Walker 1986; Walker et al. 1985), defined as

$$\mathbf{x} \triangleq [p, f, g, h, k, L]^T \quad (25.24)$$

where p is the semilatus rectum and L is the true longitude. As a result (Betts 2000), the equations of motion (25.22)–(25.23) become

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{c} \quad (25.25)$$

where

$$\mathbf{c} \triangleq \left[0, 0, 0, 0, 0, \sqrt{\mu_{\odot} p} \left(\frac{1 + f \cos L + g \sin L}{p} \right)^2 \right]^T \quad (25.26)$$

and $A \in \mathbb{R}^{6 \times 3}$ is a suitable matrix whose generic entry will be referred to as A_{ij} . It may be shown (Betts 2000) that $A_{11} = A_{13} = A_{41} = A_{42} = A_{51} = A_{52} = A_{61} = A_{62} = 0$, while

$$A_{12} = \frac{2p}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.27)$$

$$A_{21} = \sin L \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.28)$$

$$A_{22} = \frac{(2 + f \cos L + g \sin L) \cos L + f}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.29)$$

$$A_{23} = -\frac{g(h \sin L - k \cos L)}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.30)$$

$$A_{31} = -\cos L \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.31)$$

$$A_{32} = \frac{(2 + f \cos L + g \sin L) \sin L + g}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.32)$$

$$A_{33} = \frac{f(h \sin L - k \cos L)}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.33)$$

$$A_{43} = \frac{(1 + h^2 + k^2) \cos L}{2(1 + f \cos L + g \sin L)} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.34)$$

$$A_{53} = \frac{(1 + h^2 + k^2) \sin L}{2(1 + f \cos L + g \sin L)} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.35)$$

$$A_{63} = \frac{h \sin L - k \cos L}{1 + f \cos L + g \sin L} \sqrt{\frac{p}{\mu_{\odot}}} \quad (25.36)$$

It is worth remarking that the modified equinoctial elements can be transformed into the classical orbital elements (a , e , i , Ω , ω and ν), which have the advantage of possessing a clear physical meaning, and to the spacecraft position and velocity components expressed with respect to $\mathcal{T}_{\odot}(\hat{i}_x, \hat{i}_y, \hat{i}_z)$.

The Sun–spacecraft distance r required for the calculation of the gravitational and propulsive acceleration in Eqs. (25.20) and (25.23) can be written in terms of modified equinoctial elements by means of the following modified polar equation

$$r = \frac{p}{1 + f \cos L + g \sin L} \quad (25.37)$$

The introduction of the modified equinoctial elements into the equations of motion allows the computational time necessary for the solar-sail trajectory integration to be significantly reduced, and the singularities that could arise using the classical orbital elements to be avoided.

Consider now a rotating radial-tangential-normal $\mathcal{T}_{\text{RTN}}(\hat{i}_R, \hat{i}_T, \hat{i}_N)$ coordinate frame (Betts 2000), with unit vectors defined as

$$\hat{i}_R = \frac{\mathbf{r}}{r}, \hat{i}_N = \frac{\mathbf{r} \times \mathbf{v}}{\|\mathbf{r} \times \mathbf{v}\|}, \hat{i}_T = \hat{i}_N \times \hat{i}_R \quad (25.38)$$

which are sketched in Fig. 25.2a. The orientation of $\hat{\mathbf{n}}$ in the rotating frame $\mathcal{T}_{\text{RTN}}(\hat{i}_R, \hat{i}_T, \hat{i}_N)$ is univocally defined by the cone angle $\alpha \in [0, \pi/2]$ rad and the clock angle $\delta \in [-\pi, \pi]$ rad, see Fig. 25.2b, as (Wie 2005)

$$[\hat{\mathbf{n}}]_{\mathcal{T}_{\text{RTN}}} = [\cos \alpha, \sin \alpha \cos \delta, \sin \alpha \sin \delta]^T \quad (25.39)$$

Substituting Eq. (25.39) into (25.20), the propulsive acceleration components in the rotating frame $\mathcal{T}_{\text{RTN}}(\hat{i}_R, \hat{i}_T, \hat{i}_N)$ are obtained as

$$[\mathbf{a}]_{\mathcal{T}_{\text{RTN}}} = \frac{\beta_\sigma \mu_\odot}{2r^2} \cos \alpha \begin{bmatrix} b_1 + (b_2 \cos \alpha + b_3) \cos \alpha \\ (b_2 \cos \alpha + b_3) \sin \alpha \cos \delta \\ (b_2 \cos \alpha + b_3) \sin \alpha \sin \delta \end{bmatrix} \quad (25.40)$$

where r depends on the equinoctial elements according to Eq. (25.37). Finally, substituting Eq. (25.40) into Eq. (25.25), the spacecraft dynamics can be described through six first-order nonlinear differential equations, which can be formally written as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \quad (25.41)$$

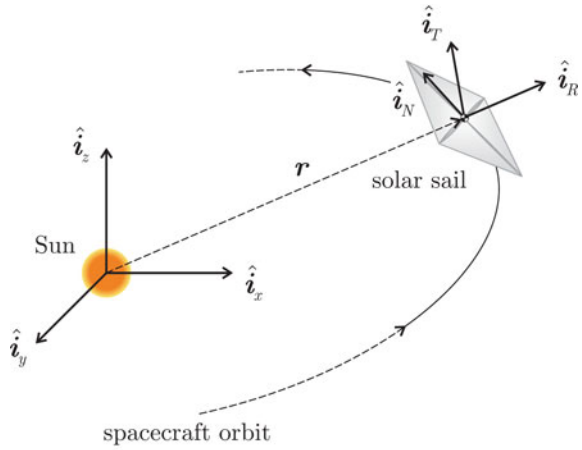
where

$$\mathbf{u} \triangleq [\alpha, \delta]^T \quad (25.42)$$

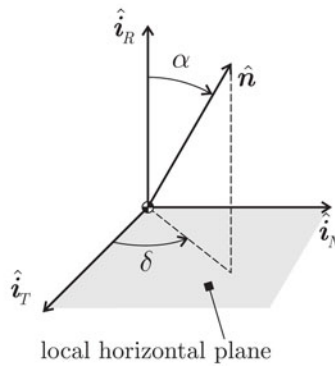
represents the solar sail control vector.

The aim of the following analysis is the determination of the optimal transfer for a solar sail-based spacecraft from Earth to a generic NEO. In this case the optimal transfer coincides with the minimum flight-time transfer since there is no propellant consumption to take into account. A direct interplanetary insertion of the spacecraft from Earth with zero hyperbolic excess energy relative to the planet is assumed. The condition of zero excess velocity is representative of a sail deployment on a parabolic

Fig. 25.2 Reference frames and control angles for a solar sail-based mission scenario



(a) Reference frames.



(b) Control angles.

Earth escape trajectory and provides conservative results as far as the rendezvous flight time is concerned.

In mathematical terms the problem is to find the optimal control law $\mathbf{u}(t)$ (where $t \in [0, t_f]$) that minimizes the time t_f necessary to transfer the spacecraft from an initial \mathbf{x}_0 to a final \mathbf{x}_f prescribed state. This amounts to maximizing the performance index:

$$J = -t_f \tag{25.43}$$

Using an indirect approach, from Eq. (25.25) the Hamiltonian of the system is

$$H \triangleq (A\mathbf{a}) \cdot \boldsymbol{\lambda} + \mathbf{c} \cdot \boldsymbol{\lambda} \tag{25.44}$$

where $\lambda \triangleq [\lambda_p, \lambda_f, \lambda_g, \lambda_h, \lambda_k, \lambda_L]^T$ is the adjoint vector, whose time derivative is given by the Euler–Lagrange equation:

$$\dot{\lambda} \triangleq -\frac{\partial H}{\partial \mathbf{x}} \quad (25.45)$$

where the derivative with respect to a vector is a gradient operator. The explicit expression of the Euler–Lagrange equation is rather involved and is not reported here for the sake of conciseness. The optimal value of the control angles is obtained by invoking the Pontryagin’s maximum principle, that is, by maximizing, at any time, the Hamiltonian H of Eq. (25.44). In particular, as far as the clock angle δ is concerned, by enforcing the necessary condition $\partial H/\partial \delta = 0$ we obtain (Bryson et al. 1975):

$$\sin \delta = A_{23}\lambda_f + A_{33}\lambda_g + A_{43}\lambda_h + A_{53}\lambda_k + A_{63}\lambda_L \quad (25.46)$$

$$\cos \delta = A_{23}\lambda_g + A_{12}\lambda_p + A_{22}\lambda_f \quad (25.47)$$

where the terms A_{ij} are defined in Eqs. (25.27)–(25.36). On the other hand, the second necessary condition $\partial H/\partial \alpha = 0$ does not provide a closed form solution for the clock angle α . Therefore, the maximization of H with respect to α is performed by means of a numerical algorithm based on golden-section search and parabolic interpolation (Forsythe et al. 1977).

The boundary-value problem associated to the variational problem is constituted by the equations of motion (25.41) and by the Euler–Lagrange Eq. (25.45). The corresponding 12 boundary conditions are related to the desired spacecraft position and velocity at the initial (t_0) and final (t_f) time. In this discussion, the angular positions on the Earth’s heliocentric orbit and on the target body’s orbit are left unconstrained (which implies that two boundary conditions are given on the adjoint variable λ_L), so that the optimization algorithm selects the optimal true anomaly on the departure and on the target orbit. This amounts to neglecting the determination of the launch window, since the actual positions of the Earth and of the target NEO are not taken into account in this preliminary analysis. The transversality condition $H(t_f) = 1$, required to determine the value of t_f , completes the differential problem (Bryson et al. 1975). In all of the simulations, the differential equations have been integrated in double precision using a variable-order Adams–Bashforth–Moulton solver with absolute and relative errors of 10^{-10} . The boundary-value problem has been solved by means of a hybrid numerical technique that combines the use of genetic algorithms to obtain a rough estimate of the adjoint variables, with gradient-based and direct methods to refine the solution (Mengali and Quarta 2005).

The results obtained with the previous optimization procedure are reported in Table 25.2 for all the asteroids included in Table 25.1. The spacecraft is assumed to be propelled by a solar sail with a characteristic acceleration $a_c = 1 \text{ mm/s}^2$, which

Table 25.2 Orbit-to-orbit optimal transfer (flight time Δt , starting true anomaly ν_0 , final true anomaly ν_f) to reach some relevant NEAs, for a solar sail-based spacecraft with an ideal force model ($a_c = 1 \text{ mm/s}^2$)

	ν_0 [deg]	ν_f [deg]	Δt [days]
1943 Anteros	73.7	177.2	387.4
3361 Orpheus	72.6	283.2	384.1
25,143 Itokawa	122.7	102.7	126.6
85,585 Mjolnir	355.5	132	163.5
99,942 Apophis	284	165.5	119.8
101,955 Bennu	328.2	145.6	171.3
162,173 Ryugu	315	117.5	184.2
367,943 Duende	200	200	263.2

corresponds to a near- or mid-term technology level, since the NEA Scout characteristic acceleration is about 0.6 mm/s^2 (McNutt et al. 2014). It is worth remarking that all of the analyzed NEOs, which have very different orbital parameters, are reachable with a flight time of less than 13 months. Clearly, a more accurate mission design should take into account the Earth’s and the target body’s ephemeris in order to determine the optimal launch window, but these results are very promising for near- and mid-term space missions equipped with solar sails and directed towards NEOs.

The optimal transfer trajectories for asteroids in Table 25.2 may be observed in Fig. 25.3, which shows their projection on the ecliptic. Finally, the optimal control laws involving the cone and the clock angles are given in Figs. 25.4 and 25.5, respectively. It can be observed from Fig. 25.4 that the time histories of the optimal cone angles are smooth and do not require large changes in a short time interval, with the transfer to asteroid Duende being the only exception. The time variations of the clock angle shown in Fig. 25.5 are all very smooth, since the only apparent discontinuity is just the variation from π rad to $-\pi$ rad (or vice versa), which is caused by a mathematical convention and does not require a physical change in the control angle. It is therefore evident that the optimal time histories of the control angles require small variations, which could in principle be generated with small control torques.

25.3.3 Electric Solar Wind Sail Thrust Model

The propulsive acceleration vector \mathbf{a} generated by an E-sail can be calculated by means of the model recently proposed by Huo et al. (2018). Using the same nomenclature as that of Eq. (25.20), the propulsive acceleration is

$$\mathbf{a} = \tau \frac{a_c}{2} \left(\frac{r_\oplus}{r} \right) [\hat{\mathbf{r}} + (\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}}] \tag{25.48}$$

where it is assumed that the E-sail is flat, axially symmetric, and its grid is made up by more than two wires. In Eq. (25.48), $\tau \in \{0, 1\}$ is a dimensionless switching parameter that allows the electron gun to be switched either on ($\tau = 1$) or off ($\tau =$

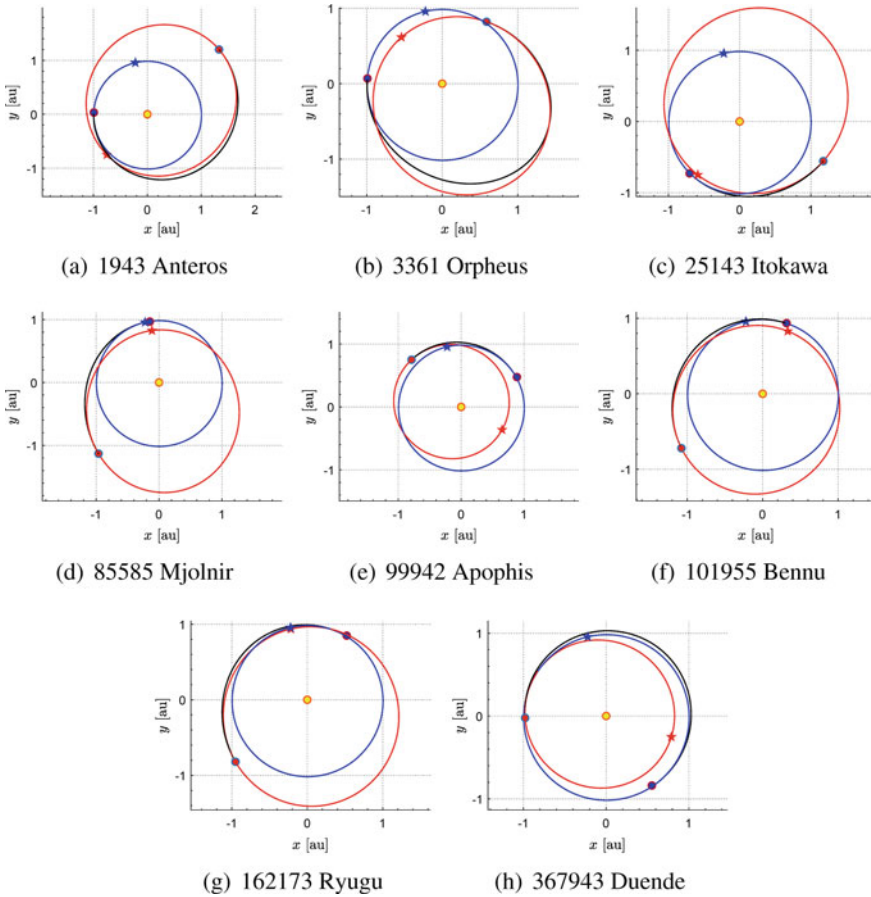


Fig. 25.3 Projection on the ecliptic of optimal ideal solar sail-based transfer trajectories towards NEOs of Table 25.2 (black line = transfer trajectory; blue line = Earth’s orbit; red line = NEO’s orbit; star = perihelion position; dot = rendezvous position)

0), and a_c is the characteristic acceleration, with the same definition as that used for a solar sail-based mission. Note that Eq. (25.48) implies that the E-sail propulsive acceleration a belongs to the plane defined by the normal unit vector \hat{n} and the radial unit vector \hat{r} . The thrust model of Eq. (25.48), which is in accordance with the results discussed by Yamaguchi and Yamakawa (2016), can be simplified to obtain a good compromise between simplicity and accuracy. In fact, using the approach proposed by Janhunen (2010), a simplified thrust model is given by

$$a = a_c \tau \left(\frac{r_\oplus}{r} \right) \hat{a} \text{ with } \arccos(\hat{a} \cdot \hat{r}) \leq \alpha_{\max} \tag{25.49}$$

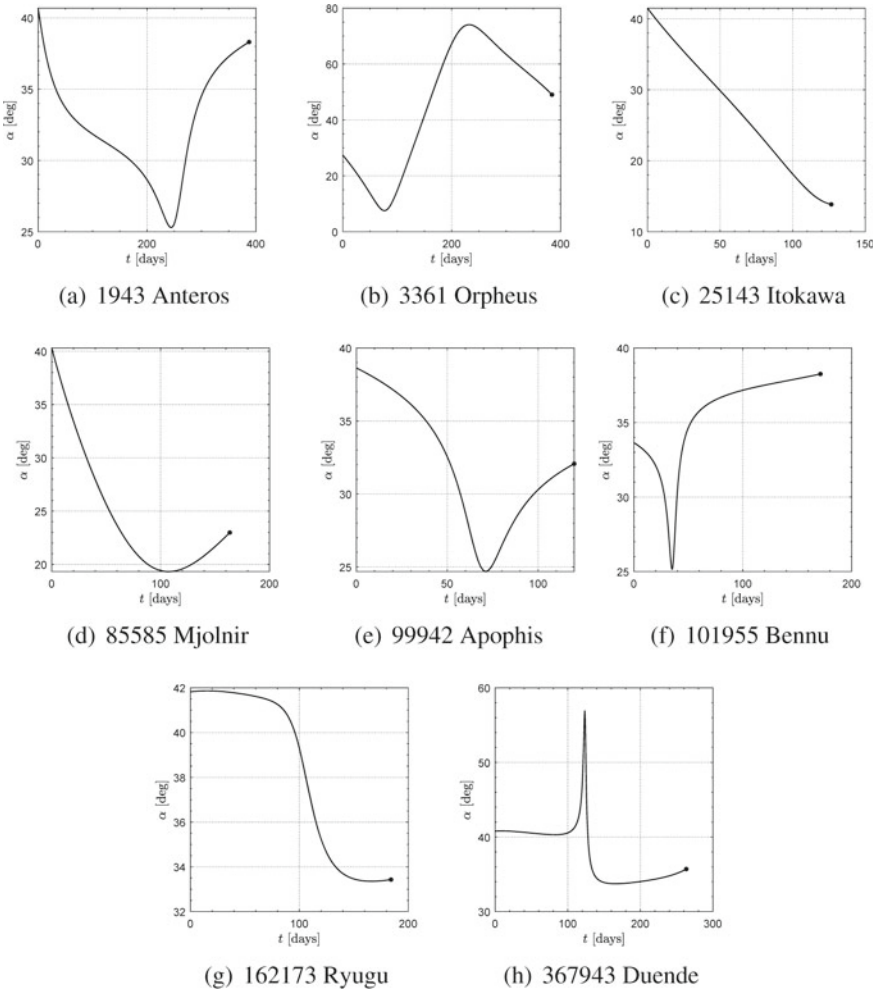


Fig. 25.4 Optimal cone angle control law $\alpha = \alpha(t)$ for ideal solar sail-based transfers ($a_c = 1 \text{ mm/s}^2$) towards NEOs of Table 25.2. The dot denotes the rendezvous condition

where $\alpha_{\max} = 35 \text{ deg}$ is the maximum admissible value of the thrust (i.e., the so-called cone) angle. The expression (25.49) will be used in the remainder of this analysis to model the propulsive acceleration of an E-sail-based spacecraft. Figure 25.6a, b show a scheme of the fundamental reference frames and of the control angles involved in the analysis, respectively.

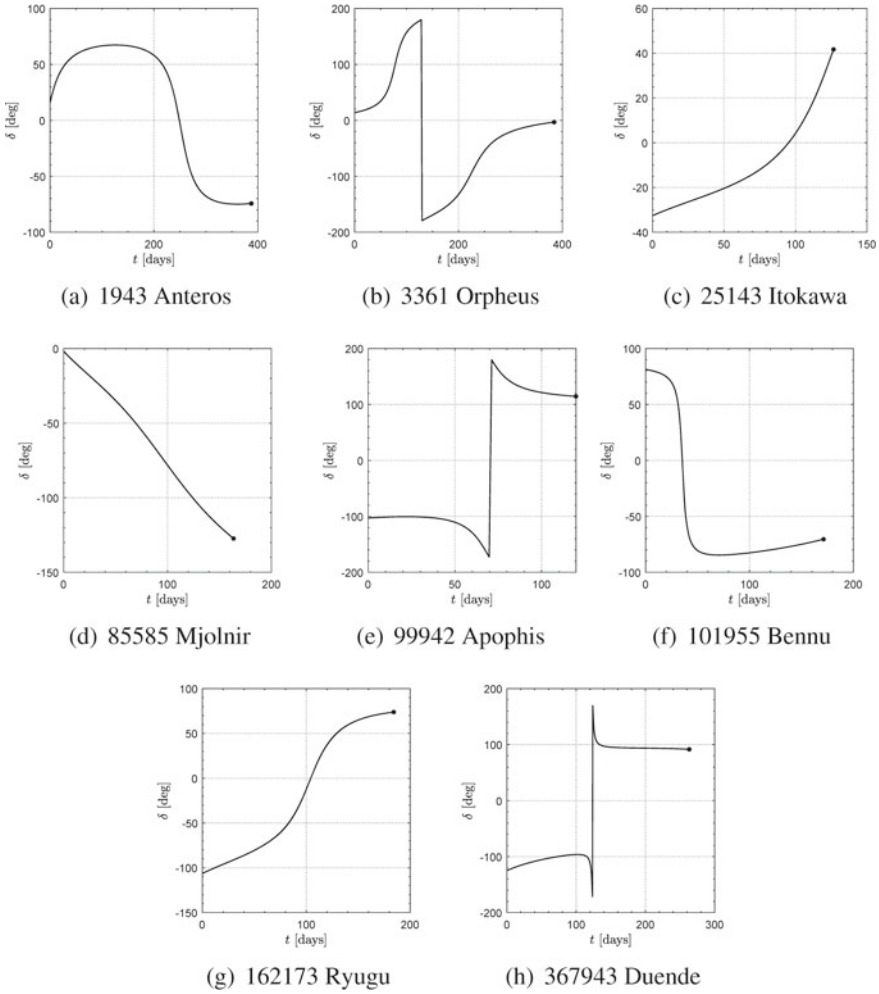
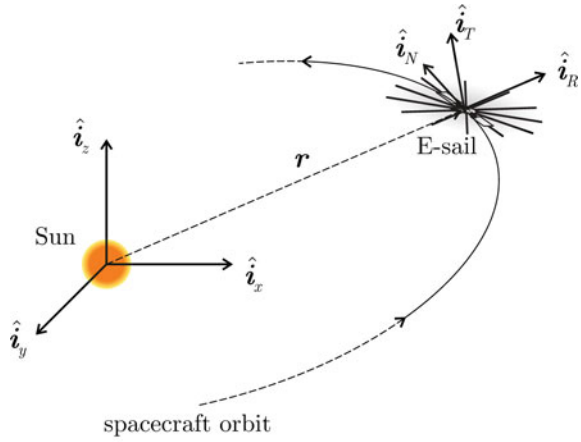


Fig. 25.5 Optimal clock angle control law $\delta = \delta(t)$ for ideal solar sail-based transfers ($a_c = 1 \text{ mm/s}^2$) towards NEOs of Table 25.2. The dot denotes the rendezvous condition

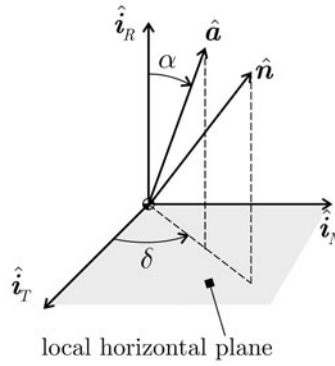
25.3.4 Electric Solar Wind Sail Optimal Transfers

When considering an optimal transfer to an NEO by means of an E-sail, we adopt the same approach as that discussed in Sect. 25.3.2. The main difference between the two scenarios is the thrust model used to express the propulsive acceleration, which is given by Eq. (25.49) in the case of an E- sail-based spacecraft. Under this assumption, the expression of \mathbf{a} in $\mathcal{T}_{RTN}(\hat{i}_R, \hat{i}_T, \hat{i}_N)$ is given by

Fig. 25.6 Reference frames and control angles for an E-sail-based mission scenario



(a) Reference frames.



(b) Control angles.

$$[a]_{\mathcal{T}_{RTN}} = a_c \tau \left(\frac{r_\oplus}{r} \right) \begin{bmatrix} \cos \alpha \\ \sin \alpha \cos \delta \\ \sin \alpha \sin \delta \end{bmatrix} \tag{25.50}$$

where, unlike the solar sail case, the control angles are the thrust angle α and the thrust clock angle δ , which denote the orientation of the propulsive acceleration vector direction with respect to $\mathcal{T}_{RTN}(\hat{i}_R, \hat{i}_T, \hat{i}_N)$

Since the propulsive acceleration, the radial unit vector, and the normal unit vector are all coplanar (see Eq. (25.48)), the thrust clock angle used in Eq. (25.50) coincides with the definition of solar-sail clock angle used in Eq. (25.40)—see also Fig. 25.2b. As regards the thrust angle, a general rule of thumb is that it is about one-half of the cone angle. However, in this analysis, the thrust angle is taken as a control variable; the relation between the thrust angle and the E-sail cone angle is discussed in depth by Yamaguchi and Yamakawa (2016) and Huo et al. (2018).

Note that in the E-sail-based scenario the switching parameter τ constitutes a third control variable. In practice, this implies that the E-sail is switched on ($\tau = 1$) only if the maximum value of the Hamiltonian obtained with the optimal combination of the thrust and the clock angles is greater than zero, while otherwise it is switched off ($\tau = 0$), and the spacecraft tracks a Keplerian arc.

The outputs of the optimization algorithm discussed in this section are listed in Table 25.3 for the same group of asteroids as in Table 25.2 and assuming an E-sail with $a_c = 1 \text{ mm/s}^2$. The results in terms of flight time are comparable with those obtained by a solar sail-based spacecraft, (see Table 25.2), suggesting that the E-sail is another promising option for the exploration of NEOs, in particular when a significant portion of the trajectory is farther from the Sun than the Earth, due to the different variation of the thrust magnitude with the Sun–spacecraft distance when compared to that of a solar sail. Figure 25.7 shows the projection of the corresponding optimal trajectories on the ecliptic. Note that in each transfer trajectory there is a branch in which the E-sail grid is switched off (i.e., $\tau = 0$ in Eq. (25.49)), which corresponds to a Keplerian arc, as previously stated. This observation is confirmed by the time histories of the thrust and clock angles, which are shown in Figs. 25.8 and 25.9, respectively, where the Keplerian arcs are denoted by gray regions. Moreover, it is interesting that the optimal value of the cone angle is almost constantly equal to its maximum allowable value of 35 deg, which corresponds to the maximum circumferential acceleration component that the E-sail can generate. Finally, it may be observed from both Figs. 25.8 and 25.9 that the optimal time histories of the control angles are quite smooth, so that they may be generated by suitable control torques.

The generation of such torques is a topic beyond the scopes of this analysis, but a possible control strategy for an axially symmetric E-sail is discussed by Bassetto et al. (2018, 2019).

Table 25.3 Orbit-to-orbit optimal transfer (flight time Δt , starting true anomaly v_0 , final true anomaly v_f) to reach some relevant NEAs, for a E-sail-based spacecraft ($a_c = 1 \text{ mm/s}^2$)

	v_0 [deg]	v_f [deg]	Δt [days]
1943 Anteros	88.1	108.8	189
3361 Orpheus	225	54.7	206
25,143 Itokawa	178	283	426
85,585 Mjolnir	202	75	221
99,942 Apophis	333.7	288	204
101,955 Bennu	331.2	119.4	127.6
162,173 Ryugu	343	111	148
367,943 Duende	92.5	88.7	245.5

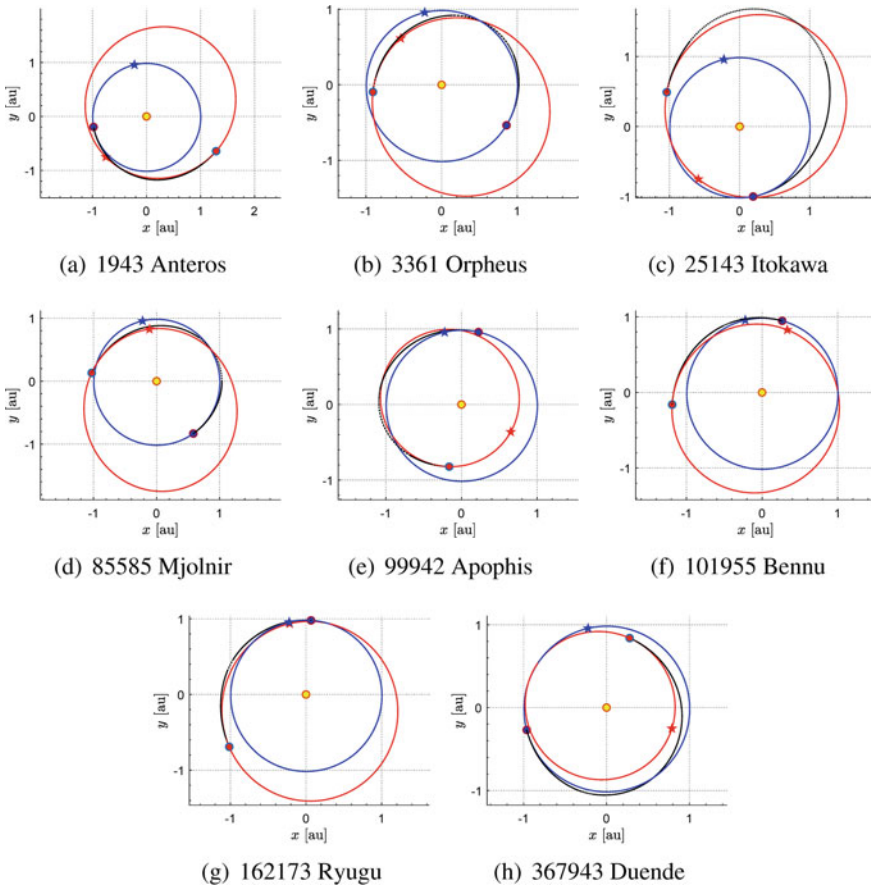


Fig. 25.7 Projection on the Ecliptic of optimal E-sail-based transfer trajectories ($a_c = 1 \text{ mm/s}^2$) towards NEOs of Table 25.3 (black continuous line = propelled transfer trajectory; black dashed line = Keplerian arc; blue line = Earth's orbit; red line = NEA's orbit; star = perihelion position; dot = rendezvous position)

25.4 Conclusions

The exploration of near-Earth objects, such as asteroids or comets, is attracting growing interest among the scientific community. However, as this analysis has shown, the values of total velocity change required for a rendezvous with these objects could be very large, due to their peculiar orbital elements. Therefore, the possibility of performing an exploration mission using propellantless propulsion systems, such as a solar sail or an electric solar wind sail, as thrust sources for the cruise phase seems very appealing.

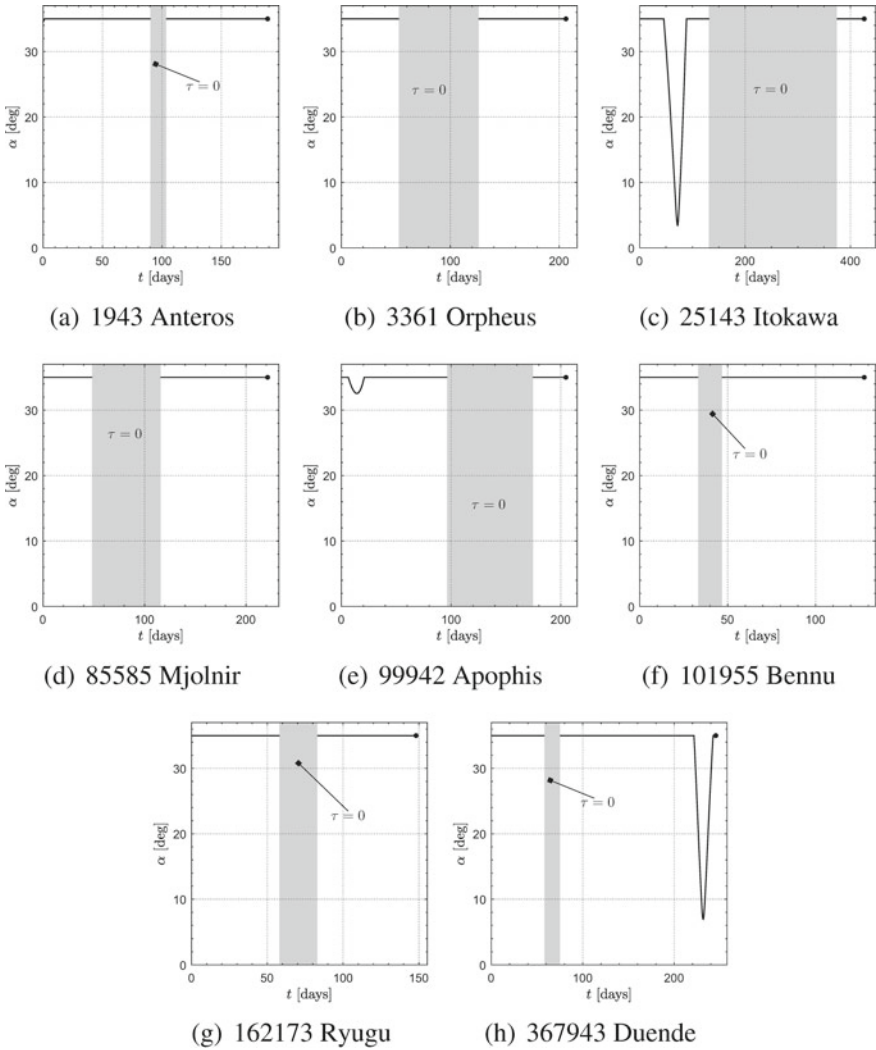


Fig. 25.8 Optimal thrust angle control law $\alpha = \alpha(t)$ for E-sail-based transfers ($a_c = 1 \text{ mm/s}^2$) towards NEOs of Table 25.3. The dot denotes the rendezvous condition

The results obtained have shown that a spacecraft equipped with a near- or mid-term solar sail or electric sail could perform a rendezvous with many near-Earth objects in a reasonable time and without requiring gravity-assist maneuvers, thus significantly simplifying the mission design phase.

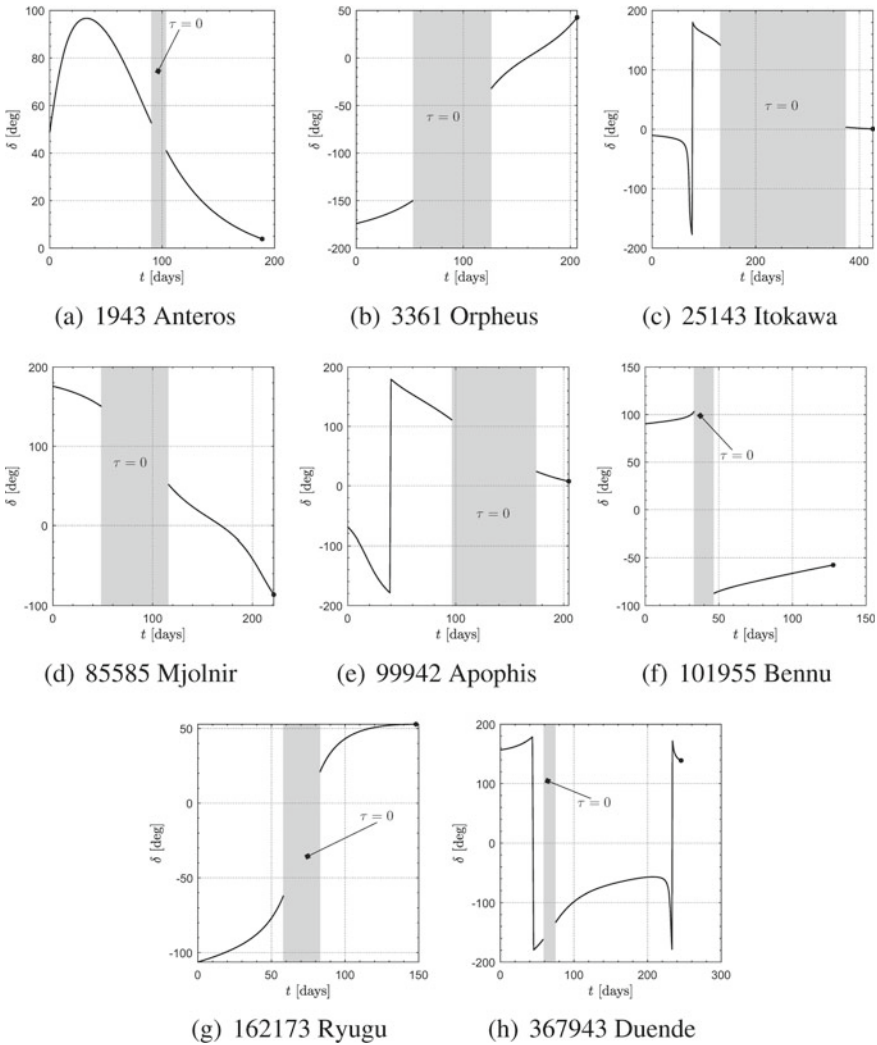


Fig. 25.9 Optimal clock angle control law $\delta = \delta(t)$ for E-sail-based transfers ($a_c = 1 \text{ mm/s}^2$) towards NEOs of Table 25.3. The dot denotes the rendezvous condition

References

Adams, E. et al. 2019. Double asteroid redirection test: The earth strikes back. In: 2019 *IEEE Aerospace Conference*. Big Sky, MT, USA. <https://doi.org/10.1109/AERO.2019.8742007>.

Ancona, E., and R.Y. Kezerashvili. 2017. Temperature restrictions for materials used in aerospace industry for the near-Sun orbits. *Acta Astronautica* 140: 565–569. <https://doi.org/10.1016/j.actastro.2017.09.002>.

Baker, J. 2006. The Falcon has landed. *Science* 312. <https://doi.org/10.1126/science.312.5778.1327>.

- Barbee, B.W. et al. 2010. A comprehensive ongoing survey of the near-earth asteroid population for human mission accessibility. In: *AIAA Guidance, Navigation, and Control Conference*. Toronto, ON, Canada. <https://doi.org/10.2514/6.2010-8368>.
- Barbee, B.W. et al. 2011. Methodology and results of the Near-Earth Object (NEO) Human Space Flight (HSF) Accessible Targets Study (NHATS). In: *AAS/AIAA Astrodynamics Specialist Conference*. Girdwood, AK, USA, pp. 595–614.
- Barucci, M., E. Dotto, and A. Levasseur-Regourd. 2011. Space missions to small bodies: Asteroids and cometary nuclei. In: *Astronomy and Astrophysics Review* 19.1. <https://doi.org/10.1007/s00159-011-0048-2>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2018. Thrust and torque vector characteristics of axially-symmetric E-sail. *Acta Astronautica* 146: 134–143. <https://doi.org/10.1016/j.actaastro.2018.02.035>.
- Bassetto, M., G. Mengali, and A.A. Quarta. 2019. Stability and control of spinning electric solar wind sail in helio- stationary orbit. *Journal of Guidance, Control, and Dynamics* 42(2), 425–431. <https://doi.org/10.2514/1.G003788>.
- Belton, M.J.S., et al. 1992. Galileo encounter with 951 Gaspra: First pictures of an asteroid. *Science* 257: 1647–1652. <https://doi.org/10.1126/science.257.5077.1647>.
- Betts, B., D.A. Spencer, J.M. Bellardo, et al. (2019). LightSail 2: controlled solar sail propulsion using a CubeSat. In: *70th International Astronautical Congress*. Washington, DC, USA.
- Betts, J.T. 2000. Very low-thrust trajectory optimization using a direct SQP method. *Journal of Computational and Applied Mathematics* 120(1): 27–40. [https://doi.org/10.1016/S0377-0427\(00\)00301-0](https://doi.org/10.1016/S0377-0427(00)00301-0).
- Boehnhardt, H. et al. 2017. The Philae lander mission and science overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375: 2097. <https://doi.org/10.1098/rsta.2016.0248>.
- Boice, D.C. and R. Wegmann. 2007. The Deep Space 1 encounter with comet 19P/Borrelly. *Advances in Space Research* 39(3): 407–412. <https://doi.org/10.1016/j.asr.2003.02.092>.
- Bottke, W.F.J. et al. 2000. Understanding the distribution of near-Earth asteroids. *Science* 288(5474): 2190–2194. <https://doi.org/10.1126/science.288.5474.2190>.
- Brownlee, D. 2014. The Stardust mission: Analyzing samples from the edge of the solar system. *Annual Review of Earth and Planetary Sciences* 42: 179–205. <https://doi.org/10.1146/annurev-earth-050212-124203>.
- Bryson, A.E., and Y.C. Ho. 1975. *Applied Optimal Control*. New York, NY: Hemisphere Publishing Corporation. Chap. 2, pp. 71–89.
- Carnelli, I. et al. 2019. HERA mission to asteroid Didymos: ESA contribution to the AIDA international collaboration. In: *70th International Astronautical Congress*. Washington, DC, USA.
- Caruso, A. et al. 2020. Solar sail optimal control with solar irradiance fluctuations. *Advances in Space Research* 67(9): 2776–2783, May 2021. <https://doi.org/10.1016/j.asr.2020.05.037>.
- Cheng, A.F. et al. 2020. DART mission determination of momentum transfer: Model of ejecta plume observations. *Icarus* 352. <https://doi.org/10.1016/j.icarus.2020.113989>.
- Dachwald, B. et al. 2006. Parametric model and optimal control of solar sails with optical degradation. *Journal of Guidance, Control, and Dynamics* 29(5), 1170–1178. <https://doi.org/10.2514/1.20313>.
- Enos, H.L., and D.S. Lauretta 2019. A rendezvous with asteroid Bennu. *Nature Astronomy* 3(4): 363. <https://doi.org/10.1038/s41550-019-0739-6>.
- Forsythe, G.E., A.M. Michael, and C.B. Moler. 1977. *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice- Hall, pp. 192–235.
- Friedman, L. et al. 1978. Solar sailing—the concept made realistic. In: *AIAA 16th Aerospace Sciences Meeting*. AIAA Paper 78–82. Huntsville, AL.
- Fu, B., E. Sperber, and F. Eke. 2016. Solar sail technology—A state of the art review. *Progress in Aerospace Sciences* 86: 1–19. <https://doi.org/10.1016/j.paerosci.2016.07.001>.

- Güttler, C. et al. 2019. Synthesis of the morphological description of cometary dust at comet 67P/Churyumov-Gerasimenko. In: *Astronomy and Astrophysics* 6330. <https://doi.org/10.1051/0004-6361/201834751>.
- Heaton, A.F. and A.B. Artusio-Glimpse. 2015. An update to NASA reference sail thrust model. In: *AIAA SPACE 2015 Conference and Exposition*. Pasadena, CA, USA. <https://doi.org/10.2514/6.2015-4506>.
- Hughes, G.W. and C.R. McInnes. 2004. Small-body encounters using solar sail propulsion. *Journal of Spacecraft and Rockets* 41(1): 140–150. <https://doi.org/10.2514/1.9277>.
- Huo, M., G. Mengali, and A.A. Quarta. 2018. Electric sail thrust model from a geometrical perspective. *Journal of Guidance, Control, and Dynamics* 41(3), 734–740. <https://doi.org/10.2514/1.G003169>.
- Itoh, T. and K. Hirao. 1986. The Sakigake and Suisei encounters with comet Halley. *Geophysical Research Letters* 13(8): 817–819. <https://doi.org/10.1029/GL013i008p00817>.
- Janhunen, P. 2004. Electric sail for spacecraft propulsion. *Journal of Propulsion and Power* 20(4): 763–764. <https://doi.org/10.2514/1.8580>.
- Janhunen, P. 2010. The electric solar wind sail status report. In: *European Planetary Science Congress 2010*. Vol. 5. Paper EPSC 2010–297. Rome, Italy.
- Johnson, L. et al. 2011. NanoSail-D: a solar sail demonstration mission. *Acta Astronautica* 68(5–6): 571–575. <https://doi.org/10.1016/j.actaastro.2010.02.008>.
- Kaneda, E. et al. 1986. Observation of comet Halley by the ultraviolet imager of Suisei. *Nature* 321(6067): 297–299. <https://doi.org/10.1038/321297a0>.
- Kestilä, A., T. Tikka, P. Peitso, et al. 2013. Aalto-1 nanosatellite—technical description and mission objectives. *Geoscientific Instrumentation, Methods and Data Systems* 2: 121–130. <https://doi.org/10.5194/gi-2-121-2013>.
- Kezerashvili, R.Y. (2008). Solar sail interstellar travel: 1. Thickness of solar sail films. *Journal of the British Interplanetary Society* 61(11): 430–439.
- Kezerashvili, R.Y. 2014. Solar sail: materials and space environmental effects. In: ed. by M. Macdonald, *Advances in Solar Sailing*. Berlin: Springer Praxis. Chap. 3, pp. 573–592. https://doi.org/10.1007/978-3-642-34907-2_36.
- Kissel, J. et al. 1986a. Composition of comet halley dust particles from giotto observations. *Nature* 321(6067): 336–337. <https://doi.org/10.1038/321336a0>.
- Kissel, J. et al. 1986b. Composition of comet halley dust particles from vega observations. *Nature* 321(6067): 280–282. <https://doi.org/10.1038/321280a0>.
- Lauretta, D.S., et al. 2017. OSIRIS-REx: Sample return from asteroid (101955) Benu. *Space Science Reviews* 212: 925–984. <https://doi.org/10.1007/s11214-017-0405-1>.
- McCoy, T.J. et al. 2002. The near Earth asteroid rendezvous mission to asteroid 433 Eros: A milestone in the study of asteroids and their relationship to meteorites. *Chemie der Erde* 62(2): 89–121. <https://doi.org/10.1078/0009-2819-00004>.
- McInnes, C.R. 1999) *Solar Sailing: Technology, Dynamics and Mission Applications*. Berlin, Germany: Springer. ISBN: 978-1-85233-102-3. <https://doi.org/10.1007/978-1-4471-3992-8>.
- McKenna-Lawlor, S.M.P. et al. 1993. Energetic ions at comet Grigg-Skjellerup measured from the Giotto spacecraft. *Nature* 363(6427): 326–329.
- McNutt, L. et al. 2014. Near-Earth Asteroid (NEA) Scout. In: *AIAA SPACE 2014 Conference and Exposition*. AIAA 2014–4435. San Diego, CA, 4–7 August. <https://doi.org/10.2514/6.2014-4435>.
- Mengali, G., and A.A. Quarta 2005. Optimal three-dimensional interplanetary rendezvous using Nonideal solar sail. *Journal of Guidance, Control, and Dynamics* 28(1): pp. 173–177. <https://doi.org/10.2514/1.8325>.
- Mengali, G. et al. 2007a. Refined solar sail force model with mission application. *Journal of Guidance, Control, and Dynamics* 30(2): 512–520. ISSN: 0731–5090. <https://doi.org/10.2514/1.24779>.
- Mengali, G., and A.A. Quarta. 2007b. Tradeoff performance of hybrid low-thrust propulsion system. *Journal of Spacecraft and Rockets* 44(6): 1263–1270. ISSN: 0022–4650. <https://doi.org/10.2514/1.30298>.

- Mengali, G., and A.A. Quarta. 2007c. Trajectory design with hybrid low-thrust propulsion system. *Journal of Guidance, Control, and Dynamics* 30(2): 419–426. <https://doi.org/10.2514/1.22433>.
- Mengali, G., and A.A. Quarta. 2014. Optimal nodal flyby with near-Earth asteroids using electric sail. *Acta Astronautica* 104(2), 450–457. <https://doi.org/10.1016/j.actaastro.2014.02.012>.
- Mori, O. et al. 2020. Solar power sail mission of OKEANOS. *Astrodynamics* 4(3). <https://doi.org/10.1007/s42064-019-0067-8>.
- Mori, O. et al. 2010. Attitude control of IKAROS solar sail spacecraft and its flight results. In: *61st International Astronautical Congress*. Paper IAC-10.C1.4.3. Prague, Czech Republic.
- Murphy, M. 2015. Delta-V to Near-Earth Asteroids: an examination of the Shoemaker-Helin equations. MA thesis. Harvard College.
- Neubeuer, F.M. et al. 1986. First results from the Giotto magnetometer experiment at comet Halley. *Nature* 321(6067): 352–355. <https://doi.org/10.1038/321352a0>.
- Niccolai, L. et al. 2019. Effects of optical parameter measurement uncertainties and solar irradiance fluctuations on solar sailing. *Advances in Space Research* 67(9): 2784–2794, May 2021. <https://doi.org/10.1016/j.asr.2019.11.037>.
- Nye, B., and E. Greeson. 2016. The Lightsail story, public outreach strategies & results. In: *67th International Astronautical Congress*. Pasadena, CA, USA.
- Oberst, J., et al. 2001. A model for rotation and shape of asteroid 9969 Braille from ground-based observations and images obtained during the Deep Space 1 (DS1) flyby. *Icarus* 153: 16–23. <https://doi.org/10.1006/icar.2001.6648>.
- Perna, D., M.A. Barucci, and M.Fulchignoni. 2013. The near-Earth objects and their potential threat to our planet. *Astronomy and Astro-physics Review* 21(1). <https://doi.org/10.1007/s00159-013-0065-4>.
- Perozzi, E., A. Rossi, and G.B. Valsecchi. 2001. Basic targeting strategies for rendezvous and flyby missions to the near-Earth asteroids. *Planetary and Space Science* 49(1): 3–22. ISSN: 0032–0633. [https://doi.org/10.1016/S0032-0633\(00\)00124-0](https://doi.org/10.1016/S0032-0633(00)00124-0).
- Rayman, M.D. 2020a. Dawn at ceres: The first exploration of the first dwarf planet discovered. *Acta Astronautica* 194: 334–352, May 2022. <https://doi.org/10.1016/j.actaastro.2019.12.017>.
- Rayman, M.D. 2020b. Lessons from the dawn mission to Ceres and vesta. *Acta Astronautica* 176: 233–237. <https://doi.org/10.1016/j.actaastro.2020.06.023>.
- Russell, C.T. et al. 2013. Dawn completes its mission at 4 Vesta. *Meteoritics and Planetary Science* 48(11), 2076–2089. <https://doi.org/10.1111/maps.12091>.
- Russell Lockett, T. et al. 2019. Near-Earth Asteroid Scout flight mission. *IEEE Aerospace and Electronic Systems Magazine* 35(3): 20–29. <https://doi.org/10.1109/MAES.2019.2958729>.
- Saiki, T., et al. 2020. Hayabusa2's kinetic impact experiment: Operational planning and results. *Acta Astronautica* 175: 362–374. <https://doi.org/10.1016/j.actaastro.2020.05.064>.
- Saito, T. et al. 1986. Interaction between comet Halley and the interplanetary magnetic field observed by Sakigake. *Nature* 321(6067): 303–307. <https://doi.org/10.1038/321303a0>.
- Sauer Jr, C.G. 1977. A comparison of Solar Sail and Ion Drive Trajectories for a Halley's Comet Rendezvous Mission. In: *Astrodynamics Specialist Conference*. Jackson Hole, WY, USA.
- Shoemaker, E.M., and E.F. Helin. 1978. Earth-approaching asteroids as targets for space exploration. Tech. rep. NASA.
- Taylor, M.G.G.T. et al. 2017. The Rosetta mission orbiter science overview: The comet phase. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375(2097). <https://doi.org/10.1098/rsta.2016.0262>.
- Tsuda, Y. et al. 2011. Achievement of IKAROS—Japanese deep space solar sail demonstration mission. In: *7th IAA Symposium on Realistic Advanced Scientific Space*. Aosta, Italy. <https://doi.org/10.1016/j.actaastro.2012.03.032>.
- Tsuda, Y., et al. 2020. Hayabusa2 mission status: Landing, roving and cratering on asteroid Ryugu. *Acta Astronautica* 171: 42–54. <https://doi.org/10.1016/j.actaastro.2020.02.035>.
- Veverka, J., et al. 1999. NEAR encounter with asteroid 253 mathilde: Overview. *Icarus* 140: 3–16. <https://doi.org/10.1006/icar.1999.6120>.

- Veverka, J., et al. 2001. Imaging of small-scale features on 433 eros from NEAR: Evidence for a complex regolith. *Science* 292: 484–488. <https://doi.org/10.1126/science.1058651>.
- Von Rosenvinge, T.T., J.C. Brandt, and R.W. Farquhar. 1986. The international cometary explorer mission to comet Giacobini-Zinner. *Science* 232(4748): 353–356. <https://doi.org/10.1126/science.232.4748.353>.
- Vulpetti, G. 2013. *Fast Solar Sailing: Astrodynamics of Special Sailcraft Trajectories*. Dordrecht, Netherlands: Springer. Chap. 6, pp. 165–254. ISBN: 978-94-007-4776-0. <https://doi.org/10.1007/978-94-007-4777-7>.
- Vulpetti, G., L. Johnson, and G.L. Matloff. 2015. *Solar Sails: A Novel Approach to Interplanetary Travel*. Chichester, UK: Springer-Praxis. ISBN: 978-1-4939-0940-7. <https://doi.org/10.1007/978-1-4939-0941-4>.
- Walker, M.J. 1986. Erratum—a set of modified equinoctial orbit elements. *Celestial Mechanics* 38(4): 391–392. <https://doi.org/10.1007/BF01238929>.
- Walker, M.J.H., J. Owens, and B. Ireland. 1985. A set of modified equinoctial orbit elements. *Celestial Mechanics* 36: 409–419. <https://doi.org/10.1007/BF01227493>.
- Wie, B. 2005. Thrust vector control of solar sail spacecraft. In: *AIAA Guidance, Navigation, and Control Conference and Exhibit*. Paper AIAA 2005–6086. San Francisco, California.
- Wright, J.L. 1992. *Space Sailing*. Philadelphia, PA, USA: Gordon and Breach Science Publishers.
- Wright, J.L., and J.M. Warmke. 1976. Solar sail mission applications. In: *AIAA/AAS Astrodynamics Conference*. San Diego, CA, USA.
- Yamaguchi, K., and H. Yamakawa. 2016. Electric solar wind sail kinetic energy impactor for asteroid deflection missions. *Journal of the Astronautical Sciences* 63(1). <https://doi.org/10.1007/s40295-015-0081-x>.
- Zola, D. et al. 2018. Photon momentum change of quasi-smooth solar sails. *Journal of the Optical Society of America A: Optics and Image Science, and Vision* 35(8): 1261–1271. <https://doi.org/10.1364/JOSAA.35.001261>.

Part VI
Ocean Worlds

Chapter 26

Ocean Worlds: Interior Processes and Physical Environments



Samuel M. Howell and Erin J. Leonard

Abstract One of the most profound insights in planetary science has been the recent discovery that bodies with current or past subsurface global water oceans are relatively common in our solar system. These ocean worlds have risen to the forefront of planetary exploration and the search for life. This chapter describes ocean-world surfaces and interior structures, as well as their constraints, and explores how they may drive future spacecraft and instrument design. While this chapter is not exhaustive, it is representative of how spacecraft data, numerical models, and geologic inferences can be employed to constrain physical, quantitative descriptions of ocean-world interior environments. We begin with an overview of the exploration history of these bodies and the scientific motivation for their study. We then explore the interior structures of ocean worlds, and how geological observation and geophysical inferences are used to quantitatively constrain icy and liquid water interior environments. We specifically focus on the physical and thermal aspects of interior environments, overviewing how observational data is used as inputs to numerical and mathematical models that predict interior state.

26.1 Introduction to Ocean Worlds

26.1.1 Exploration History

In 1977, the National Aeronautics and Space Administration (NASA) launched the dual *Voyager* spacecraft, beginning a tour of the outer solar system and beyond that continues today. In 1979, these spacecraft encountered Jupiter and its satellites, teasing images of both fire and ice in the volcanic and frozen worlds they encountered.

S. M. Howell (✉) · E. J. Leonard
NASA Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr. M/S
183-301, Pasadena, CA 91011, USA
e-mail: samuel.m.howell@jpl.nasa.gov

E. J. Leonard
e-mail: erin.j.leonard@jpl.nasa.gov

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_26

In 1989, NASA launched the *Galileo* spacecraft, bound to orbit Jupiter for the first time, which would explore the gas giant system and its satellites from 1995 to 2003. As the *Galileo* spacecraft orbited Jupiter, it made several flybys of the Galilean satellites, named for Galileo Galilei, who first documented their observation in 1610 one day before co-discoverer Simon Marius.

With increasing distance from Jupiter, *Galileo* surveyed Io, a rocky and volcanic moon driven by intense tidal interactions with Jupiter; Europa, a frozen world about the size of Earth's moon; Ganymede, the largest satellite and icy world in the solar system with a radius slightly greater than that of the planet Mercury; and the impact-battered icy world Callisto. As *Galileo* orbited Jupiter, it measured the time and spatially varying magnetic field produced by the gas giant, as well as interactions between the magnetic field and the satellites. This led to the fascinating discovery of an intrinsic magnetic field at Ganymede, the only satellite in the solar system that exhibits this behavior (Kivelson et al. 1996). However, these were not the most profound measurements of the Galilean satellites' magnetic fields.

The *Galileo* spacecraft measurements indicated that Europa and Callisto did not exhibit their own magnetic field, but rather detected induced magnetic field resulting from their passage through Jupiter's dynamic magnetic environment (Khurana et al. 1998; Kivelson et al. 2000). The induced magnetic fields of Europa, Callisto, and Ganymede could only exist if these worlds had global electrically conductive layers beneath their icy surfaces.

Further analyses of Europa's gravity and spectral data (Carr et al. 1998), and surface geology (Pappalardo et al. 1999) pointed towards an explanation—a global saltwater ocean ~100 km deep buried beneath a predominantly water–ice shell. Intense gravitational tidal interactions between Europa and Jupiter dissipate gravitational energy as frictional heat within the satellite, supporting a hypothesis that predates the *Galileo* launch, whereby a global water ocean on Europa can be sustained through time by tidal heating (Ross and Schubert 1987).

Much like Europa, there is now a scientific consensus that global saltwater oceans likely exist on two other Galilean satellites, Ganymede and Callisto, and on Saturn's small moon Enceladus and giant moon Titan (Fig. 26.1) (Hendrix et al. 2019). Myriad worlds are potential candidates for hosting past or present oceans: Neptune's large moon Triton, the Uranian satellites, additional Neptunian satellites, the largest main belt asteroid and dwarf planet, Ceres (Castillo-Rogez et al. 2020), and the nearest dwarf planet of the Kuiper Belt, Pluto (Nimmo et al. 2016).

As we explore these worlds with currently planned and potential future missions, it is necessary to build an understanding of their interior processes to place bounds and highlight uncertainties on the physical, thermal, mechanical, and compositional environments that will be encountered. Geophysical investigations contribute to the full gamut of ocean worlds science, including their potential to host past or extant life (Hand et al. 2009, 2007; Vance et al. 2016). Here, we highlight how geological and geophysical analysis of previous spacecraft and telescopic observations constrain the interior structures and environments of these worlds. Inferences of these worlds' interiors also have a crucial role in providing constraints for spacecraft and instrument design (Dachwald et al. 2021; Fleurial et al. 2019; Hockman et al. 2021; Howell et al.

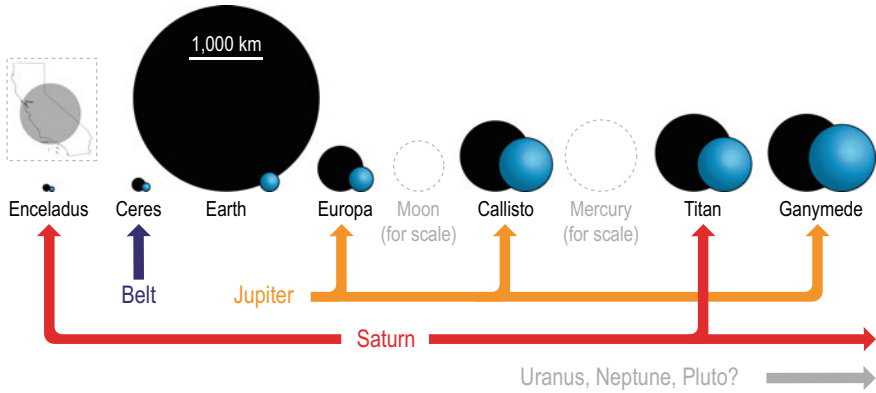


Fig. 26.1 Potential ocean worlds of the solar system ordered by estimated water volume and represented to scale (black circles), with blue circles representing a sphere with equivalent volume to that of the liquid and frozen water contained on each body. Note that this excludes the water that may be bound in hydrated minerals. California is shown as a scale comparison with Enceladus in the callout above that body. Dashed circles represent the size of the Moon and Mercury for scale. Water volumes for Enceladus, Europa, Callisto, Titan, and Ganymede are from Vance et al. (2018). The water–ice volume for Ceres is from Castillo-Rogez et al. (2020). While there are likely ocean worlds beyond Saturn, including potentially Pluto, any potential water volumes are poorly constrained and unconfirmed

2020; Klonicki et al. 2019, 2021; Nayar 2021; Palmowski et al. 2019), as well as for planetary protection considerations (McCoy et al. 2020).

Many of the methods described within this chapter are illustrated with examples for Europa, the target of NASA’s planned Europa Clipper flagship mission, launching no later than 2024 (Howell and Pappalardo 2020), one of the targets of ESA’s planned Jupiter Icy Moons Explorer (JUICE) L-class mission (Grasset et al. 2013), and one target of the extended mission for NASA’s New Frontiers-class Juno mission. While Europa is often used here as an illustrative example because of these upcoming missions, these techniques are widely applied across the myriad ocean worlds and their processes.

26.1.2 Scientific Motivation

The search for life in the universe is a priority science theme within NASA’s science strategy (NASA 2021), and the search for extant life among the ocean worlds is emphasized within the planetary science and astrobiology communities (Hand and German 2018; Moore et al. 2020; Sherwood et al. 2018; Space Studies Board 2018; Space Studies Board, National Research Council 2012). Each decade, NASA conducts a survey of outstanding questions in planetary science and astrobiology and

Table 26.1 Priority themes and questions posed by the NASA Astrobiology Program’s Network for Ocean Worlds (NOW) to the 2023–2032 Decadal Survey in Planetary Science and Astrobiology that relate fundamentally to investigations of ocean-world interior processes and environments (Howell et al. 2020)

Theme	Question
Life and Habitability	Has life emerged within ocean worlds, and does it persist today?
	What are the habitable environments and interfaces of ocean worlds?
Ocean World Evolution	What are the key interfaces that permit and regulate thermal, physical, and chemical exchange?
	How did differentiation occur for the key geologic layers: the ice shell, ocean, and rocky interior?
	What tidal interactions and other sources of energy power change within these layers?
Comparative Oceanography	What analog geological processes on Earth might provide insight into the sources of chemical energy that may render an ocean world habitable?
	How and on what timescales are biosignatures transported, modified, and preserved?

recommendations on how to address these questions through a report empaneled by the National Academies of Sciences, Engineering, and Medicine.

The 2023–2032 Decadal Survey in Planetary Science and Astrobiology solicited community input from 2020 to 2021, and outstanding questions related to exploration of ocean worlds were submitted by the NASA Astrobiology Program’s Network for Ocean Worlds (NOW) with broad community support (Howell et al. 2020). Of the priority themes and questions put forward by NOW, the majority are intricately tied to studies of the interior processes and thermomechanical environments of Ocean Worlds (Table 26.1).

26.2 Structures of Ocean Worlds

In this section, we overview the interior structure of the ocean worlds most likely to possess extant global saltwater oceans (Table 26.2), all of which are natural satellites of Jupiter and Saturn. While each of these worlds is unique, these icy satellites generally share three or more of the following interior layers (Fig. 26.2): an outer shell of ice Ih (the hexagonal-form ice found on Earth that is less dense than liquid water), an interior salty water ocean, high-pressure water ice phases that are denser than water, and a rocky interior layer that may have differentiated to form a metallic core, or whose metals remain interspersed with the silicates.

Our understanding of the interior structure of these worlds is shaped by data collected by remote sensing by robotic spacecraft, and by ground- and space-based

Table 26.2 List of example thickness ranges and interface properties of the various layers of the known ocean worlds. Body parameters come from the NASA Jet Propulsion Laboratory's Planetary Satellite Physical Parameters table (2021, and references therein)

Parameter	Europa	Ganymede	Callisto	Enceladus	Titan
Mass [kg]	4.8×10^{22}	1.5×10^{23}	1.1×10^{23}	1.1×10^{20}	1.3×10^{23}
Radius [km]	1,560.8	2,631.2	2,410.3	2,52.1	2,574.7
Gravity [m s ⁻²]	1.315	1.428	1.236	0.113	1.352
Mean density [kg m ⁻³]	3,013	1,942	1,834	1,375	1,608
Ice Ih shell thickness [km]	23–47 [2]	25–150 [1]	100–130 [1]	6–36 [3]	50–150 [1]
Ocean depth [km]	105–150 [2]	25–500 [1]	20–140 [1]	20–50 [3]	90–420 [1]
High-pressure ice thickness [km]	0 [1]	330–650 [1]	5–240 [1]	0 [3]	0–270 [1]
Rocky interior thickness [km]	600–1,000 [2]	1,700–1,780 [1]	1,970–2170 [1]	170–205 [3]	2,030–2,100 [1]
Metallic core thickness [km]	450–810 [2]	440–660 [1]	0 [1]	0 [3]	0 [1]
Surface temperature [K]	46–90 [4]	50–120 [5]	50–125 [5]	60–180 [6]	89–94 [7]
Ice-ocean interface temperature [K]	268–273 [1]	250–273 [1]	250–257 [1]	269–273 [1]	250–268 [1]
Surface pressure [bar]	10^{-11} [8]	10^{-8} – 10^{-3} [10]	10^{-8} – 10^{-7} [9]	10^{-13} – 10^{-10} [11]	1.47 [12]
Ice-ocean interface pressure [bar]	280–570	330–2,000	1,100–1,500	6–38	620–1,900
Sea-floor pressure [bar]	1,600–2,500	5,400–12,000	1450–4,700	29–61	5,500–7,600

Values of thicknesses and temperature are sourced from ¹Vance et al. (2018, and references therein), ²Howell (2021, and references therein), ³Hemingway et al. (2018, and references therein). Surface temperatures and pressures are sourced from ⁴Ashkenazy (2016), ⁵Squyers (1980), ⁶Spencer and Grinspoon (2007), ⁷Jennings et al. (2019), ⁸Hall et al. (1995), ⁹Liang et al. (2005), ¹⁰Eviatar et al. (2001, and references therein), ¹¹Waite et al. (2006), and ¹²Harri et al. (2006). Interface pressures are calculated using the layer densities of Howell (2021)

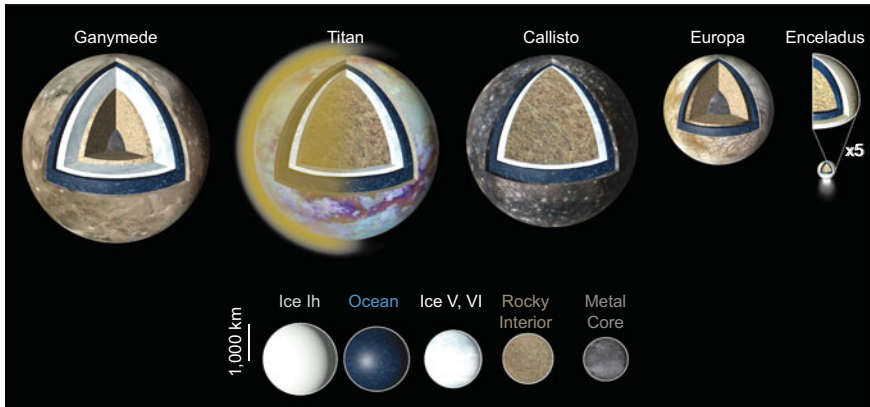


Fig. 26.2 The ocean worlds that are most likely to host present-day global interior oceans, arranged by radius. The shell cutaways depict the various layers of these bodies, including the outer ice Ih shells, interior saltwater oceans, and high-pressure inner ice shells for Ganymede, Titan, and Callisto. The interiors of Ganymede and Europa have differentiated into silicate and metal components, while the rocky-metal interiors of Titan, Callisto, and Enceladus remain undifferentiated. Layer thicknesses have high uncertainty, with example values from Vance et al. (2018) for Ganymede, Titan, and Callisto, Howell (2021) for Europa, and Hemingway et al. (2018) for Enceladus. See Table 26.1 for details

telescopic observation. An overview of planetary instrumentation, including potential applications to ocean worlds, is provided in Klonicki et al. (2021). In addition to remote and in-situ observations, laboratory studies and theoretical development of our understanding of ices and their interaction with the planetary and space environment are key to understanding ocean worlds' evolution and structure.

In this section, we provide an overview of our current understanding of these worlds' structures, including their surface geology, icy layers, the ocean, and important interfaces. Detailed landform descriptions are given in Sect. 26.3.1. We focus on the physical and thermal properties of these layers. As with laboratory studies, theoretical and numerical studies of the mechanical behavior of planetary ice shells and oceans can directly provide quantitative descriptions of ice-shell environments that can be used to derive measurement, spacecraft, and instrument requirements (Sect. 26.4).

26.2.1 Surface Geology

Visible surfaces offer the first hint that a body could be an ocean world. A young surface, or a surface that lacks impact craters, can be an indicator of resurfacing or heating, and possibly a subsurface ocean. For example, while Earth's moon has >100,000 impact craters spanning 4.5 Gyr of history (e.g., Yang et al. 2020), only 47 have been identified on Europa (Leonard et al. 2021b), indicating a global average

surface of ~60 Myr (Bierhaus et al. 2009; Zahnle et al. 2003). Surfaces that do not have preserved pronounced features can be indicative of high internal heat flow that relaxes away topography (e.g., Bland et al. 2012). Europa is the prime example of this as it is considered one of the smoothest bodies in our solar system due to the low topographic relief, $\sim\pm 1$ km (for comparison, the topographic relief on the Moon is $\sim\pm 9$ km). However, the topography and age of the surface are not always a diagnostic of a subsurface ocean world, as evidenced by the numerous craters that saturate Callisto's surface (Greeley et al. 2000).

The surface of each known ocean world has unique geologic features (Fig. 26.3). Europa is characterized by regions of chaos (shades of green), a terrain that consists of broken-up pieces of pre-existing terrain sitting in a matrix, and ridged plains (blue), terrain consisting of numerous cross-cutting ridges (e.g., Leonard et al., in press, and references therein); Ganymede is covered in bright and dark grooved terrains (Patterson et al. 2010); Callisto's cratered surface is affected by sublimation processes resulting in sharp spires known as penitentes (Greeley et al. 2000); Enceladus consists of regions of old, cratered terrain immediately adjacent to younger, tectonized terrains (Crow-Willard and Pappalardo 2015); Titan has methane liquid-filled lakes near its north pole and dunes that dominate its equatorial regions (Lopes et al. 2020). While there are some common geologic features among some of the known ocean worlds (e.g., bands on Europa and Ganymede), each body has its own, unique geologic landforms that can provide insights into the interior structure of the ice shell (see Sect. 26.3).

26.2.2 Ice Shells

Ocean world ice shells are generally regarded as comprising two fundamental layers (Howell 2021; Howell and Pappalardo 2019; McKinnon 1999; Pappalardo et al. 1998). As described below, these include a cold upper layer where geologic heat transfer occurs primarily through thermal conduction, and a warm ductile layer that may experience solid-state convection. The brittle portion of the conductive layer is typically referred to as the lithosphere, from the Greek *lithos*, meaning rock. The ductile portion of the conductive layer and the potential convective layer comprise the asthenosphere, from the Greek *asthenes*, meaning weak. See Fig. 26.4 for an illustration of temperature profiles within these layers, and Fig. 26.5 for an illustration of processes within these layers on Europa.

The relative and total thicknesses of these layers on each body are poorly constrained, and these thicknesses are critical to understanding the geological and geophysical environment (Billings and Kattenhorn 2005; Howell 2021). Estimates of the thickness of these layers depend on assumptions about the global heat budget, composition, and planetary history (Howell 2021; Vance et al. 2018). Ice-shell thickness estimates are summarized in Table 26.2. Section 26.3 describes methods for deriving quantitative descriptions of ocean-world interiors, and Sect. 26.4 describes

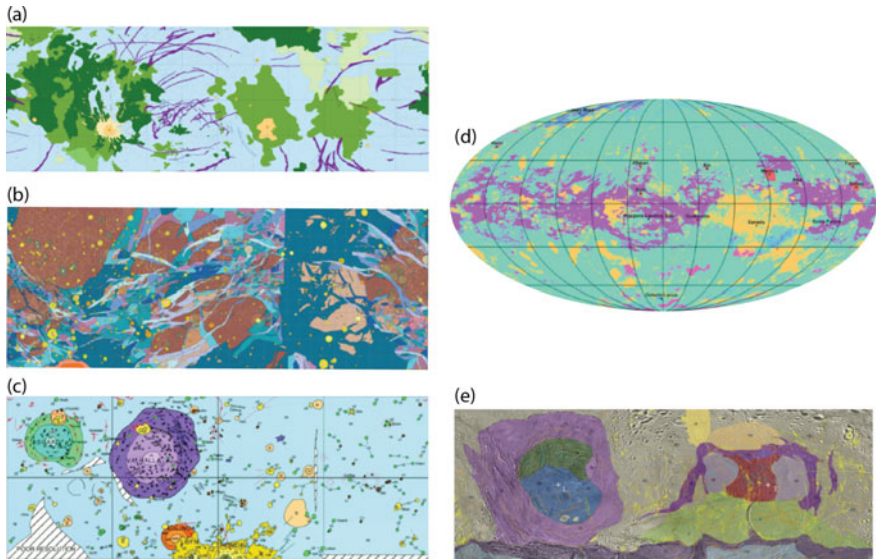


Fig. 26.3 Global geologic maps highlighting the differences in the surfaces of **a** Europa (modified from Leonard et al. 2021b); **b** Ganymede (modified from Patterson et al. 2010); **c** Callisto (modified from Greeley et al. 2000); **d** Titan (modified from Lopes et al. 2020); and **e** Enceladus (modified from Crow-Willard and Pappalardo 2015). The maps are not to scale with respect to one another. Here, colors represent various geologic units relating to specific landforms and periods in the body’s history, and their keys can be found within the citations above. Note that the distributions of major landforms vary significantly, from the approximately equatorial distribution of dunes on Titan (purple) to Ganymede’s bimodal distribution of dark (brown) and light (blue) terrains. These landforms and their distributions offer clues, unique to each body, into the internal structure and evolution of these worlds

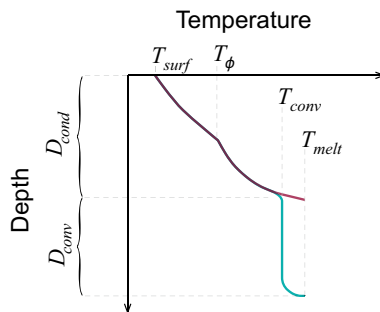


Fig. 26.4 An illustration of ice-shell thermal profiles for (red) a thinner ice shell where geologic heat transfer occurs only through conduction and (blue) a thicker ice shell that exhibits isothermal convection at depth. Here, D_{cond} and D_{conv} are respectively the conductive and convective layer thicknesses, T_{surf} is the surface temperature, T_{ϕ} is the temperature to which the porous lithosphere extends, T_{conv} is the isothermal temperature of convection, and T_{melt} is the melting temperature. Example thermal and mechanical interface values are provided for Europa in Sect. 26.4, and values of D_{cond} , D_{conv} , T_{surf} , and T_{melt} are provided for each body in Table 26.2



Fig. 26.5 Artist's depiction of geological processes within the ice shell of Europa (modified from Howell and Pappalardo 2020). Tectonic processes such as rifting, extensional band formation, and subsuspension may create major faults that span lateral distances of more than 1,000 km and extend to depths of 1–10 km. Below the brittle region of the ice shell, regions of coalesced melted water ice may form connected pore networks or interstitial water bodies. At depth, the solid ice may convect due to the thermal expansion of ice near the warm ice-ocean interface providing thermal buoyancy, and the sinking of cold, dense ice near the brittle layer. Some of these processes may result in the interaction of liquid water with the shallow subsurface, and/or the eruption of water vapor plumes

applications of different techniques for placing requirements on spacecraft and instrument design, despite these uncertainties.

26.2.2.1 Conductive Layer

Near the surface, the ice is cold and the dominant mode of geologic heat transfer across this layer is thermal conduction. The upper portion of this conductive layer deforms through brittle and shear failure in response to geologic stresses over timescales of thousands to millions of years (e.g., Howell and Pappalardo 2018), as well as through elastic bending and flexure (e.g., Nimmo et al. 2003a). At the base of the conductive layer, the ice is warmer and its viscosity is low enough that fractures and faults are impeded and healed over geologic time, and the ductile region of the conductive layer will allow the viscous relaxation of surface stresses and topography.

In addition to hosting brittle faults, the upper portions of the conductive layer of planetary ice shells may also exhibit significant porosity. At the cold surface temperatures of the outer solar system, porosity introduced through tectonic fracturing (Nimmo et al. 2003b), impacts (Wünnemann et al. 2006), snow deposition from water vapor plumes (Spencer et al. 2009), and the formation of ice shell may be retained through geologic time. At depth within the conductive layer, the ice is able to relax and close porosity at higher temperatures and for lower viscosity ice, so that porosity persists only to some critical temperature within the ice shell (Besserer et al. 2013).

Unique within the solar system, the upper portion of the conductive layer of Titan experiences rapid weathering due to the presence of an atmosphere and three-phase methane cycle. Here, organic weathering can produce many features familiar to Earth, including lakes, dunes, and caverns (Jaumann et al. 2010). Methane and ethane trapped within the porous structure of Titan's conductive icy layer may additionally affect the thermal state of the ice shell, providing both a thermally insulating layer that greatly increases the ice-shell thickness, and a density contrast that might drive geologic activity (Kalousová and Sotin 2020).

Equally as interesting, many researchers have inferred the presence of shallow liquid water bodies on Europa (Hammond 2020; Manga and Michaut 2017; Schmidt et al. 2011). However, no consensus yet exists on the plausibility of such reservoirs or their longevity, due to arguments that such a configuration is energetically unfavorable, as evidenced by the rapid freezing timescales of subsurface reservoirs (Buffo et al. 2020), and the tendency of any melt in the shallow subsurface to quickly advect to the ocean below (Kalousová et al. 2016). Further investigations of Europa by *Juno*'s Microwave Radiometer (MWR) and by the REASON and RIME radar instruments of *Europa Clipper* and *JUICE*, respectively, may shed light on the presence and fate of shallow liquid reservoirs.

26.2.2.2 Convective Layer

Beneath the conductive layer may be a warm, nearly isothermal layer experiencing solid-state convection (Barr and McKinnon 2007a; Besserer et al. 2013; Grott et al. 2007; Pappalardo et al. 1998). Here, ice near the ice–ocean interface is warm, and at a lower density than ice near the base of the conductive layer, driving buoyant material from near the interface upwards and dense material from beneath the cool conductive layer downwards. Additionally, the ice may be heated from within by the frictional dissipation of tidal energy from orbital eccentricity (Ross and Schubert 1987; Sotin et al. 2002; Vilella et al. 2020) and obliquity (Jankowski et al. 1989; Nimmo and Spencer 2015).

The convective layer is likely free of vacuum-filled pores because the warm interior ice quickly relaxes over geologic timescales (e.g., Howell 2021). However, small amounts of melt may be produced and trapped within the convecting layer, providing up to a few percent water by volume trapped within the ice at depth (Vilella et al. 2020).

Freezing processes at the base of the convective layer in a thick ice shell are likely too slow to allow for the incorporation of significant non-ice materials (e.g., salts) (Buffo et al. 2020), and therefore the convecting interior water ice may be relatively pure. However, it is possible for brines to form and persist at the ice–ocean interface, and where seawater has been injected directly into the ice shell (Buffo et al. 2021b; Vance et al. 2019).

26.2.3 *Ice–Ocean Interfaces*

The ice–ocean interfaces of ocean worlds are poorly understood, in large part because of the difficulty in obtaining unique constraints on their structure and behavior. On Earth, this interface is crucial for regulating the global exchange of heat and salts (McPhee et al. 2008). On the ocean worlds, the ice–ocean interface may be one of the most crucial for understanding not only heat and material transport, but planetary habitability. Active geologic processes within planetary ice shells are critical for exchanging thermal and chemical energy between the surface and the ocean (e.g., Howell and Pappalardo 2019, 2018), potentially establishing or maintaining the chemical disequilibrium required for life to emerge or persist (Hand et al. 2009, 2007; Vance et al. 2016).

Studies of the ice–ocean interface on Europa provide a wide range of predictions. Energy and mass balances adopted from Earth provide some constraints on interface and compositional evolution of planetary ice shells (Buffo et al. 2020, 2021b). These interfaces may host regions of crystalline mush up to several meters in thickness that may drive requirements for seismic or radar detection (Buffo et al. 2021b). If the interfaces are indeed sharp, long-wavelength topography may arise from thermal convection within the ice shell, and short wavelength topography at the interface could include salty “brinicles” that protrude up to meters into the ocean (Vance et al. 2019). These brinicles form because salty H₂O mixtures have lower freezing temperatures, allowing brine to drain downwards before freezing, and may provide local habitats at the ice–ocean interface (Vance et al. 2019).

Because of their potential as habitable environments, planetary ice–ocean interfaces are seen as one of most crucial to explore in situ in the search for life, and their access is driving considerations for the future of ocean worlds exploration (Dachwald et al. 2021; Hockman et al. 2021; Howell et al. 2020; Klonicki et al. 2021; Moore et al. 2020; Nayar 2021; Sherwood et al. 2018). In the past decade, NASA has invested heavily in ocean access technologies through their funding programs, including Scientific Exploration Subsurface Access Mechanism for Europa (SESAME), Concepts for Ocean Worlds Life Detection Technology (COLDTech), and Planetary Science and Technology from Analog Research (PSTAR).

26.2.4 Interior Oceans

The oceans of the solar system are central to the search for life in the universe (Space Studies Board 2018; Space Studies Board, National Research Council 2012), spanning likely depths of a few tens to several hundreds of kilometers (Table 26.2). They provide plentiful liquid water, the potential for sustained chemical disequilibrium at their icy and rocky interfaces, and a warm habitat in a cold region of the solar system, and facilitate the transport of chemical constituents and thermal energy.

Understanding the physical state of the ocean is a key objective of ocean-world exploration. Planetary oceans might be stratified, where they are poorly mixed and temperature and composition are strong functions of depth, or they may be vigorously convecting, quickly transporting material from the sea-floor and ice–ocean interfaces and producing a homogeneous layer (Hay and Matsuyama 2019; Lobo et al. 2021; Thomson and Delaney 2001). Broadly, salinities are generally considered to span approximately one order of magnitude above and below a few weight percent, which is the approximate salinity of Earth’s oceans (Hand and Chyba 2007).

The presence of oceans on the icy satellites and minor planets are inferred through measurements of magnetic induction, gravity, and orbital properties and geometry. No planetary ocean has been directly observed, and therefore the composition, state, and behavior of planetary oceans are poorly understood. At Saturn’s moon Enceladus, subsurface water originating from the ice shell or ocean is ejected to space from south polar jets, and the resulting plume was investigated by the *Cassini* spacecraft.

Because of Enceladus’ relatively low gravity (0.113 m s^{-2}), diurnal tidal interactions with Saturn may be sufficient to provide a link between the subsurface ocean and space environment via deep tectonic fractures (Spitale and Porco 2007). The ocean water may rapidly ascend towards the surface and erupt from the observed jets, or evolve through time within ice-shell reservoirs, providing insight into the reservoir composition and history (Postberg et al. 2018a). Results from in-situ instruments aboard the *Cassini* spacecraft during its fly-through of the south polar jets indicate the presence of organics (Postberg et al. 2018b) in an alkaline (pH 9–11) ocean (Glein et al. 2015, 2020).

Plumes have additionally been observed on Triton (Kirk et al. 1995), the target of the proposed (but unselected) Discovery-class Trident flyby mission. They have also tentatively been observed on Europa (e.g., Jia et al. 2018; Sparks et al. 2016) (Fig. 26.6), the prime target of NASA’s Europa Clipper flagship-class mission, and one target of ESA’s JUICE L-class mission and NASA’s Juno New Frontiers-class mission. Future exploration and observation of these plumes may offer similar insight into the composition and state of other planetary oceans on geologically active worlds.

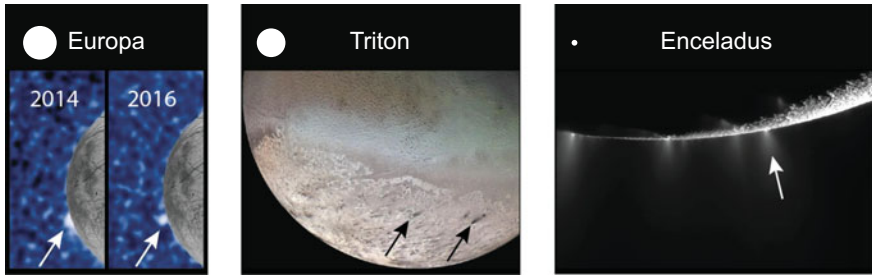


Fig. 26.6 Example observations of (left) a potential plume detection on Europa using the Hubble space telescope, indicated with a white arrow (modified from Sparks et al. 2016), (middle) geyser vents on Triton, indicated by black arrows associated with dark streaks of plume fallout (NASA/JPL-Caltech), and (right) an arrow indicating the largest of several visible water vapor plumes originating from the south polar vents of Enceladus. White circles depict the relative sizes of the three bodies

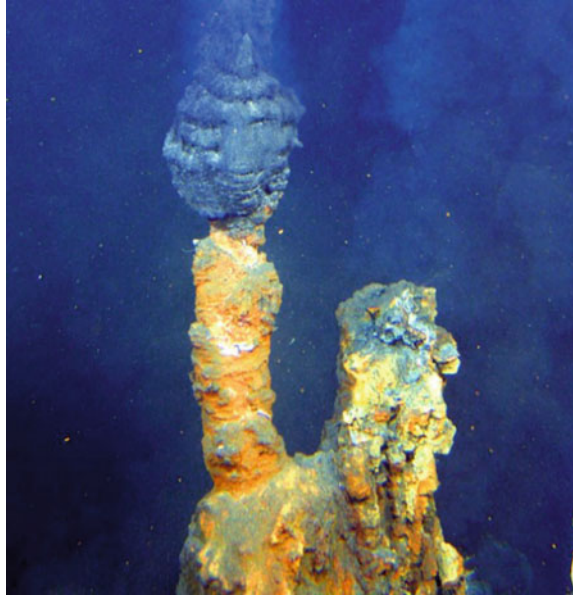
26.2.5 Sea-Floors

Two types of sea-floor are likely found beneath planetary saltwater oceans. For moons with thinner cryospheres or smaller masses, like Europa and Enceladus, oceans are likely in direct contact with the rocky interiors, as on Earth (Table 26.2, Fig. 26.2). For more massive moons and moons with thicker cryospheres, like Ganymede and Triton, the pressure and temperature conditions at the base of the ocean are favorable to formation of high-pressure ice phases that are denser than the seawater, buffering interaction of their interior oceans with their rocky sea-floors.

In cases where the ocean directly interacts with the rocky interior, water rock reactions, such as serpentinization, may act to reduce the ocean and free volatile and soluble species from the rock into the water column (e.g., Hand et al. 2009). Serpentinization is one such water–rock reaction that may provide hydrogen to reduce oxidants coming from the ice shell, and is thus highlighted as a critical reaction in planetary habitability (e.g., Vance and Melwani Daswani 2020). All known life relies on redox potentials as the primary currency for energy, enabling metabolism and reproduction. Serpentinization and periodic volcanism ice on Europa, for example, may have allowed reduction oxidation (or redox) potentials to persist through time (Vance and Melwani Daswani 2020). Enceladus, however, has a small and porous core, and serpentinization may therefore have been rapid and short-lived, or ongoing serpentinization may be indicative of a young ocean (Zandanel et al. 2021). Sulfate-reducing bacteria have been found on Earth’s sea-floor that fully subsist on serpentinization reactions (Glombitza et al. 2021).

Additionally, sea-floor volcanic processes on bodies with active rocky interiors, as might be the case for Europa (Běhounková et al. 2021), may produce structures similar to hydrothermal sea-floor vents found on Earth (Fig. 26.7). Hydrothermal processes act to concentrate minerals into localized hot thermal plumes that inject into the ocean. These vents thus act to concentrate mineral and thermal energy, and

Fig. 26.7 An approximately 2 m-tall alkaline hydrothermal vent on Earth's sea-floor showing a chimney-like sea-floor structure where mineral-rich hydrothermal fluids are ejected from subsurface channels, concentrating thermal and chemical energy [NOAA]. Similar structures may create habitable environments on the sea-floors of ocean worlds



are suggested as one candidate environment for the emergence of life on Earth and ocean worlds (Martin et al. 2008).

In cases where the rocky surface is separated from the ocean by a high-pressure ice layer, the sea-floor environment is governed by a deep, secondary ice-ocean interface. While high-pressure ice layers may produce a barrier that inhibits water-rock reaction, compositional evolution and mixing may be facilitated by the melting of high-pressure ice phases. At these pressures and temperatures, water is buoyant within the high-pressure ice phases, allowing melt produced at the warm rocky surface to quickly percolate through the high-pressure ices and reach the ocean (Kalousová et al. 2018).

26.3 Constraining Interior Environments

Due to the paucity of data measuring the properties of ocean-world interiors, it is necessary to use different predictive and deductive tools to infer the structure, state, and environmental properties of the icy shells and oceans. Chapter 6 of this volume, 'Instrumentation for Planetary Exploration' by Klonicki et al. provides a description of methods for acquiring planetary data. In this section, we overview some of the methods that combine spacecraft and laboratory data with a physical understanding of planetary layers and their processes to define interior environments.

26.3.1 Geologic Inference and Analytical Approaches

The study of geological landforms on planetary surfaces provides information about the history, age, and thermomechanical environments at depth. Example surface terrains from high-resolution image mosaics are shown for each known ocean world in Fig. 26.8. Geologic processes on ocean worlds provide windows into the icy interior because their structure, length, and characteristics are functions of the stresses, strains, and timescales of formation, as well as the thermal, brittle, elastic, and viscous properties of the geologic layers they form within. Studies of landforms thus tend to focus on the size, shape, and textures of the features that comprise their morphology, and on the distributions of features in space and time.

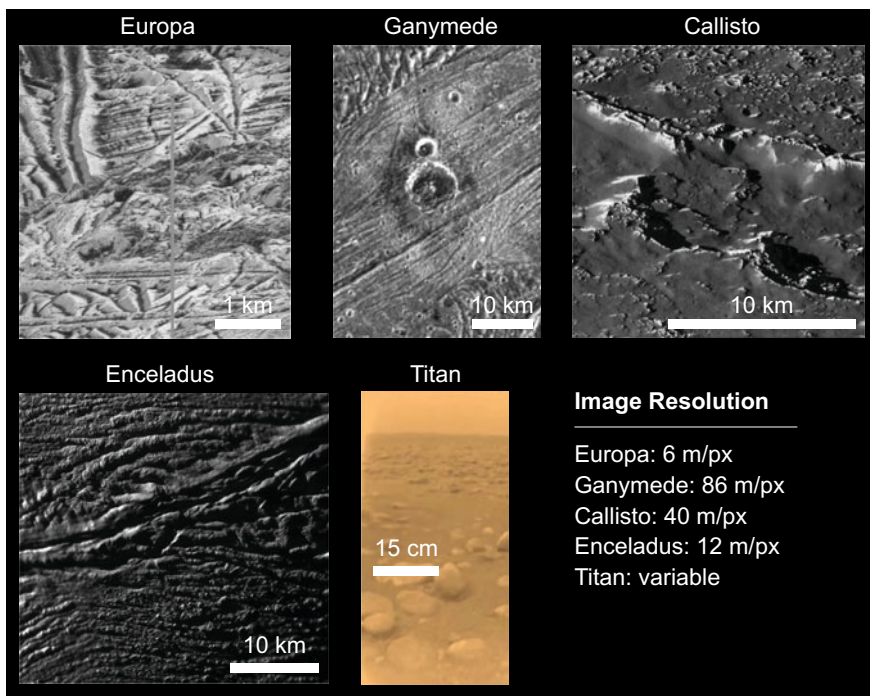


Fig. 26.8 Among the highest resolutions obtained, these spacecraft images and mosaics show landforms characteristic of the known ocean worlds with extant oceans. *Galileo* spacecraft imagery (top) of the Galilean ocean worlds (left) Europa, exhibiting a prominent double-ridge structure in the upper left that has been cross-cut by tectonic activity; (middle) Ganymede, exhibiting an impact crater at the center of a smooth extensional band where the icy lithosphere separated to expose warm interior ice; and (right) Callisto, exhibiting an ancient fault scarp that has been heavily modified by cratering processes (NASA/JPL-Caltech). The Saturnian ocean worlds (bottom) include (left) Enceladus, imaged by *Cassini*, highlighting one of the prominent Tiger Stripe fractures from which the south polar water vapor jets emanate (NASA/JPL-Caltech); and (right) Titan, exhibiting a rock-like icy and organic surface at the *Huygens* landing site (ESA/NASA/JPL-Caltech) (see also Fig. 26.10)

26.3.1.1 Impact Features

As described in Sect. 26.2, our understanding of the ages of planetary surfaces largely derives from the distribution of impact craters on their surfaces. Geologically active worlds like Europa, Enceladus, and Titan are resurfaced, or undergo geologic processes that erase older craters and landforms. Worlds with relatively young surfaces, like Europa, have very few impact craters, while worlds with relatively old surfaces, like Callisto, possess surfaces dominated by impact craters. In addition to providing a metric for surface age estimation, impact processes also sample the subsurface of planetary bodies. Because impacts occur very quickly on geologic timescales and provide large instantaneous stresses, they can cause brittle failure to great depths, may produce melting and subsequent cryovolcanic activity, and can introduce fracture porosity into their target layer.

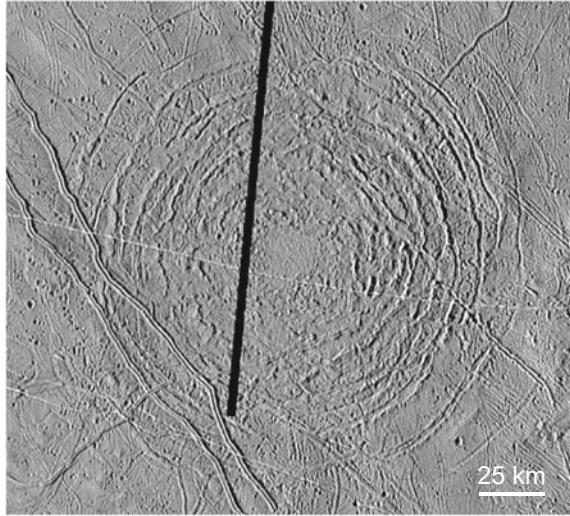
Cratering records are highly sensitive to estimating the relevant impactor populations, the most prominent of which at present are the Jupiter Family Comets (e.g., Zahnle et al. 2003). Without in-situ instruments that can accurately provide absolute ages through isotopic age dating (as have been used in studies of Earth's moon and Mars), the absolute ages of planetary surfaces derived from crater age dating can have order-of-magnitude uncertainties.

Geologic inferences of ice-shell structure can also be made from the cratering record. Of the most incredible examples, geological analyses of Europa's largest impact craters suggest that the impactors penetrated the full lithosphere, reaching a low-viscosity target such as warm ice or even the subsurface ocean (Fig. 26.9) (Moore et al. 1998). Beyond Europa, Leonard et al. (2021c) use the morphology of the relic cratered blocks on Enceladus to understand the potential brittle ice-shell thickness at the time of formation. The depth-to-diameter ratios of craters, such as those studied by Bland et al. (2012, 2018) on Enceladus and Ganymede respectively, are often used across the solar system to understand the thermal history of the body. On Ceres, whose relic ocean is now likely 98–99% frozen (Castillo-Rogez et al. 2019, 2020), a central brine peak and brine deposits known as *faculae* within the prominent Occator crater may point to the remobilization of ancient brines (remnants of the ancient subsurface ocean) and their subsequent surface eruption, proving key insights into the past evolution and current state of Ceres' rocky and icy crust (Scully et al. 2020).

26.3.1.2 Tectonic Features

Tectonic deformation is common throughout the ice-covered ocean worlds of the solar system. Here, tectonic deformation refers to the permanent brittle or plastic failure associated with faults and fractures that relieve large-scale geologic stresses and strains. These processes are central to the study of the thermal and mechanical substrates in which tectonic features form, because they are sensitive to the rate of deformation of the subsurface material, the temperature at depth, frictional properties,

Fig. 26.9 The Tyre impact structure on Europa (NASA/JPL-Caltech). Among the largest impact remnants on Europa, the flat, multi-ring structure of Tyre may indicate a large impactor that penetrated through the icy lithosphere, reaching warm interior ice and potentially the subsurface ocean below. The center ~15 km smooth patch of material may be related to impact melt or a liquid water source beneath the icy crust. The black stripe is a region of missing data



crystal structure, thermal properties, the duration of activity, and the time elapsed since the cessation of activity.

These tectonics may be organized as rigid, plate-like deformation, sharing many of the qualitative aspects of plate tectonics on Earth (Patterson et al. 2006). However, the sources of stresses in ocean-world ice shells are quite different from the gravity-driven global convection cycle responsible for Earth's resurfacing (Howell and Pappalardo 2019). From inferences of surface cracks, we understand that on these worlds, gravitational tidal interactions with the giant planets may induce cyclic shear stresses of ~1 bar each orbit, driving failure to depths of <1 km on the larger ocean worlds (Cameron et al. 2018, 2019, 2020; Hoppa et al. 1999; Hurford et al. 2007), and potentially driving crack opening to the ocean or interstitial reservoir feeding the plumes of Enceladus (e.g., Hedman et al. 2013).

Failure at greater depths may occur if the ice-shell thickness changes, changing the radius and surface area of the body as water freezes and thaws. For example, analytical modeling has demonstrated that freezing can induce shear stresses of up to 100 bar on Europa, driving failure through the full lithosphere and stalling only after reaching the ductile asthenosphere (Nimmo 2004). Similar stresses may be generated on worlds where the ice shell is decoupled in its rotation from the silicate interior, experiencing non-synchronous rotation if the poles remain aligned and true polar wander if the poles misalign through time (Wahr et al. 2009).

Additional studies of individual landforms on the surfaces of planetary ice shells use topographic heights and visible landform measurements to probe the interior. For example, analytical modeling of the flexure surrounding elastically supported topography is often employed to determine the elastic thickness and modulus of an ice shell at the time the analyzed feature formed, as well as the temperature at depth and the time elapsed since formation (Billings and Kattenhorn 2005, and references

therein; Giese et al. 2008; Hurford et al. 2005; Nimmo et al. 2003a; Nimmo and Pappalardo 2004).

26.3.1.3 Other Landforms

While impacts and tectonic deformation most commonly modify ocean-world surfaces to produce landforms, many worlds exhibit strange and unique geology that offers special insight into the processes at depth. Among the most prominent uncommon landforms are those associated with potential cryovolcanism, where liquid water, warm ice, or a slurry of the two erupt onto a planetary surface to produce constructional features and lobate flows. On Europa, cryovolcanism is suggested as a potential source for chaos formation—a landform comprising rotated and translated blocks of surface material mobilized within a fine ice matrix (Schmidt et al. 2011). Additionally, cryovolcanism has been invoked to explain smooth deposits and lobate flows on Ganymede (Patterson et al. 2010) and Pluto (Cruikshank et al. 2019), and the prominent Ahuna Mons mound on Ceres (Ruesch et al. 2016, 2019). Despite widespread evidence of cryovolcanism, no active cryovolcano has been definitively observed on any ocean world.

Another prominent source of surface deformation may be the solid-state convection of ductile ice within the deeper regions of outer planetary ice shells. Convection produces long-wavelength cells of alternating upwelling and downwelling whose size scales with the thickness of the icy shell (Peddinti and McNamara 2019). Thermal convection has been proposed as one potential source of Europa's chaos, pits, and domes (Collins and Nimmo 2009), as well as the source of long-wavelength topography on Enceladus (Besserer et al. 2013). The spectacular nitrogen-ice plates of Pluto's Sputnik Planum may be the surface expression of convection (Trowbridge et al. 2016), and the "cantaloupe terrain" of Triton may be the result of compositional diapirism (Schenk and Jackson 1993).

Among the most unique landforms in the solar system are the aeolian and fluvial landscapes of Titan, whose atmospheric surface pressure exceeds that of Earth (Table 26.2) and is subject to a three-phase methane cycle. Aeolian and fluvial weathering on Titan have produced a porous upper icy crust that experiences lake and cavern formation, exhibits vast swaths of wind-blown organic dunes, and dissects plateaus of organic-rich ices to form the labyrinth terrains (Jaumann et al. 2010; Lopes et al. 2020). Of all of the landforms decorating the surfaces of the ocean worlds, only those of Titan have been touched by a robotic spacecraft. In 2005, ESA's *Huygens* lander was deployed by NASA's *Cassini* spacecraft, where it descended through the atmosphere by parachute before a soft landing (Fig. 26.10). Titan's surface is the target of NASA's recently selected New Frontiers-class Dragonfly mission, which will deploy a rotorcraft, flight-capable lander into the atmosphere for reconnaissance of the organic dunes (Lorenz et al. 2018).

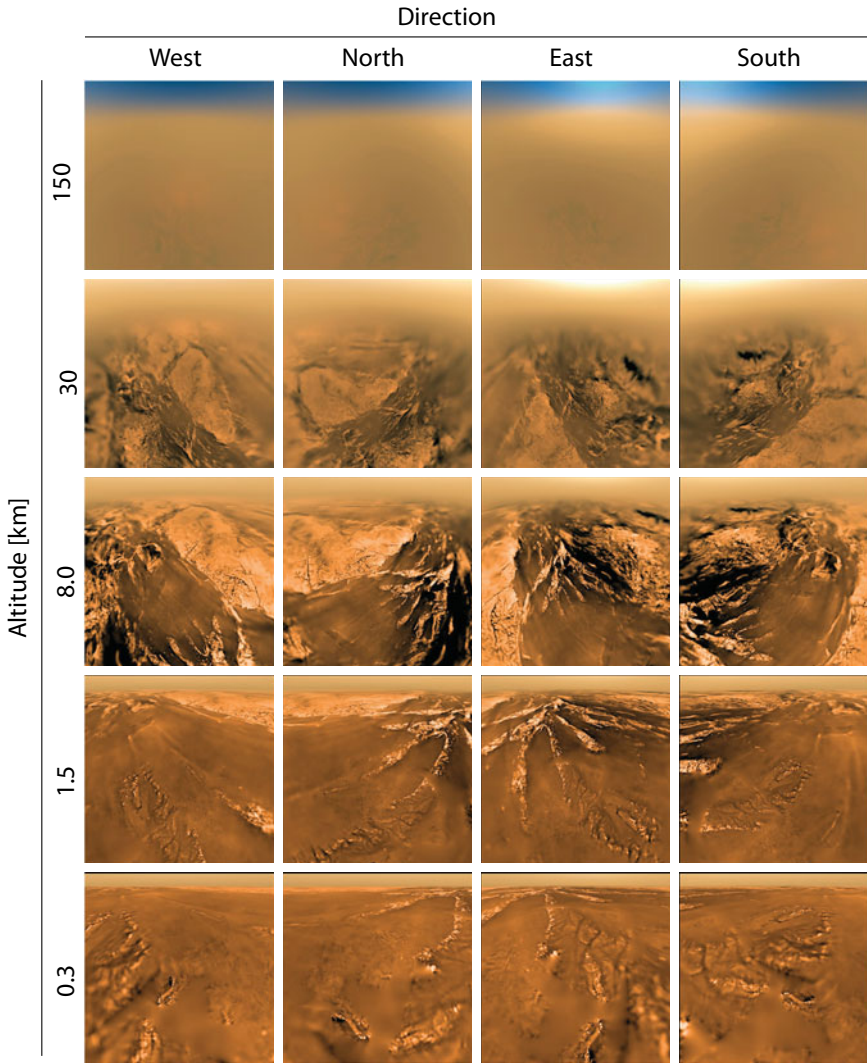


Fig. 26.10 Images from the *Huygens* lander taken on descent to the surface of Titan, showing light-colored uplifted terrain, drainage channels from fluvial methane erosion, and dark low-lying regions (ESA/NASA/JPL-Caltech). Note that these images are highly distorted by the imager's optics

26.3.2 Numerical Methods

Numerical modeling of ocean-world interiors is a primary method of deriving constraints about the icy and ocean environments. In its most basic form, planetary numerical modeling incorporates spacecraft and laboratory data and inferences

to place constraints on the physical relationships that govern the mass, momentum, and energy balances of geologic media. Numerical models of ocean-world interiors are applied to a wide range of pertinent problems from ice-shell impacts to tectonic deformation to surface–ocean interaction. In the following subsections, we briefly review some of these methods and applications of numerical modeling to the geologic state, deformation, and change within ocean-world interiors.

26.3.2.1 Modeling Approaches

To understand how the energy associated with impacts is partitioned into material deformation and heat, and the influence of these processes on the surface structure and interior state of a body, planetary physicists employ a wide range of tools. Familiar continuum models that follow numerical finite element and finite difference schemes may be used to study the response of a perturbed ice shell through geologic time after an impact event. For example, numerical models of an icy continuum have been applied to understand how large impacts into ice shells, like those observed on Europa, may form long-lived melt reservoirs that later erupt cryovolcanically and create water vapor plumes (Steinbrügge et al. 2020).

Among the most prevalent tools applied to the numerical modeling of impacts are hydrodynamical simulations produced by hydrocodes (Pierazzo and Collins 2004). Unlike continuum models that consider the post-impact evolution of craters and subsurface structures, hydrocodes are better suited to study the physics associated with impact shocks because they resolve small-timescale, high-stress and strain, high-temperature processes. Hydrocodes of different complexity may consider an impact target that undergoes shock melting, vaporization, breakup into discontinuous media, deformation, and subsequent relaxation.

Sophisticated hydrocodes have been used to infer the post-impact thermal and mechanical state of icy shells across the ocean worlds, including the formation of Pluto’s Sputnik Planum, where modeling is used to infer subsurface ocean thickness (Johnson et al. 2016). On Titan, researchers have investigated impacts into a hydrocarbon-rich surface, the creation of subsurface melt reservoirs, and the ultimate fate of surface organics (Artemieva and Lunine 2003). At Ganymede, hydrocodes have been used to interpret geologic interpretations of an ancient giant impact that formed a global system of tectonic features known as furrows, changing both the surface and interior structure of the ocean world. Geological observations have also relied on hydrocodes for interpretations of the relic ocean world Ceres, where it has been inferred that the large Occator crater-forming impact remobilized ancient frozen brine (Bowling et al. 2019; Scully et al. 2020).

26.3.2.2 Tectonic and Geodynamic Modeling

Geodynamic numerical modeling is one of the most powerful tools for gaining insight into unexplored depths of the ocean worlds, and indeed all solid bodies throughout

the solar system, including Earth. Geodynamic models generally consider continuous media, applying governing equations that enforce the conservation of mass, energy, and momentum (e.g., Gerya 2010). Depending on their complexity, these models consider various laws regarding the rate at which a geologic material deforms when a stress is applied, including the viscous flow of solid, crystalline media over geologic timescales through various “creep” mechanisms, the permanent plastic deformation of these media, and the storage and release of elastic energy. These relationships are included as a rheological constitutive law. Additionally, geodynamic numerical models may consider the thermal convection, fracturing, faulting, melting, freezing, internal heating, and multiphase flow of geologic materials.

Among the first geodynamic models of ocean worlds were those of thermal convection, simulating the gravitationally driven motion of solid ice in response to the positive buoyancy of deep warm ice and negative buoyancy of cold surface ice. As convection modeling advances, current work considers when the onset of convection may occur in ocean-world ice shells (Barr and McKinnon 2007b; Mitri and Showman 2005), how thermal properties affect convective states (Carnahan et al. 2021), the evolution of convection in a planetary ice shell as it changes thickness (Peddinti and McNamara 2019), and even the fully three-dimensional occurrence and evolution of melts that form in tidally heated ice shells (Vilella et al. 2020). Models of multiphase and thermo-compositional convection are also being employed to understand the unique and complex challenges associated with ice–ocean interface dynamics (Buffo et al. 2021a) and the presence of non-ice materials (Leonard and Howell 2019).

Geodynamic models of icy failure have been widely used to interpret subsurface processes and structures related to faults and fractures visible on icy satellite surfaces. Coupled convective-tectonic models have been used to explore the role of extensional tectonics on Ganymede and Europa, predicting the entrainment, distribution, and surface exposure of frozen, “fossilized” ocean material (Fig. 26.11) (Howell and Pappalardo 2018). Tectonic models have also been constrained by analyses of fault and fracture spacing (Bland and McKinnon 2015) and employed to understand the thermal implications at depth of active fracture activity (Abramov and Spencer 2009; Hammond 2020). Such modeling can help us understand even poorly observed worlds, such as the Uranian moon Triton, where fracture heating modeling and comparative planetology help offer a glimpse of subsurface processes and structure (Prockter et al. 2005). One-dimensional geodynamic models of Pluto’s thermal evolution have been used to explore how phase changes can drive tectonic activity at the energy-poor edges of the solar system (Hammond et al. 2016).

The role of water in ice-shell interior is central to studies of ocean worlds’ habitability and to descriptions of their interior environments. Simplistic models of the pressurization of freezing reservoirs, for example, have been used to understand the potential for the surface eruption of subsurface water melts (Lesage et al. 2020). Geodynamic models of two-phase flow that consider the motion of liquid water and the evolution of solid ice have been used to investigate the downward drainage of liquid water through outer ice shells (Kalousová et al. 2016), the upward buoyant rise

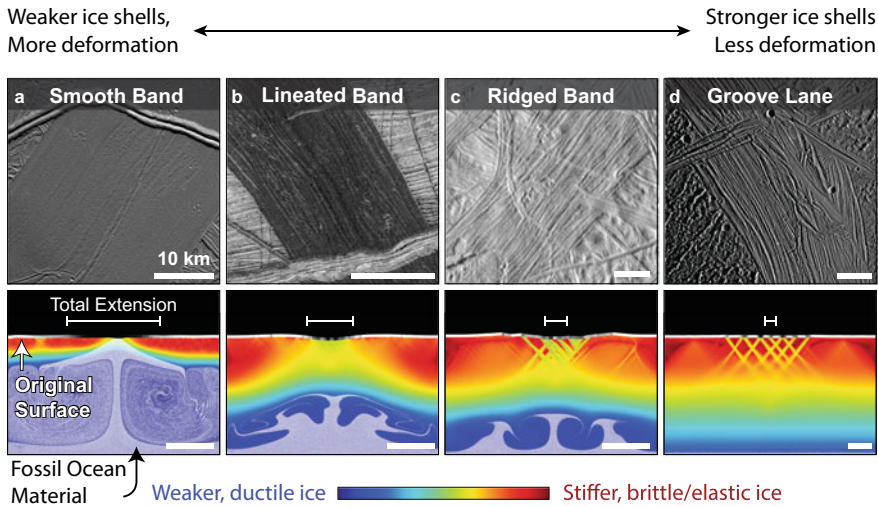


Fig. 26.11 Shown are (top) Galileo spacecraft images in map view with (bottom) associated geodynamic numerical models in cross-section, modified from Howell and Pappalardo (2018). White bars represent a 10 km scale. Images show extensional bands of different morphologies, where the icy lithosphere has pulled apart to expose warm interior ice. Images are of **a** Europa's smooth band Thynia Linea, **b** Europa's lineated band Yelland Linea, **c** a ridged band in Europa's northern leading hemisphere, and **d** Ganymede groove lane Tiamat Sulcus. Model outputs are colored by ice viscosity, with warmer colors related to stiffer, more brittle ice, and showing frozen, fossilized ocean material (white dots within ductile ice). Green and tallow bands represent deep tectonic faults. For the smooth band shown in **(a)**, fossil ocean material is exposed at the icy surface, but not in the models of more tectonized terrains shown in **(b–d)**. The model surfaces are colored white where the original surface has not been significantly disrupted, and black where the model surface strain is the greatest and fresh interior ice is exposed

of liquid water through dense high-pressure ice phases at the sea-floor (Kalousova et al. 2018), and the local-scale evolution of ice–ocean interfaces (Buffo et al. 2020).

26.3.2.3 Ocean Modeling

The field of comparative oceanography is rapidly developing, and scientists are widely utilizing tools developed for the study of Earth's oceans to study oceans throughout our solar system. These modeling methods may employ conservation of mass, energy, and momentum, sometimes expressed in the form of the Laplace tidal equations. These tidal equations explain how the gravitation potential of two bodies can drive changes in the orbit and the figure of the body, and produce tides. Modeling of planetary oceans helps to infer where stratified oceans that promote compositional and thermal variations as a function of depth may occur, and where oceans may be well mixed compositionally and thermally. These modeling methods

are key to understanding the habitability of ocean worlds, and couple the rocky and icy environments at depth.

Numerical models of interior oceans have helped constrain global energy balances, contributing to our understanding of how orbital energy is dissipated as heat (Hay et al. 2020; Hay and Matsuyama 2019). Additionally, models of planetary oceans predict the time and length scales of ocean convection (Soderlund 2019). For example, modeling the circulation within the ocean of Enceladus, constrained by observational evidence, produces ocean model predictions of pole-to-equator convective overturn at depth, with profound implications for the distribution of non-ice materials, ice–ocean interface melting, and freezing (Lobo et al. 2021).

Within NASA and the broader planetary science and astrobiology communities, efforts are ramping up to leverage new and broader expertise in modeling and understanding planetary oceans from terrestrial oceanographers. Groups like NASA's Network for Ocean Worlds connect modelers across bodies (Howell et al. 2020), and these exchanges of knowledge across fields will surely enable new breakthroughs in ocean modeling and comparative oceanography in the decades to come.

26.3.2.4 Statistical Methods

Throughout studies of ocean worlds, statistical methods are widely employed to improve geologic inferences, provide an interface for disparate modeling disciplines, and to produce quantitative predictions. Statistical methods applied to icy interiors can provide insight into geologic landform formation, interior environments and processes, temporal and spatial distributions of energy and change within planetary ice shells and oceans, and extend even to studies of planetary protection. One of the most powerful advantages of statistical modeling is that uncertainties are often quantifiable, and can be used to derive requirements and margin for future observations, spacecraft, and instrument design.

One key approach for making inferences about processes at depth from spacecraft observations is to characterize the spatial, temporal, and size-frequency distributions of surface features, including craters, faults, fractures, chaos, pits, domes, lakes, and more. These techniques are common and varied, with examples including inferences about how Titan's orbital interactions drive surface and subsurface change made from asymmetric distributions of lakes on the surface (Aharonson et al. 2009), comparisons of chaotic terrain size-frequency distributions across the ocean worlds to infer their internal properties (Skjetne et al. 2021) and the energy of formation, which may highlight their thermal and mechanical structure at depth (Leonard et al. 2021a).

Statistical modeling has also allowed the quantification of uncertainties in key parameters that govern the interior processes and environments of ocean worlds. For example, Howell (2021) creates a statistical model of Europa's internal structure, including probability distributions and current best estimates for the total ice shell and layer thicknesses, the depth to which porosity extends, the thermal and mechanical properties of the icy shell, and the effects of salinity. By both identifying current best estimate values and characterizing uncertainty, these results have been directly

incorporated into the preliminary design of potential future spacecraft (Fleuriel et al. 2019; Woerner et al. 2020), and may provide useful constraints for future observations and instrument design. This use case for constraining spacecraft design and operations is explored in Sect. 26.4.

Applying statistical approaches to characterizing geological and geophysical problems also allows for the synthesis of geological predictions with other fields. One unique example has been in the planetary protection approach of Europa Clipper, which uses a probabilistic risk assessment to explore whether the spacecraft conforms with the legal requirement that the probability of contaminating Europa's long-lived liquid water bodies within 1,000 years of the mission is below 10^{-5} (McCoy et al. 2020). In this case, statistical models of Europa's resurfacing, driven by its interior processes, were coupled to statistical predictions of potential inadvertent spacecraft impacts, the resulting distribution of debris, and subsequent biological exploration.

26.4 Example Application to Europa

In this section, we show how scientific inferences of ocean-world processes can be used to build descriptions of the pressure, temperature, porosity, stress, and strain rate environments within planetary ice shells. Howell (2021) performs a Monte Carlo analysis to estimate a probability for ice-shell thickness on Europa (Fig. 26.12a), and ingest ~ 20 prior distributions to simulate 10^7 potential ice-shell configurations. During this analysis, underlying distributions are produced for the thickness of the constituent conductive and convective layer (Fig. 26.12b), their thermal conductivity, their density, the porosity structure with depth, the temperature of transition between conductive and convective states, the isothermal temperature of convection, and the ice–ocean interface temperature.

From the distributions derived within this Monte Carlo model, temperature and pressure profiles can be determined, and previous studies can be used to predict ice velocity and stress conditions as a function of depth. For example, a current best estimate (CBE) ice shell of Europa is shown in Fig. 26.13.

These interior profiles offer constraints that allow for spacecraft, instrument, and observation design and planning. For example, chief considerations in the design of melt probes, a conceptual technology for reaching the subsurface ocean of an ocean world, include the total heat density required to melt the probe's surroundings and make progress towards the ocean. These design considerations are a function of temperature, thermal conductivity, local pore pressure, and pressure vessel design, which depends on the icy and water overburden pressure (Dachwald et al. 2021; Fleuriel et al. 2019; Hockman et al. 2021).

This approach to interior modeling can also be used to derive challenge cases for the design of spacecraft, instruments, and observations. For example, the Monte Carlo study of Howell (2021) shown above has been used to help size cryobot concept designs (Fleuriel et al. 2019), size power sources for ocean worlds (Woerner

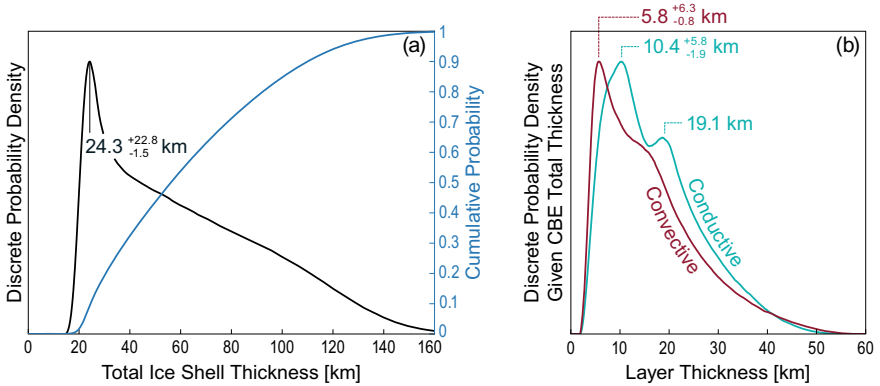


Fig. 26.12 Icy thickness probability distributions predicted by the Europa Monte Carlo model of Howell (2021). **a** The highly-skewed discrete probability density distribution (black line) for total icy shell thickness, and cumulative probability distribution (blue line). **b** Conditional discrete probability distributions for convective and conductive thickness, given the current best estimate total shell thickness

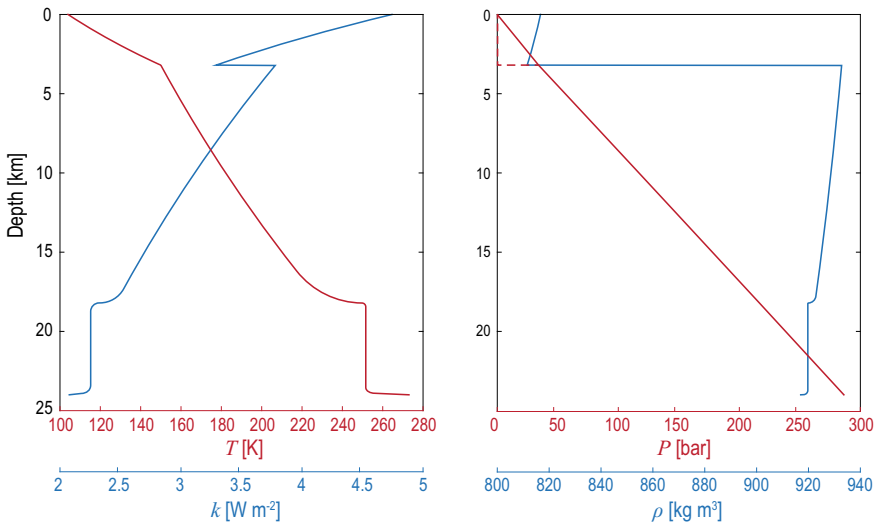


Fig. 26.13 Current best estimate interior model for Europa, derived from the Monte Carlo approach of Howell (2021), exhibiting a porous lithosphere within a conductive layer and convection at depth. Plots show example (left) thermal profiles (red) and thermal conductivity profiles (blue), and (right) physical profiles of pressure (red) and density (blue). The dashed red pressure line shows the pore pressure at depth in this model, rather than the lithostatic pressure. High porosity keeps icy pores and fractures connected to the vacuum over the first ~3 km

et al. 2020), and define science payload and operations requirements for in-situ observations (Klonicki et al. 2021).

Thus, the analyses and inferences of ocean-world interior processes have utility beyond their geological and geophysical science value. From mapping surface geology to numerical models of icy tectonics, to studies of the formation of ocean worlds, each individual area of contribution builds towards a practical understanding of these environments.

26.5 Summary

Ocean worlds exploration is an exciting new development in planetary science. From the myriad ocean worlds that we already know exist in our solar system, it is clear that each is unique and worthy of focused study. The numerous unanswered questions regarding the evolution and current state of ocean worlds bring together scientists and engineers from a wide range of planetary, oceanographic, geological, geochemical, and astrobiological disciplines—from aquatic geomicrobiology, to high-pressure material science, to fracture mechanics, to theoretical and computational modeling, and many others. By enhancing the potential to discover life beyond Earth within the next few decades, rather than centuries, ocean worlds exploration has risen to the forefront of planetary exploration.

Acknowledgements We thank M Melwani Daswani and J Scully for their review of this chapter. This work was carried out at the NASA Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Portions of this work were funded by the JPL Probe using Radioisotopes for Icy Moons Exploration (PRIME) technology development program, and by NASA Scientific Exploration and Subsurface Access Mechanism for Europa (SESAME) grant #80NSSC19K0614.

References

- Abramov, O., and J.R. Spencer. 2009. Endogenic heat from Enceladus' south polar fractures: New observations, and models of conductive surface heating. *Icarus* 199: 189–196. <https://doi.org/10.1016/j.icarus.2008.07.016>.
- Aharonson, O., A.G. Hayes, J.I. Lunine, R.D. Lorenz, M.D. Allison, and C. Elachi. 2009. An asymmetric distribution of lakes on Titan as a possible consequence of orbital forcing. *Nature Geoscience* 2: 851–854. <https://doi.org/10.1038/ngeo698>.
- Artemieva, N., and J. Lunine. 2003. Cratering on Titan: Impact melt, ejecta, and the fate of surface organics. *Icarus* 164: 471–480.
- Ashkenazy, Y. 2016. The surface temperature of Europa. ArXiv160807372 Astro-Ph.
- Barr, A.C., and W.B. McKinnon. 2007a. Convection in ice I shells and mantles with self-consistent grain size. *Journal Geophysical Research Planets* 112. <https://doi.org/10.1029/2006JE002781>.
- Barr, A.C., and W.B. McKinnon. 2007b. Convection in Enceladus' ice shell: Conditions for initiation. *Geophysical Research Letter* 34. <https://doi.org/10.1029/2006GL028799>.

- Běhounková, M., G. Tobie, G. Choblet, M. Kervazo, M. Melwani Daswani, C. Dumoulin, and S.D. Vance. 2021. Tidally Induced Magmatic Pulses on the Oceanic Floor of Jupiter's Moon Europa. *Geophysical Research Letters*, 48(3), e2020GL090077.
- Besserer, J., F. Nimmo, J.H. Roberts, and R.T. Pappalardo. 2013. Convection-driven compaction as a possible origin of Enceladus's long wavelength topography. *Journal Geophysical Research Planets* 118: 908–915.
- Bierhaus, E.B., K. Zahnle, C.R. Chapman, R.T. Pappalardo, W.R. McKinnon, and K.K. Khurana. 2009. Europa's crater distributions and surface ages. *Europa*, 161–180.
- Billings, S.E., and S.A. Kattenhorn. 2005. The great thickness debate: Ice shell thickness models for Europa and comparisons with estimates based on flexure at ridges. *Icarus. Europa Icy Shell* 177: 397–412. <https://doi.org/10.1016/j.icarus.2005.03.013>.
- Bland, M.T., and W.B. McKinnon. 2018. Viscous relaxation as a prerequisite for tectonic resurfacing on Ganymede: Insights from numerical models of lithospheric extension. *Icarus* 306: 285–305. <https://doi.org/10.1016/j.icarus.2017.10.017>.
- Bland, M.T., and W.B. McKinnon. 2015. Forming Ganymede's grooves at smaller strain: Toward a self-consistent local and global strain history for Ganymede. *Icarus* 245: 247–262. <https://doi.org/10.1016/j.icarus.2014.09.008>.
- Bland, M.T., K.N. Singer, W.B. McKinnon, and P.M. Schenk. 2012. Enceladus' extreme heat flux as revealed by its relaxed craters. *Geophysical Research Letter* 39. <https://doi.org/10.1029/2012GL052736>.
- Bowling, T.J., F.J. Ciesla, T.M. Davison, J.E. Scully, J.C. Castillo-Rogez, S. Marchi, and B.C. Johnson. 2019. Post-impact thermal structure and cooling timescales of Occator crater on asteroid 1 Ceres. *Icarus* 320: 110–118.
- Buffo, J.J., C.R. Meyer, and J.R.G. Parkinson. 2021a. Dynamics of a Solidifying Icy Satellite Shell. *Journal Geophysical Research Planets* n/a, e2020JE006741. <https://doi.org/10.1029/2020JE006741>.
- Buffo, J.J., B.E. Schmidt, C. Huber, and C.R. Meyer. 2021b. Characterizing the ice-ocean interface of icy worlds: A theoretical approach. *Icarus* 360: 114318.
- Buffo, J.J., B.E. Schmidt, C. Huber, and C.C. Walker. 2020. Entrainment and dynamics of ocean-derived impurities within Europa's Ice Shell. *Journal Geophysical Research Planets* 125, e2020JE006394.
- Cameron, M.E., B.R. Smith-Konter, L. Burkhard, G.C. Collins, F. Seifert, and R.T. Pappalardo. 2018. Morphological mapping of Ganymede: Investigating the role of strike-slip tectonics in the evolution of terrain types. *Icarus* 315: 92–114. <https://doi.org/10.1016/j.icarus.2018.06.024>.
- Cameron, M.E., B.R. Smith-Konter, G.C. Collins, D.A. Patthoff, and R.T. Pappalardo. 2020. Ganymede, then and now: How past eccentricity may have altered tidally driven coulomb failure. *Journal Geophysical Research. Planets* 125, e2019JE005995. <https://doi.org/10.1029/2019JE005995>.
- Cameron, M.E., B.R. Smith-Konter, G.C. Collins, D.A. Patthoff, and R.T. Pappalardo. 2019. Tidal stress modeling of Ganymede: Strike-slip tectonism and Coulomb failure. *Icarus* 319: 99–120. <https://doi.org/10.1016/j.icarus.2018.09.002>.
- Carnahan, E., N.S. Wolfenbarger, J.S. Jordan, and M.A. Hesse. 2021. New insights into temperature-dependent ice properties and their effect on ice shell convection for icy Ocean Worlds. *Earth and Planet. Science Letter* 563: 116886. <https://doi.org/10.1016/j.epsl.2021.116886>.
- Carr, M.H., M.J.S. Belton, C.R. Chapman, M.E. Davies, P. Geissler, R. Greenberg, A.S. McEwen, B.R. Tufts, R. Greeley, R. Sullivan, J.W. Head, R.T. Pappalardo, K.P. Klaasen, T.V. Johnson, J. Kaufman, D. Senske, J. Moore, G. Neukum, G. Schubert, J.A. Burns, P. Thomas, and J. Veverka. 1998. Evidence for a subsurface ocean on Europa. *Nature* 391: 363–365. <https://doi.org/10.1038/34857>.
- Castillo-Rogez, J.C., M.A. Hesse, M. Formisano, H. Sizemore, M. Bland, A.I. Ermakov, and R.R. Fu. 2019. Conditions for the long-term preservation of a deep brine reservoir in Ceres. *Geophysical Research Letters* 46 (4): 1963–1972.

- Castillo-Rogez, J.C., M. Neveu, J.E. Scully, C.H. House, L.C. Quick, A. Bouquet, K. Miller, M. Bland, M.C. De Sanctis, and A. Ermakov. 2020. Ceres: Astrobiological target and possible Ocean World. *Astrobiology* 20: 269–291.
- Collins, G.C., and F. Nimmo. 2009. Chaotic terrain on Europa, in: *Europa*. University of Arizona Press.
- Crow-Willard, E.N., and R.T. Pappalardo. 2015. Structural mapping of Enceladus and implications for formation of tectonized regions. *Journal Geophysical Research Planets* 120: 928–950. <https://doi.org/10.1002/2015JE004818>.
- Cruikshank, D.P., O.M. Umurhan, R.A. Beyer, B. Schmitt, J.T. Keane, K.D. Runyon, D. Atri, O.L. White, I. Matsuyama, J.M. Moore, W.B. McKinnon, S.A. Sandford, K.N. Singer, W.M. Grundy, C.M. Dalle Ore, J.C. Cook, T. Bertrand, S.A. Stern, C.B. Olkin, H.A. Weaver, L.A. Young, J.R. Spencer, C.M. Lisse, R.P. Binzel, A.M. Earle, S.J. Robbins, G.R. Gladstone, R.J. Cartwright, and K. Ennico. 2019. Recent cryovolcanism in Virgil Fossae on Pluto. *Icarus* 330: 155–168. <https://doi.org/10.1016/j.icarus.2019.04.023>.
- Dachwald, B., J. Kowalski, J. Biele, F. Baader, and N. Komle. 2021. Ice Melting Probes, in: *Handbook of Space Resources*. Springer.
- Eviatar, A., M. Vasyliūnas, V.A. and Gurnett, D. 2001. The ionosphere of Ganymede. *Planet. Space Science, Magnetospheres of the Outer Planets* (Part I) 49, 327–336. [https://doi.org/10.1016/S0032-0633\(00\)00154-9](https://doi.org/10.1016/S0032-0633(00)00154-9).
- Flourial, J.-P., S. Howell, D. Woerner, D. Landau, J. Gayle, M. Smith, J. Mueller, T. Hendricks, S. Roberts, and B. Nesmith. 2019. Notional concept of operations and system capability definition for enabling scientific ocean access missions on icy worlds. EPSC 2019, EPSC-DPS2019.
- Gerya. 2010. *Introduction to Numerical Geodynamic Modelling*. Cambridge University Press, New York.
- Giese, B., R. Wagner, H. Hussmann, G. Neukum, J. Perry, P. Helfenstein, and P.C. Thomas. 2008. Enceladus: An estimate of heat flux and lithospheric thickness from flexurally supported topography. *Geophysical Research Letter* 35. <https://doi.org/10.1029/2008GL036149>.
- Glein, C.R., J.A. Baross, and J.H. Waite Jr. 2015. The pH of Enceladus' ocean. *Geochimica Et Cosmochimica Acta* 162: 202–219.
- Glein, C.R., and J.H. Waite. 2020. The carbonate geochemistry of Enceladus' ocean. *Geophysical Research Letters* 47 (3), e2019GL085885.
- Glombitza, C., L.I. Putman, K.R. Rempfert, M.D. Kubo, M.O. Schrenk, A.S. Templeton, and T.M. Hoehler. 2021. Active microbial sulfate reduction in fluids of serpentinizing peridotites of the continental subsurface. *Communication Earth Environment* 2: 1–9. <https://doi.org/10.1038/s43247-021-00157-z>.
- Grasset, O., M.K. Dougherty, A. Coustenis, E.J. Bunce, C. Erd, D. Titov, M. Blanc, A. Coates, P. Drossart, and L.N. Fletcher. 2013. JUPITER ICy moons Explorer (JUICE): An ESA mission to orbit Ganymede and to characterise the Jupiter system. *Planetary and Space Science* 78: 1–21.
- Greeley, R., J.E. Klemaszewski, and R. Wagner. 2000. Galileo views of the geology of Callisto. *Planetary and Space Science* 48: 829–853. [https://doi.org/10.1016/S0032-0633\(00\)00050-7](https://doi.org/10.1016/S0032-0633(00)00050-7).
- Grott, M., F. Sohl, and H. Hussmann. 2007. Degree-one convection and the origin of Enceladus' dichotomy. *Icarus* 191: 203–210. <https://doi.org/10.1016/j.icarus.2007.05.001>.
- Hall, D.T., D.F. Strobel, P.D. Feldman, M.A. McGrath, and H.A. Weaver. 1995. Detection of an oxygen atmosphere on Jupiter's moon Europa. *Nature* 373: 677–679. <https://doi.org/10.1038/373677a0>.
- Hammond, N.P. 2020. Estimating the magnitude of cyclic slip on strike-slip faults on Europa. *Journal Geophysical Research. Planets* n/a, e2019JE006170. <https://doi.org/10.1029/2019JE006170>.
- Hammond, N.P., A.C. Barr, and E.M. Parmentier. 2016. Recent tectonic activity on Pluto driven by phase changes in the ice shell. *Geophysical Research Letters* 43: 6775–6782. <https://doi.org/10.1002/2016GL069220>.
- Hand, K.P., R.W. Carlson, and C.F. Chyba. 2007. Energy, chemical disequilibrium, and geological constraints on Europa. *Astrobiology* 7: 1006–1022. <https://doi.org/10.1089/ast.2007.0156>.

- Hand, K.P., and C.F. Chyba. 2007. Empirical constraints on the salinity of the European ocean and implications for a thin ice shell. *Icarus* 189: 424–438.
- Hand, K.P., C.F. Chyba, J.C. Priscu, R.W. Carlson, and K.H. Nealson. 2009. Astrobiology and the Potential for Life on Europa. In *Europa*, ed. R.T. Pappalardo, W.B. McKinnon, and K. Khurana, 589. University of Arizona Press, Tucson: The University of Arizona Space Science Series.
- Hand, K.P., and C.R. German. 2018. Exploring ocean worlds on earth and beyond. *Nature Geoscience* 11: 2–4. <https://doi.org/10.1038/s41561-017-0045-9>.
- Harri, A.-M., T. Mäkinen, A. Lehto, H. Kahanpää, and T. Siili. 2006. Vertical pressure profile of Titan—observations of the PPI/HASI instrument. *Planet. Space Sci. Surfaces and Atmospheres of the Outer Planets, Their Satellites and Ring Systems from Cassini-Huygens Data* 54: 1117–1123. <https://doi.org/10.1016/j.pss.2006.05.037>.
- Hay, H.C.F.C., and I. Matsuyama. 2019. Nonlinear tidal dissipation in the subsurface oceans of Enceladus and other icy satellites. *Icarus* 319: 68–85. <https://doi.org/10.1016/j.icarus.2018.09.019>.
- Hay, H.C.F.C., A. Trinh, and I. Matsuyama. 2020. Powering the Galilean satellites with moon-moon tides. *Geophysical Research Letter* 47, e2020GL088317. <https://doi.org/10.1029/2020GL088317>.
- Hedman, M.M., C.M. Gossmeier, P.D. Nicholson, C. Sotin, R.H. Brown, R.N. Clark, K.H. Baines, B.J. Buratti, and M.R. Showalter. 2013. An observed correlation between plume activity and tidal stresses on Enceladus. *Nature* 500: 182–184. <https://doi.org/10.1038/nature12371>.
- Hemingway, D., L. Iess, R. Tajeddine, and G. Tobie. 2018. The interior of Enceladus. *Enceladus Icy Moons Saturn* 57.
- Hendrix, A.R., T.A. Hurford, L.M. Barge, M.T. Bland, J.S. Bowman, W. Brinckerhoff, B.J. Buratti, M.L. Cable, J. Castillo-Rogez, and G.C. Collins. 2019. The NASA roadmap to ocean worlds. *Astrobiology* 19: 1–27.
- Hockman, B.J., Y. Bar-Cohen, M. Smith, and J.P. Fleurial. 2021. Navigation and control of Cyrobots as robotic penetrators to reaching water under extremely deep ice crest in Ocean Worlds, in: *Handbook of Space Resources*. Springer.
- Hoppa, G.V., B.R. Tufts, R. Greenberg, and P.E. Geissler. 1999. Formation of cycloidal features on Europa. *Science* 285: 1899–1902. <https://doi.org/10.1126/science.285.5435.1899>.
- Howell, S.M. 2021. The likely thickness of Europa's Icy shell. *The Planetary Science Journal* 2 (4): 129.
- Howell, S.M., and R.T. Pappalardo. 2020. NASA's Europa Clipper—a mission to a potentially habitable Ocean World. *Nature Communications* 11: 1311. <https://doi.org/10.1038/s41467-020-15160-9>.
- Howell, S.M., and R.T. Pappalardo. 2019. Can Earth-like plate tectonics occur in Ocean World ice shells? *Icarus*. <https://doi.org/10.1016/j.icarus.2019.01.011>.
- Howell, S.M., and R.T. Pappalardo. 2018. Band formation and ocean-surface interaction on Europa and Ganymede. *Geophysical Research Letters*. <https://doi.org/10.1029/2018GL077594>.
- Howell, S.M., W.C. Stone, K. Craft, C. German, A. Murray, A. Rhoden, and K. Arrigo. 2020. Ocean worlds exploration and the search for life. ArXiv200615803 Astro-Ph.
- Hurford, T.A., R.A. Beyer, B. Schmidt, B. Preblich, A.R. Sarid, and R. Greenberg. 2005. Flexure of Europa's lithosphere due to ridge-loading. *Icarus, Europa Icy Shell* 177: 380–396. <https://doi.org/10.1016/j.icarus.2005.06.019>.
- Hurford, T.A., A.R. Sarid, and R. Greenberg. 2007. Cycloidal cracks on Europa: Improved modeling and non-synchronous rotation implications. *Icarus* 186: 218–233. <https://doi.org/10.1016/j.icarus.2006.08.026>.
- Jankowski, D.G., C.F. Chyba, and P.D. Nicholson. 1989. On the obliquity and tidal heating of Triton. *Icarus* 80: 211–219.
- Jaumann, R., R.L. Kirk, R.D. Lorenz, R.M.C. Lopes, E. Stofan, E.P. Turtle, H.U. Keller, C.A. Wood, C. Sotin, L.A. Soderblom, and M.G. Tomasko. 2010. Geology and surface processes on Titan. In: Brown, R.H., Lebreton, J.-P., Waite, J.H. (Eds.), *Titan from Cassini-Huygens*. Springer Netherlands, Dordrecht, pp. 75–140. https://doi.org/10.1007/978-1-4020-9215-2_5.

- Jennings, D.E., T. Tokano, V. Cottini, C.A. Nixon, R.K. Achterberg, F.M. Flasar, V.G. Kunde, P.N. Romani, R.E. Samuelson, M.E. Segura, N.J.P. Goriun, E. Guandique, M.S. Kaelberer, and A. Coustenis. 2019. Titan surface temperatures during the Cassini mission. *The Astrophysical Journal* 877: L8. <https://doi.org/10.3847/2041-8213/ab1f91>.
- Jia, X., M.G. Kivelson, K.K. Khurana, and W.S. Kurth. 2018. Evidence of a plume on Europa from Galileo magnetic and plasma wave signatures. *Natural Astronomy* 2: 459–464. <https://doi.org/10.1038/s41550-018-0450-z>.
- Johnson, B.C., T.J. Bowling, A.J. Trowbridge, and A.M. Freed. 2016. Formation of the Sputnik Planum basin and the thickness of Pluto’s subsurface ocean. *Geophysical Research Letters* 43: 10068–10077. <https://doi.org/10.1002/2016GL070694>.
- Kalousová, K., and C. Sotin. 2020. The insulating effect of methane clathrate crust on Titan’s thermal evolution. *Geophysical Research Letters* 47, e2020GL087481.
- Kalousová, K., C. Sotin, G. Choblet, G. Tobie, and O. Grasset. 2018. Two-phase convection in Ganymede’s high-pressure ice layer—Implications for its geological evolution. *Icarus* 299: 133–147. <https://doi.org/10.1016/j.icarus.2017.07.018>.
- Kalousová, K., O. Souček, G. Tobie, G. Choblet, and O. Čadež. 2016. Water generation and transport below Europa’s strike-slip faults. *Journal Geophysical Research Planets* 121: 2444–2462. <https://doi.org/10.1002/2016JE005188>.
- Khurana, K.K., M.G. Kivelson, D.J. Stevenson, G. Schubert, C.T. Russell, R.J. Walker, and C. Polansky. 1998. Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780. <https://doi.org/10.1038/27394>.
- Kirk, R.L., L.A. Soderblom, R.H. Brown, S.W. Kieffer, and J.S. Kargel. 1995. Triton’s plumes: discovery, characteristics, and models. *Neptune Triton*, 949–989.
- Kivelson, M.G., K.K. Khurana, C.T. Russell, M. Volwerk, R.J. Walker, and C. Zimmer. 2000. Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.
- Kivelson, M.G., K.K. Khurana, C.T. Russell, R.J. Walker, J. Warnecke, F.V. Coroniti, C. Polansky, D.J. Southwood, and G. Schubert. 1996. Discovery of Ganymede’s magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.
- Klonicki, E., B.G. Clement, R. Kidd, M. Gonzalez, J.-P. Fleurial, D. Woerner, T.J. Hendricks, and S.M. Howell. 2019. Evaluation and Integration of Potential Instruments for Subsurface Ocean Worlds Missions. *AGUFM* 2019: P51B–P5.
- Klonicki, E.F., M.J. Malaska, V.J. Scott, M.P. Panning, S.E. Waller, and P.J. Gasda. 2021. Instrumentation for planetary exploration. In: *Handbook of Space Resources*. Springer.
- Leonard, E.J., and S. Howell. 2019. Compositional variations within Europa’s Ice shell: Implications for surface geology. In *50th Lunar and Planetary Science Conference 2019*, 2631.
- Leonard, E.J., S.M. Howell, A. Mills, D.A. Senske, D.A. Patthoff, and R.T. Pappalardo. 2021a. Bringing order to chaos: insights on the formation of chaos terrain from geologic mapping of Europa at the regional scale. *Lunar Planet. Science Conference* 2269.
- Leonard, E.J., D.A. Patthoff, and D.A. Senske. 2021b. The geologic map of Europa at 1:15M. *U. S. Geol. Surv.* in press.
- Leonard, E.J., A. Yin, and R.T. Pappalardo. 2021c. Forming relic cratered blocks: Left-lateral shear on Enceladus inferred from ice-shell deformation in the leading hemisphere. *Journal Geophysical Research Planets* 126, e2020JE006499. <https://doi.org/10.1029/2020JE006499>.
- Lesage, E., H. Massol, and F. Schmidt. 2020. Cryomagma ascent on Europa. *Icarus* 335: 113369. <https://doi.org/10.1016/j.icarus.2019.07.003>.
- Liang, M.-C., B.F. Lane, R.T. Pappalardo, M. Allen, and Y.L. Yung. 2005. Atmosphere of Callisto. *Journal Geophysical Research Planets* 110. <https://doi.org/10.1029/2004JE002322>.
- Lobo, A.H., A.F. Thompson, S.D. Vance, and S. Tharimena. 2021. A pole-to-equator ocean overturning circulation on Enceladus. *Nature Geoscience* 14: 185–189. <https://doi.org/10.1038/s41561-021-00706-3>.
- Lopes, R.M.C., M.J. Malaska, A.M. Schoenfeld, A. Solomonidou, S.P.D. Birch, M. Florence, A.G. Hayes, D.A. Williams, J. Radebaugh, T. Verlander, E.P. Turtle, A. Le Gall, and S.D. Wall. 2020.

- A global geomorphologic map of Saturn's moon Titan. *Natural Astronomy* 4: 228–233. <https://doi.org/10.1038/s41550-019-0917-6>.
- Lorenz, R.D., E.P. Turtle, J.W. Barnes, M.G. Trainer, D. Adams, K. Hibbard, C.Z. Sheldon, K. Zacny, P.N. Peplowski, D.J. Lawrence, M.A. Ravine, T.G. McGee, K. Sotzen, S.M. MacKenzie, J. Langelaan, S. Schmitz, L.S. Wolfarth, and P.D. Bedini. 2018. Dragonfly: A rotorcraft lander concept for scientific exploration at titan. Johns Hopkins APL Tech. *Dig. Application Physical Laboratory* 34: 374–387.
- Manga, M., and C. Michaut. 2017. Formation of lenticulae on Europa by saucer-shaped sills. *Icarus* 286: 261–269. <https://doi.org/10.1016/j.icarus.2016.10.009>.
- Martin, W., J. Baross, D. Kelley, and M.J. Russell. 2008. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* 6: 805–814. <https://doi.org/10.1038/nrmicro1991>.
- McCoy, K.J., M. DiNicola, C. Everline, H. Burgoyne, K. Reinholtz, and B. Clement. 2020. Europa Clipper planetary protection probabilistic risk assessment summary? *Planet. Space Science*, 105139. <https://doi.org/10.1016/j.pss.2020.105139>.
- McKinnon, W.B. 1999. Convective instability in Europa's floating ice shell. *Geophysical Research Letters* 26: 951–954.
- McPhee, M.G., J.H. Morison, and F. Nilsen. 2008. Revisiting heat and salt exchange at the ice-ocean interface: Ocean flux and modeling considerations. *Journal Geophysical Research Oceans* 113.
- Mitri, G., and A.P. Showman. 2005. Convective–conductive transitions and sensitivity of a convecting ice shell to perturbations in heat flux and tidal-heating rate: Implications for Europa. *Icarus* 177: 447–460. <https://doi.org/10.1016/j.icarus.2005.03.019>.
- Moore, J., L. Spilker, M. Cable, S. Edgington, A. Hendrix, M. Hofstadter, T. Hurford, K. Mandt, A. McEwen, and C. Paty. 2020. Exploration Strategy for the Outer Planets 2023–2032: Goals and Priorities. ArXiv Prepr. ArXiv200311182.
- Moore, J.M., E. Asphaug, R.J. Sullivan, J.E. Klemaszewski, K.C. Bender, R. Greeley, P.E. Geissler, A.S. McEwen, E.P. Turtle, and C.B. Phillips. 1998. Large impact features on Europa: Results of the Galileo nominal mission. *Icarus* 135: 127–145.
- NASA. 2021. Science 2020–2024: A Vision for Scientific Excellence. NASA Sci. Strategy. https://science.nasa.gov/science-red/s3fs-public/atoms/files/2020-2024_Science.pdf.
- Nayar, H. 2021. Robotics for Ocean Worlds Exploration. In: *Handbook of Space Resources*. Springer.
- Nimmo, F. 2004. Stresses generated in cooling viscoelastic ice shells: Application to Europa. *Journal Geophysical Research Planets* 109.
- Nimmo, F., B. Giese, and R.T. Pappalardo. 2003a. Estimates of Europa's ice shell thickness from elastically-supported topography. *Geophysical Research Letter* 30, n/a-n/a. <https://doi.org/10.1029/2002GL016660>.
- Nimmo, F., D.P. Hamilton, W.B. McKinnon, P.M. Schenk, R.P. Binzel, C.J. Bierson, R.A. Beyer, J.M. Moore, S.A. Stern, and H.A. Weaver. 2016. Reorientation of Sputnik Planitia implies a subsurface ocean on Pluto. *Nature* 540: 94–96.
- Nimmo, F., and R.T. Pappalardo. 2004. Furrow flexure and ancient heat flux on Ganymede. *Geophysical Research Letter* 31 <https://doi.org/10.1029/2004GL020763>.
- Nimmo, F., R.T. Pappalardo, and B. Giese. 2003b. On the origins of band topography, Europa. *Icarus* 166: 21–32. <https://doi.org/10.1016/j.icarus.2003.08.002>.
- Nimmo, F., and J.R. Spencer. 2015. Powering Triton's recent geological activity by obliquity tides: Implications for Pluto geology. *Icarus* 246: 2–10.
- Palmowski, J., B. Mellerowicz, K. Zacny, K.P. Hand, C. Sotin, S.M. Howell, T.A. Cwik, J. Mueller, B.L. Ehlmann, S. Nagihara, M. Tipton, and S. Liller. 2019. SLUSH: Search for life using submersible heated drill. Presented at the *AGU Fall Meeting 2019*, AGU.
- Pappalardo, R.T., M.J. Belton, H.H. Breneman, M.H. Carr, C.R. Chapman, G.C. Collins, T. Denk, S. Fagents, P.E. Geissler, and B. Giese. 1999. Does Europa have a subsurface ocean? Evaluation of the geological evidence. *Journal Geophysical Research Planets* 104: 24015–24055.

- Pappalardo, R.T., J.W. Head, R. Greeley, R.J. Sullivan, C. Pilcher, G. Schubert, W.B. Moore, M.H. Carr, J.M. Moore, M.J.S. Belton, and D.L. Goldsby. 1998. Geological evidence for solid-state convection in Europa's ice shell. *Nature* 391: 365–368. <https://doi.org/10.1038/34862>.
- Patterson, G.W., G.C. Collins, J.W. Head, R.T. Pappalardo, L.M. Prockter, B.K. Lucchitta, and J.P. Kay. 2010. Global geological mapping of Ganymede. *Icarus* 207: 845–867. <https://doi.org/10.1016/j.icarus.2009.11.035>.
- Patterson, G.W., J.W. Head, and R.T. Pappalardo. 2006. Plate motion on Europa and nonrigid behavior of the icy lithosphere: The Castalia Macula region. *Journal Structure Geology. Faulting and Fault-Related Processes on Planetary Surfaces* 28: 2237–2258. <https://doi.org/10.1016/j.jsg.2006.03.032>.
- Peddinti, D.A., and A.K. McNamara. 2019. Dynamical investigation of a thickening ice-shell: Implications for the icy moon Europa. *Icarus*. <https://doi.org/10.1016/j.icarus.2019.03.037>.
- Pierazzo, E., and G. Collins. 2004. A brief introduction to Hydrocode modeling of impact cratering. In *Cratering in Marine Environments and on Ice*.
- Planetary Satellite Physical Parameters [WWW Document], 2021. URL https://ssd.jpl.nasa.gov/?sat_phys_par. Accessed 16 May 2021.
- Postberg, F., R.N. Clark, C.J. Hansen., A.J. Coates, C.D. Ore, F. Scipioni, M.M. Hedman, and J.H. Waite. 2018a. Plume and surface composition of Enceladus. *Enceladus and the Icy Moons of Saturn*, 129.
- Postberg, F., N. Khawaja, B. Abel, G. Choblet, C.R. Glein, M.S. Gudipati, B.L. Henderson, H.-W. Hsu, S. Kempf, and F. Klenner. 2018b. Macromolecular organic compounds from the depths of Enceladus. *Nature* 558: 564–568.
- Prockter, L.M., F. Nimmo, R.T. Pappalardo. 2005. A shear heating origin for ridges on Triton. *Geophysical Research Letter* 32. <https://doi.org/10.1029/2005GL022832>.
- Ross, M.N., and G. Schubert. 1987. Tidal heating in an internal ocean model of Europa. *Nature* 325: 133–134.
- Ruesch, O., T. Platz, P. Schenk, L.A. McFadden, J.C. Castillo-Rogez, L.C. Quick, S. Byrne, F. Preusker, D.P. O'Brien, N. Schmedemann, D.A. Williams, J.-Y. Li, M.T. Bland, H. Hiesinger, T. Kneissl, A. Neesemann, M. Schaefer, J.H. Pasckert, B.E. Schmidt, D.L. Buczkowski, M.V. Sykes, A. Nathues, T. Roatsch, M. Hoffmann, C.A. Raymond, and C.T. Russell. 2016. Cryovolcanism on Ceres. *Science* 353. <https://doi.org/10.1126/science.aaf4286>.
- Ruesch, O., A. Genova, W. Neumann, L.C. Quick, J.C. Castillo-Rogez, C.A. Raymond, .. and M.T. Zuber. 2019. Slurry extrusion on Ceres from a convective mud-bearing mantle. *Nature Geoscience* 12(7), 505–509.
- Schenk, P., and M.P.A. Jackson. 1993. Diapirism on Triton: A record of crustal layering and instability. *Geology* 21: 299–302. [https://doi.org/10.1130/0091-7613\(1993\)021%3c0299:DOTARO%3e2.3.CO;2](https://doi.org/10.1130/0091-7613(1993)021%3c0299:DOTARO%3e2.3.CO;2).
- Schmidt, B.E., D.D. Blankenship, G.W. Patterson, and P.M. Schenk. 2011. Active formation of 'chaos terrain' over shallow subsurface water on Europa. *Nature* 479: 502–505. <https://doi.org/10.1038/nature10608>.
- Scully, J.E.C., P.M. Schenk, J.C. Castillo-Rogez, D.L. Buczkowski, D.A. Williams, J.H. Pasckert, K.D. Duarte, V.N. Romero, L.C. Quick, M.M. Sori, M.E. Landis, C.A. Raymond, A. Neesemann, B.E. Schmidt, H.G. Sizemore, and C.T. Russell. 2020. The varied sources of faculae-forming brines in Ceres' Occator crater emplaced via hydrothermal brine effusion. *Nature Communications* 11: 3680. <https://doi.org/10.1038/s41467-020-15973-8>.
- Sherwood, B., J. Lunine, C. Sotin, T. Cwik, and F. Naderi. 2018. Program options to explore Ocean Worlds. *Acta Astronautica* 143: 285–296. <https://doi.org/10.1016/j.actaastro.2017.11.047>.
- Skjetne, H.L., K.N. Singer, B.M. Hynek, K.I. Knight, P.M. Schenk, C.B. Olkin, O.L. White, T. Bertrand, K.D. Runyon, W.B. McKinnon, J.M. Moore, S.A. Stern, H.A. Weaver, L.A. Young, and K. Ennico. 2021. Morphological comparison of blocks in chaos terrains on Pluto, Europa, and Mars. *Icarus, Special Issue, Pluto System, Kuiper Belt, and Kuiper Belt Objects* 356: 113866. <https://doi.org/10.1016/j.icarus.2020.113866>.

- Soderlund, K.M. 2019. Ocean dynamics of outer solar system satellites. *Geophysical Research Letters* 46: 8700–8710. <https://doi.org/10.1029/2018GL081880>.
- Sotin, C., J.W. Head III., and G. Tobie. 2002. Europa: Tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting. *Geophysical Research Letters* 29: 74–81.
- Space Studies Board. 2018. *Astrobiology Science Strategy for the Search for Life in the Universe*. The National Academies Press.
- Space Studies Board, National Research Council. 2012. *Vision and voyages for planetary science in the decade 2013–2022*. National Academies Press.
- Sparks, W.B., K.P. Hand, M.A. McGrath, E. Bergeron, M. Cracraft, and S.E. Deustua. 2016. Probing for evidence of plumes on Europa with HST/STIS. *The Astrophysical Journal* 829: 121. <https://doi.org/10.3847/0004-637X/829/2/121>.
- Spencer, J., and D. Grinspoon. 2007. Inside Enceladus. *Nature* 445: 376–377. <https://doi.org/10.1038/445376b>.
- Spencer, J.R., A.C. Barr, L.W. Esposito, P. Helfenstein, A.P. Ingersoll, R. Jaumann, C.P. McKay, F. Nimmo, and J.H. Waite. 2009. Enceladus: An active cryovolcanic satellite. In: *Saturn from Cassini-Huygens*. Springer, pp. 683–724.
- Spitale, J.N., and C.C. Porco. 2007. Association of the jets of Enceladus with the warmest regions on its south-polar fractures. *Nature* 449: 695–697.
- Squyres, S.W. 1980. Surface temperatures and retention of H₂O frost on Ganymede and Callisto. *Icarus* 44: 502–510. [https://doi.org/10.1016/0019-1035\(80\)90040-8](https://doi.org/10.1016/0019-1035(80)90040-8).
- Steinbrügge, G., J.R.C. Voigt, N.S. Wolfenbarger, C.W. Hamilton, K.M. Soderlund, D.A. Young, D.D. Blankenship, S.D. Vance, and D.M. Schroeder. 2020. Brine migration and impact-induced Cryovolcanism on Europa. *Geophysical Research Letter* 47, e2020GL090797. <https://doi.org/10.1029/2020GL090797>.
- Thomson, R.E., and J.R. Delaney. 2001. Evidence for a weakly stratified European ocean sustained by seafloor heat flux. *Journal Geophysical Research Planets* 106: 12355–12365. <https://doi.org/10.1029/2000JE001332>.
- Trowbridge, A.J., H.J. Melosh, J.K. Steckloff, and A.M. Freed. 2016. Vigorous convection as the explanation for Pluto’s polygonal terrain. *Nature* 534: 79–81. <https://doi.org/10.1038/nature18016>.
- Vance, S.D., L.M. Barge, S.S. Cardoso, and J.H. Cartwright. 2019. Self-assembling ice membranes on Europa: Brinicle properties, field examples, and possible energetic systems in icy Ocean Worlds. *Astrobiology* 19: 685–695.
- Vance, S.D., K.P. Hand, and R.T. Pappalardo. 2016. Geophysical controls of chemical disequilibria in Europa. *Geophysical Research Letters* 43: 4871–4879. <https://doi.org/10.1002/2016GL068547>.
- Vance, S.D., and M. Melwani Daswani. 2020. Serpentinite and the search for life beyond Earth. *Philosophical Transaction Royal Society Mathematical Physical Engineering Science* 378, 20180421. <https://doi.org/10.1098/rsta.2018.0421>.
- Vance, S.D., M.P. Panning, S. Stähler, F. Cammarano, B.G. Bills, G. Tobie, S. Kamata, S. Kedar, C. Sotin, W.T. Pike, R. Lorenz, H.-H. Huang, J.M. Jackson, and B. Banerdt. 2018. Geophysical investigations of habitability in ice-covered ocean worlds. *Journal Geophysical Research Planets* 123: 180–205. <https://doi.org/10.1002/2017JE005341>.
- Vilella, K., G. Choblet, W.-E. Tsao, and F. Deschamps. 2020. Tidally heated convection and the occurrence of melting in icy satellites: Application to Europa. *Journal Geophysical Research Planets* 125, e2019JE006248. <https://doi.org/10.1029/2019JE006248>.
- Wahr, J., Z.A. Selvans, M.E. Mullen, A.C. Barr, G.C. Collins, M.M. Selvans, and R.T. Pappalardo. 2009. Modeling stresses on satellites due to nonsynchronous rotation and orbital eccentricity using gravitational potential theory. *Icarus* 200: 188–206. <https://doi.org/10.1016/j.icarus.2008.11.002>.
- Waite, J.H., M.R. Combi, W.-H. Ip, T.E. Cravens, R.L. McNutt, W. Kasprzak, R. Yelle, J. Luhmann, H. Niemann, D. Gell, B. Magee, G. Fletcher, J. Lunine, and W.-L. Tseng. 2006. Cassini ion and

- neutral mass spectrometer: Enceladus plume composition and structure. *Science* 311: 1419–1422. <https://doi.org/10.1126/science.1121290>.
- Woerner, D., S. Johnson, J.-P. Fleurial, S.M. Howell, B. Bairstow, and M. Smith. 2020. Radioisotope heat sources and power systems for enabling ocean worlds subsurface and ocean access missions. *National Academy Planet. Science Astrobiology Decadal Survey, 2023–2032*.
- Wünnemann, K., G.S. Collins, and H.J. Melosh. 2006. A strain-based porosity model for use in hydrocode simulations of impacts and implications for transient crater growth in porous targets. *Icarus* 180: 514–527.
- Yang, Chen, Haishi Zhao, Lorenzo Bruzzone, Jon Atli Benediktsson, Yanchun Liang, Bin Liu, Xingguo Zeng, Renchu Guan, Chunlai Li, and Ziyuan Ouyang. 2020. Lunar impact crater identification and age estimation with Chang'E data by deep and transfer learning. *Nature Communications* 11 (1): 1–15.
- Zahnle, K., P. Schenk, H. Levison, and L. Dones. 2003. Cratering rates in the outer Solar System. *Icarus* 163: 263–289. [https://doi.org/10.1016/S0019-1035\(03\)00048-4](https://doi.org/10.1016/S0019-1035(03)00048-4).
- Zandanel, A., L. Truche, R. Hellmann, A. Myagkiy, G. Choblet, and G. Tobie. 2021. Short lifespans of serpentinization in the rocky core of Enceladus: Implications for hydrogen production. *Icarus* 364: 114461. <https://doi.org/10.1016/j.icarus.2021.114461>.

Chapter 27

Robotic Mobility and Sampling Systems for Ocean-World Bodies



**Hari Nayar, Junggon Kim, Gareth Meirion-Griffith,
Brendan Chamberlain-Simon, Kalind Carpenter, Anna Boettcher,
Michael Hans, Brian Wilcox, Jason Carlton, and Justin Jenkins**

Abstract Ocean-world bodies like Europa and Enceladus are likely destinations for future space missions because they have conditions that could promote production of complex molecules and potentially life. Surface exploration of these bodies will require new types of robotics systems because their environments are unlike any we have encountered before. We have investigated several technologies for mobility,

H. Nayar · J. Kim · A. Boettcher · M. Hans
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS
198-219, Pasadena, CA 91109, USA
e-mail: hari.d.nayar@jpl.nasa.gov

J. Kim
e-mail: junggon.kim@jpl.nasa.gov

A. Boettcher
e-mail: anna.boettcher@jpl.nasa.gov

M. Hans
e-mail: michael.a.hans.jr@jpl.nasa.gov

G. Meirion-Griffith
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena,
CA 91109, USA
e-mail: gareth.meirion-griffith@jpl.nasa.gov

B. Chamberlain-Simon (✉)
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS 82-112,
Pasadena, CA 91109, USA
e-mail: Brendan.K.Chamberlain-Simon@jpl.nasa.gov

K. Carpenter
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS
107-102, Pasadena, CA 91109, USA
e-mail: kalind.c.carpenter@jpl.nasa.gov

B. Wilcox
Marine Bioenergy, Inc, 4408 Union Avenue, La Canada, CA 91011, USA
e-mail: Brian.Wilcox@marinebiomass.com

J. Carlton
3D Design Concepts, Inc, 1035 E. Vista Way #245, Vista, CA 92084-4606, USA
e-mail: jcarlton@3ddesignconcepts.com

This is a U.S. government work and not under copyright protection in the U.S.; foreign
copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_27

manipulation and sampling for these missions. For surface mobility, robotics systems will have to overcome terrain that has a wide range of potential terra-mechanical and topographical properties. We developed an optimized design through analysis, simulation and experimental demonstration to perform well in these conditions. For manipulation and sampling, it will be desirable to be able to collect samples over a wide area as far from the lander as possible. Alternative concepts were developed for reaching and collecting samples a very long distance from a lander, far beyond the capabilities of current manipulator arms. The alternative designs include (1) a foldable arm with truss components with a cable-and-pulley concept for transporting samples from the sample collector to the lander, and (2) a tethered projectile launched from the lander. Sampling material on the surface of these bodies will be a challenge because of the very cold temperatures. A three-pronged heated claw gripper was designed to address this challenge. These concept technologies have been demonstrated in simulation and in experiments. They will be capable of long-range mobility and sampling over varied terrain for exploration of ocean-world bodies.

27.1 Introduction

Several bodies among the outer planets in our solar system are attractive destinations for future space missions because of discoveries of liquid water below their icy crusts (Porco et al. 2006; Kivelson et al. 2000). Europa, a satellite of Jupiter, and Enceladus, a satellite of Saturn, are prominent among these (Franck et al. 2000) because they have (1) large subsurface oceans, (2) rocky cores in contact with the oceans that would provide salts and other chemicals needed for life, (3) energy fed into the oceans either from hydrothermal sources in the mantle or tidal activity, and (4) stable environments that could promote production of complex molecules and potentially life.

While astrobiology would be the primary science goal for missions to Europa and Enceladus, understanding the origins and geology of these bodies is also of interest (Sykes 2002). Spacecraft imagery of Europa, for example, has shown a varied surface with distinctive features that indicate an active ice tectonic history (Kattenhorn and Hurford 2009). A better understanding of the processes that led to Europa's current surface conditions will give insight into its formation and point to further questions that need to be investigated.

There is little information available on the surface geometry and material properties at the scale of robotics systems for ocean worlds bodies (Moore et al. 2009). The surface topography is expected to be rugged. Material on the ground may range from hard ice to loose regolith. On Europa, the surface material may be corrosive due to

J. Jenkins

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS 92-112, Pasadena, CA 91109, USA

e-mail: justin.m.jenkins@jpl.nasa.gov

sulfuric acid and other sulfur salts, and it is exposed to radiation so it is desirable to collect samples at least 20 cm below the surface.

27.2 Design Considerations

Among the ocean worlds within our solar system (including Earth, Europa, Ganymede, Callisto, Titan, Enceladus, Mimas, Tethys, Iapetus, Rhea, Triton, Ceres, and Pluto), Europa and Enceladus are two moons currently garnering significant interest from the scientific community and are the focus of this section. The surface topographic, mechanical, radiation and thermal environments of these moons pose a significant challenge to their robotic exploration. Surface temperatures range from approximately 76 K to 130 K on Europa (Spencer et al. 1999), and 65 K to 140 K on Enceladus (Brown et al. 2006). While similar in many respects (Patthoff et al. 2013), the significant differences between the two are: (1) the surface of Europa is exposed to high levels of radiation while there is none on Enceladus, and (2) Europa is significantly larger with a surface gravity of 1.3 m/s^2 (slightly less than Earth's moon) than Enceladus with a surface gravity of 0.11 m/s^2 (Paranicas et al. 2009). Perhaps the greatest challenges to multi-site surface operations are the topographic and mechanical properties of surfaces of ocean worlds. The majority of ocean worlds are covered by icy crusts, kilometers in depth. These crusts are composed largely of pure water ice but also contain salts and potentially organics, deposited on the surface by endogenous processes. Images taken by the Cassini and Galileo missions show distinct surface morphologies in the form of double ridges, bands, chaos, and lenticulae, each of which pose challenges to the concept of surface mobility. Plume vents detected on Enceladus by Cassini (Porco et al. 2006) and more recently on Europa by Hubble (Sparks et al. 2016), show that significant deposition of subsurface material exists, likely yielding a somewhat localized, fine-grained cryogenic ice regolith, that likely holds properties emplacing unique design requirements for rover designers.

A geology science mission to Europa or Enceladus with the ability to perform in-situ science investigations over varied terrain types would be of great value. There are a number of options for achieving this capability. The requirements derived for a surface mobility geology mission are to be able to (1) cover a distance of at least 10 km over the large range of terrain types to be found on ocean worlds, (2) handle the range of material types from hard ice rock to non-cohesive granular ice, and (3) operate in the environmental conditions of cryogenic temperatures, and vacuum and meet mission constraints of power, mass, and volume. The 10 km drive goal was derived from a study of images of Europa. Figure 27.1, for example, shows a 35 km by 50 km patch of varied terrain, where it was determined that with an appropriate landing site it would be possible to visit at least three different terrain types within a 10 km travel distance. Among the design choices, a vehicle that drives over the



Fig. 27.1 Mosaic of Conamara Chaos region on Europa (Image credit: NASA/JPL/U Arizona)

terrain is an attractive solution. The details of the design and implementation of this concept are described in the remainder of this section.

The reach goal for the design of a manipulation system on ocean-world bodies is driven by the expected pin-point precision capabilities of vision-aided inertial navigation for targeting landing sites (Johnson et al. 2016) of about 20 m in the 2030 or later time-frame for future missions. Assuming the landing system is able to place the lander within 20 m of a science target of interest, a manipulation system with a reach of 20 m will enable access to that target. For Europa, samples would have to be collected at least 20 cm below the surface due to radiation processing of materials at the surface (Pathhoff et al. 2013).

In addition to satisfying environmental constraints, the concepts developed should meet the requirements on available power, mass and stowed volume. Assuming a radioisotope thermoelectric generator (RTG) power source as the baseline, the available power would be in the range of 100 W. Assuming the availability of the space launch system (SLS) (Donahue et al. 2016), the design limit on mass for a manipulation system was capped at 100 kg, dimension for the manipulator system was capped at 3 m, and a maximum volume of 6 m³. Additional considerations for any surface system on Europa and Enceladus are the need to meet planetary protection requirements, perform operations autonomously and maintain the integrity of samples between sampling and analysis.

27.2.1 Surface Environment

The successful design of a robotic platform for ocean worlds requires a priori knowledge of the surface topography and mechanical properties. From a mobility perspective, topography yields an understanding of the surface roughness as well as the range of obstacle sizes and gradients the vehicle will likely encounter. To date, surface imagery from the Galileo and Cassini missions has yielded images with pixel sizes only on the order of ~6 m and ~20 m, respectively, which while globally informative, do not provide insight into the surface characteristics at the vehicle scale. Any proposed vehicle design is required to be robust to a range of surface conditions. As part of this task, several pathological cases have been identified as providing the most stringent requirements on robotic mobility: cryogenic ice, penitentes, regolith, salt evaporites, and chaos terrain, as illustrated in Fig. 27.2.

Cryogenic Ice. Cryogenic ice (Fig. 27.2a) forms the major component of the surfaces of ocean worlds and with the exception of salt evaporites, provides the basis from which other surface morphologies are formed. Unlike Earth-borne ice, cryogenic ice is exceptionally hard, with a compressive strength of 110–120 MPa in laboratory tests (Wu and Prakash 2015).

Penitentes: Penitentes (Fig. 27.2b) are near-vertical ice “spikes” that may be found in equatorial latitudes and are made possible due to the extremely low surface pressure, which is far below the triple point of water. Near-perpendicular incident solar energy at these latitudes can readily result in the ablation of surface material, yielding 1–5 m spikes (Hobley et al. 2013), similar to those found in regions of the Andes Mountain on Earth. Mobility in a penitente-strewn environment is difficult to conceive and represents an extreme topographic roughness that may preclude mobility in certain equatorial latitudes.

Cryogenic Ice Regolith: Multiple remote sensing measurements have been taken that are informative of the characteristics of the top layer of ice and existence of an ice regolith (Fig. 27.2c). Photometry, performed in the visible spectrum of 380–770 nm, can penetrate into the top tens of microns of a surface and has suggested particle sizes in the range of 20 to several hundreds of microns (Moore et al. 2009). Regolith compaction varies between the leading and trailing sides, yielding void ratios of 25% on the leading side and 79% on the trailing side (Buratti et al. 1988). In addition, polarization phase curves of very fine-grained particles closely match the photometric

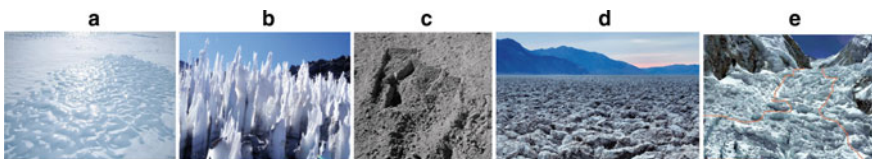


Fig. 27.2 From left to right: **a** cryogenic ice, **b** penitentes, **c** regolith, **d** salt evaporites, **e** chaos

observations and suggest that the uppermost layer of the surface is likely formed of <1.5-micron particles with void spaces of up to 90% (Moore et al. 2009). Results from RADAR (Nelson et al. 2015), show that ocean worlds, and both Europa and Enceladus in particular, are bright to backscatter at a wide range of wavelengths. Although the interpretation of this is still under debate, a brightness to backscatter typically implies that particles or void spaces of equivalent size to the illuminating source exist. A very high backscatter to a wide range of wavelengths would thus imply the likelihood of a very well graded regolith on the order of meters deep.

Salt Evaporites: Salt evaporites (Fig. 27.2d) are non-ice species on the surfaces of ocean worlds and are visible in orbital imagery as brown, orange, or yellow surface discolorations. These non-ices are deposited salts, believed to be associated with the upwelling of subsurface ocean materials. As the liquid phase of the upwelled material sublimates, the salt components remain behind. Salt evaporites, found in dry lake basins of Earth (e.g., the Devil’s Golf Course in Death Valley), are formed by very similar processes. The crystalline, evaporite structures range from very brittle and weak to strong enough to support the heaviest of vehicles. Primarily, these structures offer a topographic challenge, potentially offering trap conditions.

Chaos: Chaos terrains (Fig. 27.2e), a pathological case describing the heavily disrupted ice fields that dominate regions such as Europa’s Thera Macula, are formed due to tectonic activity and driven by tidal swelling. Chaos represents one of the most challenging terrains for surface mobility due to its extreme topography. Given the aforementioned measurements of a very fine-grained, porous surface, it is likely that chaos terrains found on any ocean world will have a layer (of indeterminate thickness) of regolith. In this case, the highly rugged surface combined with loose surface material is very likely to pose challenges to mobility and navigation, and the risk of entrapment.

One of this task’s principal goals has been to evaluate the performance of various mobility architectures. Through the use of in-house modeling and simulation tools such as M3TK (Mukherjee et al. 2014), it is possible to evaluate the ability of a vehicle to traverse regions of ocean worlds as a function of their topography. In order to understand the role of the regolith that is believed to cover the surfaces of these bodies, it is necessary to understand the geotechnical properties of a cryogenic ice regolith under vacuum. To the first order, the properties of interest are the shear strength according to the Mohr–Coulomb failure criterion, and its bearing strength under normal load. By studying these properties, it is possible to estimate the thrust (a function of shear strength) and sinkage (a function of bearing capacity) that a vehicle may achieve on a granular medium. To this end, the authors have investigated the properties of a granular, cryogenic ice regolith inside a cryogenic vacuum chamber, shown in Fig. 27.3. Both direct shear and bearing capacity tests were performed inside the chamber, which is capable of maintaining a regolith temperature of 100 K, inside a near-vacuum of ~10⁻⁵ Torr. Tests were performed on a range of both particle sizes (10 μm to 2 mm) and shapes. To simulate the ice grains formed during the fracturing of tectonic processes, angular grains were generated by mechanical fracturing and sieving directly into a liquid nitrogen bath. To replicate the spherical, amorphous

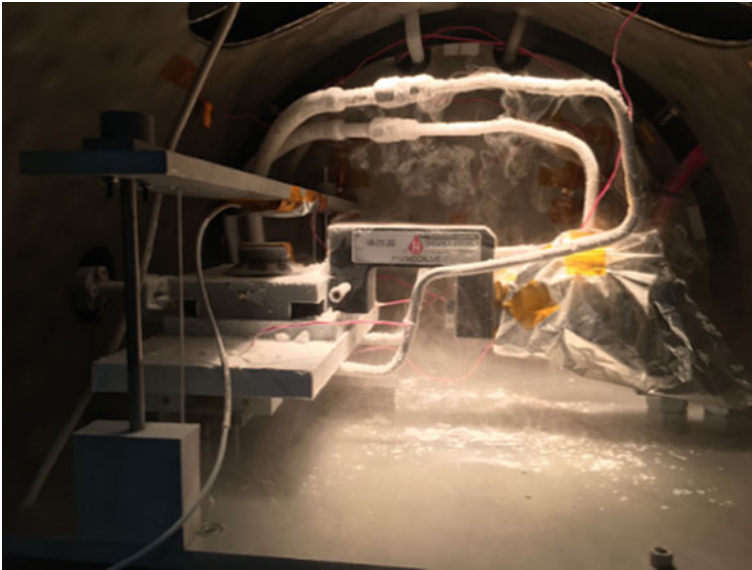


Fig. 27.3 Direct shear tests performed in a cryogenic vacuum chamber

particles likely found on the flanks of the Enceladus plume, liquid water droplets were dispersed from an atomizing nozzle directly into a liquid nitrogen bath. Note that it is important to either sieve or disperse samples directly into liquid N_2 , as at no time during sample preparation should the material be exposed to ambient conditions, which would allow for vapor deposition and sintering, leading to inter-particle cohesions that would not be present on the surface of ocean worlds.

27.3 Vehicle Design

Little is known about the surface of Europa and Enceladus at the scale of a robotic vehicle that would operate on their surfaces. Therefore,, given the knowledge we currently have, a vehicle capable of driving on these bodies should be designed to handle a wide range of terrain types in terms of both topography and material properties. There was a time in the early 1960s (De Fries 1967) when a similar problem was posed with regard to driving on our Moon. A conservative solution developed was the GM rover (De Fries 1967) shown in Fig. 27.4. The recommendation made at that time (De Fries 1967) was to design a vehicle with large compliant wheels for the non-cohesive soils expected on the Moon and a suspension system capable of handling the rough topography. Since then, NASA has gained significant practical

experience with mobility on planetary bodies from lunar rovers and the Mars rocker-bogie mobility systems (Bickler 1988). Figure 27.5 shows the first of the Mars rovers, Sojourner.

The two challenging aspects of mobility on the surface of ocean worlds are rugged terrain and potentially soft regolith (Moore et al. 2009). The solutions chosen to address these challenges are: (1) a large range-of-motion rocker-bogie suspension system to comply with a wide range of topography, (2) large-diameter deformable wheels relative to the vehicle size to float on soft regolith, and (3) a low center of mass for stability. A hybrid concept that combines the advantages of the GM rover and the Mars rocker-bogie rovers was chosen for analysis, optimization and prototyping for experimental evaluation. The concept has large compliant wheels to drive over non-cohesive terrain as was implemented on the GM rover. It also has a rocker-bogie suspension system to enable driving over rugged terrain with a relatively simple mechanical suspension. A planar kinematic model with three wheels representing



Fig. 27.4 GM rover from 1960s (De Fries 1967)

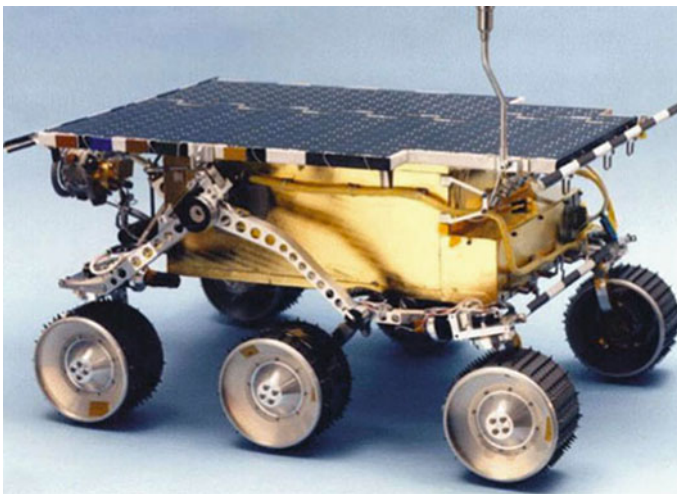


Fig. 27.5 Pre-flight image of Mars Pathfinder rover, Sojourner (Image credit: JPL)

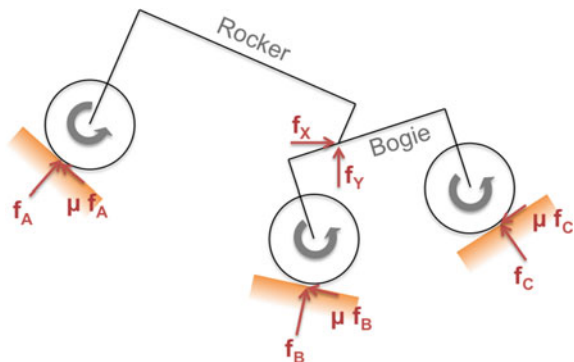
one side of this vehicle chassis was analyzed in Matlab. The analysis modeled the quasi-static forces at representative configurations along the trajectory for climbing over large (twice wheel diameter) obstacles and also for avoiding trap conditions where an obstacle between wheels prevents further forward or reverse driving. An optimization was performed to find the best kinematic parameters for the vehicle for obstacle climbing and trap-free driving.

27.3.1 Vehicle Design Optimization

We predict traverse performance of a rocker-bogie rover using 2D kinematics and static force analysis, find an optimal rover design to maximize the performance, and validate our design method using full 3D dynamic simulation results. Figure 27.6 illustrates a force diagram of a rocker-bogie rover at an arbitrary state on an uneven terrain where (f_x, f_y) is the reaction force at the free rotating hinge, $f_A, f_B,$ and f_C are the normal contact forces at the wheels, and μ is the Coulomb friction coefficient between the terrain and wheels. The direction of the friction forces is determined by the wheel rotation direction, depicted as the gray arrows. Assuming a slow vehicle speed, the six unknown values (the forces and the friction coefficient) are determined from the force and moment equilibrium equations for the two rigid bodies, i.e., rocker and bogie. Note that the rover design parameters such as the vehicle mass and hinge location (see Fig. 27.7 for all parameters), the current rover configuration (e.g., the hinge joint angle) conformed to the terrain geometry, and the wheel rotation direction are all incorporated into the nonlinear force/moment equations and affect the solution.

In particular, we focus on the solution for the friction coefficient, or critical friction coefficient, because it can be used as a metric of the vehicle's performance. In order to achieve the current quasi-static vehicle state, the effective friction between the terrain and the wheels must be at least equal to or greater than the critical value. Otherwise, the vehicle would not be able to move in a desired direction. For a given terrain geometry, one can calculate the critical friction coefficients along a trajectory

Fig. 27.6 Force diagram of a rocker-bogie rover located on an uneven terrain



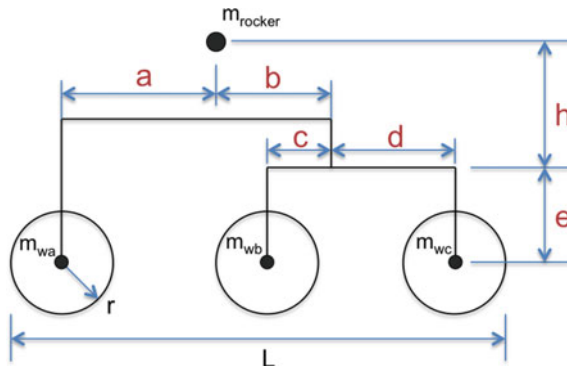


Fig. 27.7 Design parameters of the rocker-bogie rover

across the terrain and find the maximum value. This represents the minimum required friction coefficient for the vehicle to traverse and indicates how well the vehicle can traverse terrains with such geometry. A lower critical value results in a better traverse performance. Note that the critical friction coefficient is a purely geometric value and depends only on the rover design and the geometry of the terrain being considered, so we can decouple our rover design optimization problem from the unknown material properties of the terrain surface.

Three types of terrain geometries shown in Fig. 27.8 have been considered as representatives of the surface of ocean worlds in our analysis. The first and second terrains (Plateau and Valley) are to test climbing performance of the rover. Their height is set to four times the wheel radius. The third one (Trap) is to test a trap-free driving capability of the rover; we seek to avoid an instance in which the front or middle wheel has easily passed over the obstacle but the ensuing wheel cannot overcome it. This instance becomes a high failure-potential case if the rover cannot back away from the obstacle it traversed in its forward direction.

Figure 27.9 shows the critical friction coefficients of an arbitrarily designed rover at six key states while traversing the terrain with the obstacle (Trap). Assume the rover is moving in bogie-facing direction and the actual frictional coefficient between the terrain and the wheels is 0.6, as an example. The front wheel can pass over the obstacle because the friction is higher than required (0.45), but the middle wheel cannot because it needs a higher friction (0.82) to overcome the same obstacle. Furthermore, the rover cannot back off because the friction coefficient value is too low (<0.77) for the front wheel to retreat over the obstacle. Thus, the rover is likely

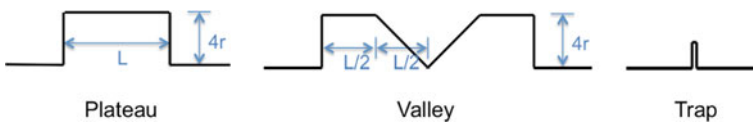


Fig. 27.8 Types of terrain geometries considered in our analysis

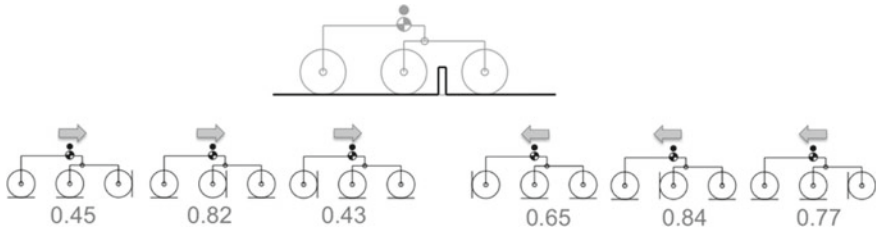


Fig. 27.9 An arbitrarily designed rover that can be trapped by an obstacle

to be trapped when it traverses rough terrains with a frictional coefficient of between 0.45 and 0.77 in bogie-facing driving. Such a trapped situation does not occur in rocker-facing driving because the front wheel can back off when the middle wheel cannot pass over the obstacle.

We use an optimization technique to determine the rover design parameters shown in Fig. 27.7. The design goal, or the cost function to be minimized in our optimization, is the maximum critical frictional coefficient, or the minimally required friction coefficient for the vehicle to traverse across a terrain. In order to calculate the cost of the current rover design, we sample multiple rover states in both driving directions by solving inverse kinematics of the rover at different locations on the terrain (see Fig. 27.10), calculate the critical frictional coefficient value at each state by solving the nonlinear force/moment equilibrium equations, and finally take the maximum value. Some of the design parameters, such as wheel radius (11 cm), mass (including the drive motor), and the load mass (for control box and batteries) attached onto the rocker frame, and the total vehicle length (set to 100 cm in the analysis), were given from the hardware specifications we chose. The masses of the chassis frames were lumped together to the load mass and the wheel mass to simplify the problem in our analysis. In order to avoid the aforementioned trap situation, a set of inequality constraints on the critical friction coefficients of the six rover states was added to the problem set-up for Trap terrain.

We used the Matlab optimization toolbox to solve the problem numerically. The gradient-based optimization algorithm finds a local optimum rather than a globally optimal design. Due to the discrete rover state sampling on the terrain and an iterative method for solving the nonlinear force/momentum equations for the critical friction coefficients, the cost function is not only highly nonlinear but also non-smooth, which causes poor convergence in the numerical optimization. Therefore, for each terrain type, we ran the design optimization twice with different initial designs, and took the better result. Figure 27.11 shows the two initial rover designs and the optimized designs we found for the three terrain types.

Figure 27.12 compares traverse performance of the five rover designs predicted from the critical friction coefficients of the rovers. The top figure presents climb performance of the vehicles predicted on Plateau terrain, and the bottom one is the trap-free driving capability of them on Trap terrain. The horizontal axis indicates the rover designs being compared. There are two bars for each design; the left is

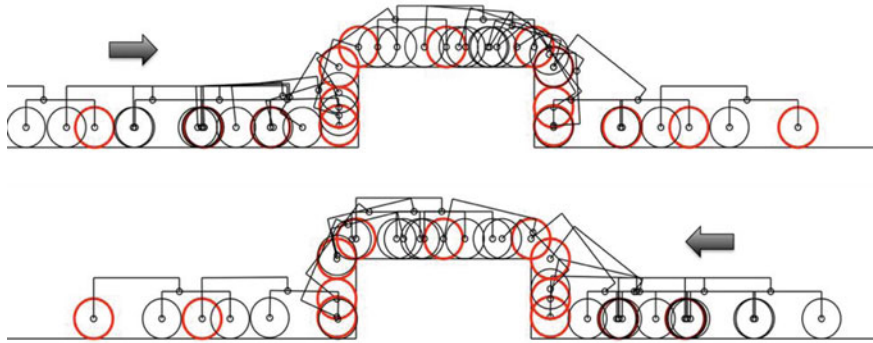


Fig. 27.10 An example of sampled rover states on Plateau terrain (The front wheel is colored red for visual clarity.)

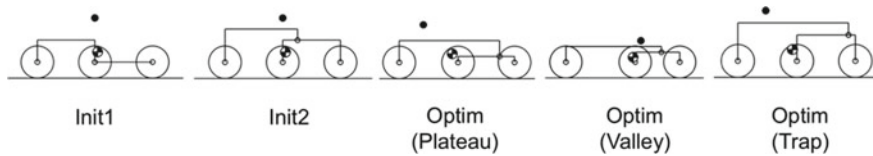


Fig. 27.11 Two initial designs and the optimized designs for the three terrain types

for bogie-facing driving and the right one is for rocker-facing driving. The vertical axis indicates the frictional coefficient of the contacts between the wheels and the tested terrains. The green color in the bar graphs represents the range of frictional coefficient in order for the rovers to traverse the terrain successfully. The friction range with no color (the area below the bars) indicates that the friction is too low to climb up the cliff of Plateau terrain or for the front wheel to pass over the obstacle in Trap terrain. In the case of Trap terrain, the blue color indicates that the rover can back off when the middle or rear wheel cannot overcome the obstacle. The red color means the rover is trapped when the obstacle is between the wheels, and cannot escape from the obstacle in both directions.

The two initial designs showed the worst performance in both climbing and trap-free driving, and in both driving directions. The third rover, marked as ‘Optim (Plateau)’, showed the best climb performance (Fig. 27.12, top). This is natural because the design has been optimized for that terrain (Plateau). The fifth rover, which has been optimized for Trap terrain, also showed a very good climb performance in bogie-facing driving mode, although its rocker-facing driving is not as adept as the third rover. This rover also showed the best trap-free driving capability on Trap terrain (Fig. 27.12, bottom). We predict that, in bogie-facing driving, the fifth rover will have almost as good a climb capability as the third rover, with the smallest chance of being trapped among the five designs being compared. In rocker-facing mode, the rover can safely drive on rough terrain without being trapped, although its

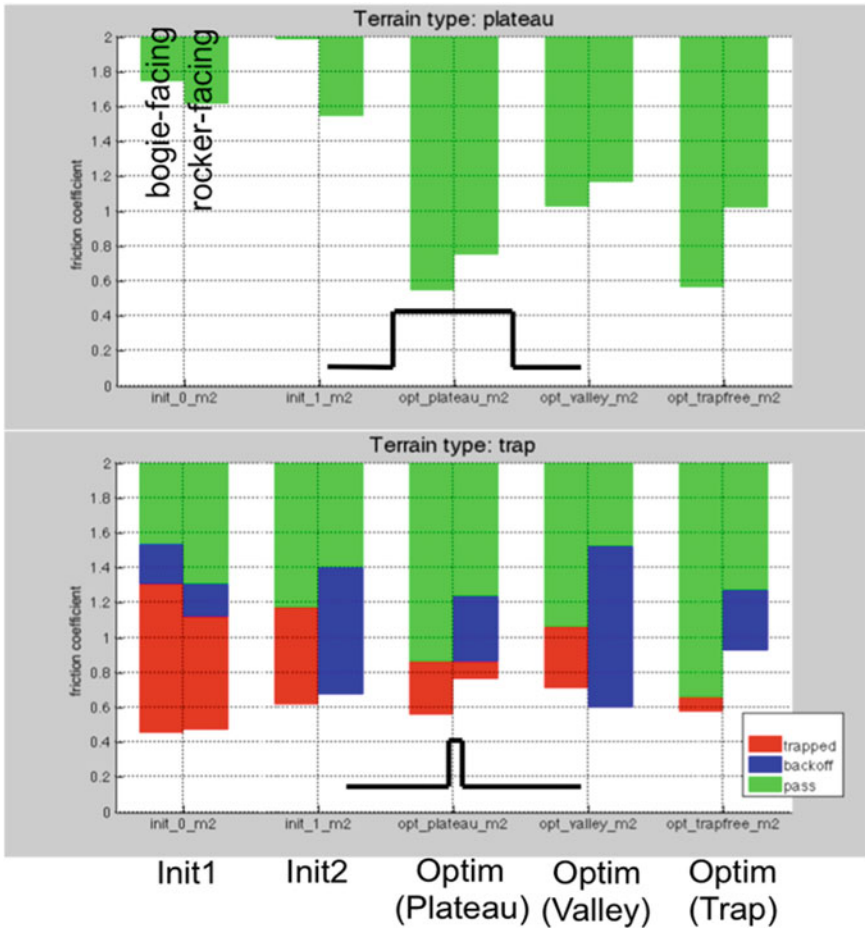


Fig. 27.12 Traverse performance of the initial and optimal rover designs predicted on Plateau (top) and Trap (bottom) terrains

climb performance is less capable than the third rover. Therefore, we chose the fifth rover design as the model for our initial hardware fabrication.

27.3.2 Vehicle Simulation Analysis

Following the optimization of the rocker-bogie configuration, 3D simulations were performed in M3tk (Mukherjee et al. 2014) to validate the 2D model and to expose issues which may not be accounted for in the quasi-static optimization. The 3D simulation was parameterized in the same way as the optimization model, including

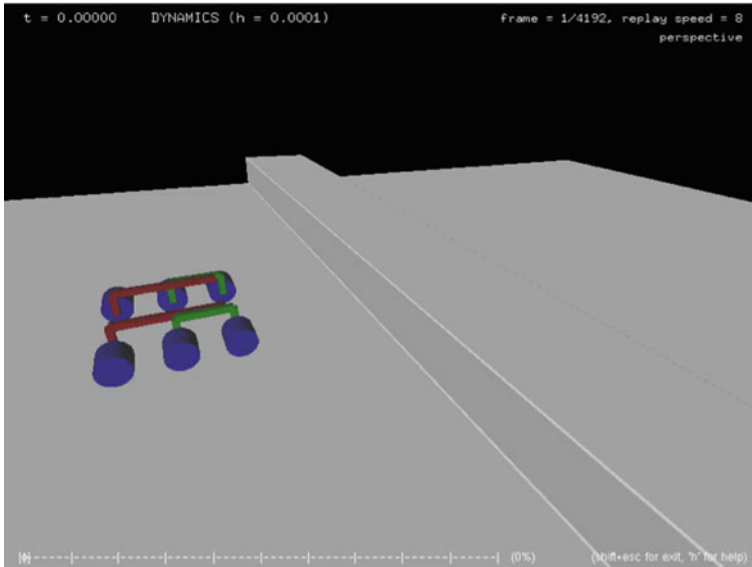


Fig. 27.13 A rover model on a terrain in our full 3D dynamics simulation

link lengths and mass properties, so that the found optimum could be directly ported to the 3D model. The three-dimensional terrain also exactly matched the 2-dimensional model, where primitive shapes such as cylinders and rectangular prisms were used to create the considered terrain types (an example is shown in Fig. 27.13). The wheels of the rover were also represented using primitive shapes for collisions with the terrain, where each wheel consisted of two spheres, offset from each other by a representative wheel width of 0.2 m. The Kelvin–Voigt contact model (Meyers and Chawla 2008) was used for wheel-ground normal force calculations, while a linear Coulomb friction model was used for finding tangential forces. The wheels of the rover were kept at a constant angular velocity of 0.5 radians per second.

3D simulations began with a parameter study to find a stable simulation set-up. This included ground contact parameters such as stiffness and damping, control gain values for the simulated controllers keeping the wheels at a constant speed, and the desired velocity of the rover. When the simulation output was found to be stable, work began on verifying the optimization output on the trap-free, Plateau, and Valley terrain types. Each rover model was simulated driving over their respective obstacle multiple times, with ground contact friction varying in each simulation from 0.05 to 27.3 in increments of 0.05. Using this procedure, the minimum required friction coefficient for traversal of the obstacle can be found for that rocker-bogie configuration, enabling a direct comparison with the minimum friction coefficient found in the optimization. This test set-up also allowed for verification of the trap-free condition, where only one or two wheels capable of traversing the obstacle would leave the rover trapped.

The 3D simulations closely matched the results of 2D analysis. Again, the two initial designs turned out to have very poor traverse performance. We also found that the optimized design for Trap terrain, named ‘Optim (Trap),’ still showed the best trap-free driving performance in 3D simulation, and there is a tradeoff between trap-free driving and climbing capabilities, as implied by the Matlab analysis. In the simulation we did not implement a control scenario for backing off the rover when the middle or rear wheel could not pass over the obstacle on Trap terrain, and instead, marked as ‘blocked’ the cases depicted in magenta color in the bar graphs. Therefore, the magenta region can actually correspond to either the ‘back-off’ or ‘trapped’ situation in Fig. 27.9. Although there is some minor inconsistency between the two bar graphs in Figs. 27.12 and 27.14, their overall qualitative trends are very similar, which supports our analysis and optimization methods.

The 3D dynamics simulations also uncovered potential issues not seen in the static analysis. One such issue is that the rover can have a tendency to flip onto its back when traversing a tall obstacle. This would occur on approach to the obstacle, when the front two sets of wheels were on top of the obstacle, and the back wheels were still on the ground. Due to the height of the obstacle, the center of mass of the rocker-bogie would move behind the last set of wheels, causing the vehicle to become unstable and flip over. This issue was eventually corrected with new and more accurate mass properties that showed the wheels to be much more massive than originally estimated. This moved the center of mass lower on the rover, keeping the vehicle stable even at extreme angles. The new mass properties were also used in our analysis and design optimization described in the previous section. A second issue that was discovered was that the vehicle had a tendency to do a “wheelie” when driving in rocker-facing mode. The set of wheels attached to the rocker would come into contact with the obstacle first. Instead of these front wheels beginning to climb the obstacle, sometimes they would slip, holding the vehicle back. The middle wheel set would lift off the ground as the rear set of wheels continued to drive forward. This effectively made it so that only the front wheels were contributing force toward overcoming the obstacle, where the middle wheels were not on the ground, and the rear wheels were only rotating the rocker-bogie joint. This was solved in simulation through a simple traction control algorithm which would stop driving the rear wheels when the middle wheels left the ground. This allowed the middle wheels to contribute force to the vehicle to start the process of ascending the obstacle. This solution was later extended to physical hardware, where the ratio of the velocities of the bogie wheels can be set to prevent the wheelie from occurring.

27.3.3 Vehicle Hardware Implementation

A detailed vehicle design with six fully steered and driven wheels reflecting the optimized configuration from analysis and simulation studies was prototyped. The prototype was designed to be reconfigurable in order to experimentally determine performance of a range of vehicle kinematic configurations. The approach taken was

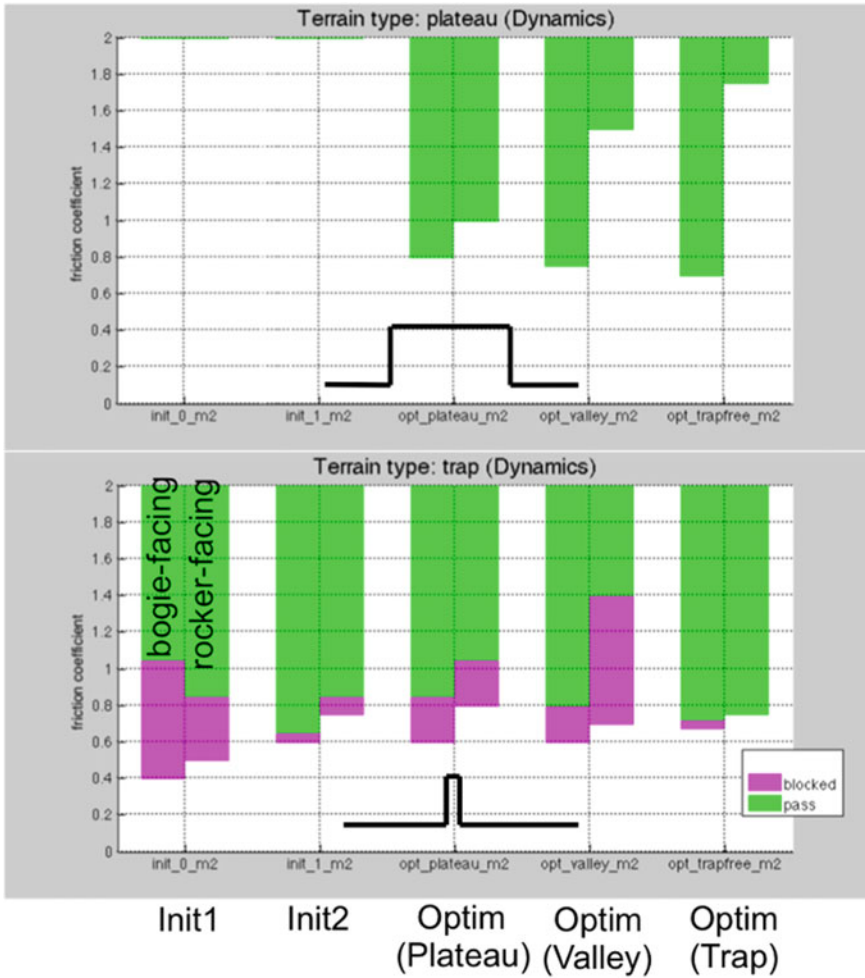


Fig. 27.14 Traverse performance of the initial and optimal rover designs predicted from a full 3D dynamics simulation

to design several modular elements connected by tubular links that could be varied in length. One of the primary modular elements in the design is the wheel-steer assembly. All six wheel-steer assemblies on the prototype are identical. In order to minimize cost, a commercial off-the-shelf compliant wheel was chosen. A custom-drive motor and gear train sub-assembly were designed to fit within the wheel hub. The drive wheel sub-assembly was mounted to the steer sub-assembly to form the steer-drive unit as shown in Fig. 27.15. Two rocker joints and two bogie joints were also designed to be mounted between tubular elements on the two sides of the chassis. Each of these joints has an adjustable hard stop, allowing for a parametric evaluation

Fig. 27.15 Wheel-steer assembly



of the effect of joint limits on the mobility of the vehicle. The rocker-bogie passive suspension mechanism that constrains the rocker motion between the left- and right-side rocker joints was implemented with a mechanical lever pivoted in the middle of the tube connecting the left and right sides of the vehicle. The bogie joints were free to rotate constrained by wheel contact on the ground. Figure 27.16 shows the assembled vehicle.

The drive electronics consists of an Arduino micro-controller driving the 12 motors through 12 H-bridge motor controller boards. Optical encoders sense the wheel-drive velocities and absolute magnetic encoders are used to measure the wheel-steer angles. The motor controller also implements motor current sensing. Two lithium-ion polymer (LiPo) batteries, mounted within the electronics box, are used to power the system. For testing, a radio-controlled joystick is used to send commands for forward, sideways and turn velocities to the Arduino. The 2-D velocity vector and angular velocity commands from the joystick are translated into wheel-steer angles and drive velocities, based on the Ackerman steering algorithm, that are then used as a set point for a 2-DOF PID controller for the steering motor controllers. The drive motors are controlled in an open loop.

27.4 Sampling System Design

We have developed two options for deployment of a sampling system that allow for a ten-meter or more reach. They are (1) a deployable boom composed of truss segments with a rail for transporting a sampling system, and (2) a projectile launcher



Fig. 27.16 Prototype rover being exercised at JPL test yard

to lob a sampling system to a desired target and retrieve it. In addition, a unique sampling system designed specifically for ocean-world bodies has been developed.

27.4.1 Component Technologies

Robotics systems operating on the surface of ocean worlds depend on the availability of a host of component technologies that survive extreme conditions. An investigation was performed on the state of component technologies that would be needed for surface robotics systems on Europa and Enceladus.

For structural and housing elements and fixtures, steel and titanium will work in cryogenic conditions (DellaCorte 2010). High-torque gearboxes are also needed for cold-temperature applications. A development based on bulk metallic glass (BMG) materials is underway at JPL (Roberts 2014). Some composite materials exhibit better performance at lower temperatures. Shear strength increases with decreasing temperature in composite materials. However, micro-cracks due to thermal expansion mismatch between the matrix and the fiber degrade performance. The performance of composites for specific applications will have to be experimentally evaluated.

Heated actuators are currently used in space applications and are a fallback solution for ocean-world missions. Commercial off-the-shelf cold-temperature actuators are available for short-duration missions. A NASA-funded research effort (Mojaraddi et al. 2011) to develop a brushless DC cryogenic motor and associated control electronics achieved 120 million cycles at temperatures of 110 K.

Wet lubricants will require heating for ocean-world applications. Dry film lubricants are available for low-temperature applications. Molybdenum disulfide is the typical dry lubricant used for space applications. Dicronite (Tungsten disulfide) has also been used for space applications and is rated for the nominal thermal conditions at the surface of Europa and Enceladus. However, new BMG materials may obviate the need for lubricants.

Low-temperature electronics technology is being developed under NASA funding for ocean worlds applications. These are expected to become available in the time-frame for extended surface missions that this section targets. The fallback and conventional solution is to use heaters in a warm electronics box to maintain temperatures within the operating range for the respective components.

A short-duration mission would be feasible with heated electrochemical battery technology. New battery technology is being developed under NASA funding for extending the life and temperature range for batteries in order to reduce the power needed for heaters. Long-duration missions would have to use very large solar arrays or RTGs. Use of RTGs would place increased constraints on planetary protection for surface missions on ocean worlds.

Perception and sensing systems are an integral element of robots. For ocean-world applications, protection casings would have to be developed for cameras for the cryogenic and vacuum conditions. For visible-spectrum sensors, the ambient lighting from the Sun at the outer planets will be significantly less than on Earth and Mars (Lee et al. 2016). Artificial lighting may be needed for robotics surface operations.

Radio-frequency communication back to Earth from ocean worlds will require large, power-hungry antennas. Optical communication could be a mature technology in the time-frame considered with relatively high bandwidth and low power.

27.4.2 Deployable Boom Sampling Concept

One approach to place a sampling system 10 m or more from a lander is to deploy a lightweight boom. To demonstrate this concept, a prototype 1/3-scale arm composed of four truss segments that unfold from a stowed configuration on a lander was developed. Three trusses deployed in the forward direction extend the boom for positioning the sampling system. A fourth truss is deployed in the backward direction to counter the change in the center of mass of the combined lander-boom system and maintain it within the footprint of the lander. Control elements would be housed on this truss to provide the counter-weight. When mounted on a turntable in its deployed configuration, the boom can slew in the horizontal plane over its workspace, sweeping an arc and enabling access to a sampling system on the arm to any location on the ground within that arc. In the design, a pulley-based positioning system is used to translate a cart carrying a sampling system along the length of the arm and drop and retrieve the device to the surface for sampling operations. The prototype in its

un-stowing sequence is illustrated in Fig. 27.17. The fully deployed truss boom with the payload at its tip is shown in Fig. 27.18.

During the design evaluation and optimization phase of this concept, bending performance of a boom arm under gravitational conditions on Europa was analyzed.

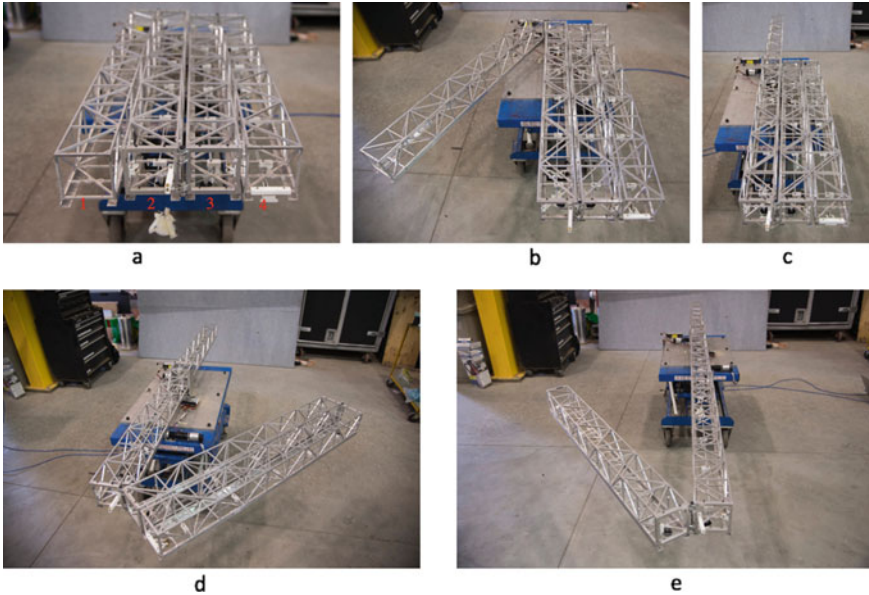


Fig. 27.17 Computer-controlled truss boom deployment sequence (a–e). Top left panel shows truss numbering (1–4)

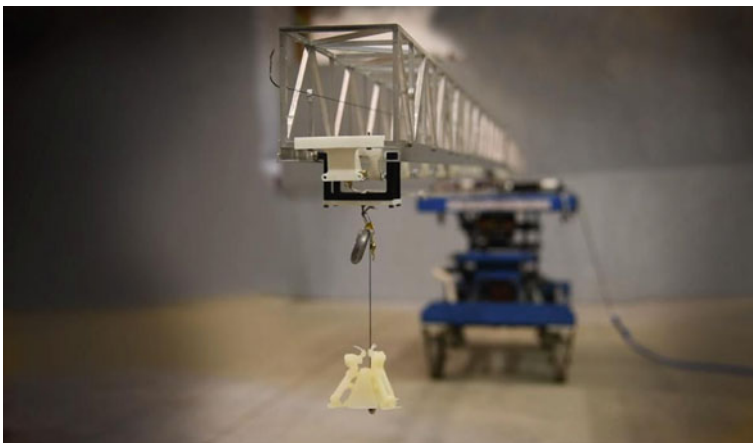


Fig. 27.18 Prototype of a fully deployed truss boom with payload at tip

The nominal geometry for the design was a 10 m-long arm with a 0.15 m square cross-sectional dimension and a 1/8-inch wall thickness. Trusses made of steel, aluminum and composite materials were evaluated. It was found that an aluminum arm with a mass of 7 kg with an 8 kg sampling payload at its tip deflected only 0.03 m at its tip. A carbon-fiber composite material had three times better bending performance. Arm deflection on Enceladus would be much better than on Europa because of the significantly weaker gravitational field.

In the scale-model prototype developed, each truss is 1 m long and was machined from 4-inch square tubular aluminum stock with a 3/16-inch-thick wall. The four trusses are connected serially by three hinges on diagonally opposite edges so they can fold to lie beside each other and unfold to form a long boom. The trusses are numbered on the left-most panel of Fig. 27.17. In the demonstration of this concept, a base-grounded truss (Truss 2) is mounted on a pedestal. A backward-rotating truss (Truss 1) is hinged to the back on one side of Truss 2. Two connected forward-rotating trusses (Truss 3 is hinged to the front of Truss 2 and Truss 4 hinged to Truss 3) are mounted on the other side.

Two independent sets of cables, routed internally through pulleys within the trusses, are used for the forward and backward deployment. The cables are pulled onto two spools that are driven independently by two brushless DC motors. A latching mechanism is used to lock the trusses into place upon completion of deployment. In the intended application scenario, the boom will be deployed once, locked in place and never retracted. The sequence for forward deployment of Truss 3 and 4 is controlled by setting the relative friction in the hinges between trusses. The motors are turned off after deployment. The high gear ratio transmissions on the outputs of the motors enable tension to be maintained in the cables even in the unpowered state to help keep the boom stiff. In the design, a sampling system is suspended from a cart below the boom. The cart travels along the length of the boom on a rail attached to the bottom of the trusses. It is pulled along the rail by cables routed over pulleys and driven by a third brushless DC motor. The cart can be positioned at any point along the rail to lower the sampling system to the ground below.

Two RoboClaw motor controllers (Ion Motion Control 2020) interfaced to a laptop computer through a USB port were used to control the motors. Power for the set-up was supplied by a lithium-ion polymer (LiPo) battery, making it a very portable system. A graphics user interface (GUI) and Python software on the laptop were used to control the motors manually by toggling switches on the GUI. For the prototype, the deployment motors were set up to allow retraction of the boom in order to demonstrate deployment repeatedly.

27.4.3 Projectile Launcher Concept

An alternative concept for long-range sampling from a lander is to launch a sampling system as a tethered projectile to its target sampling site and retrieve it after the sampling operation. A number of options are available for launching a projectile

sampler. These include (1) an Onager catapult, (2) a motor driving a linear transmission, (3) compressed gas driving a piston, and (4) a compressed spring. Analyses of each of these options were performed to determine the feasible distance a projectile could be launched within the constraints of a notional lander mission to Europa. All the designs easily exceeded the minimum desired launch distance of 10 m. For practical and safety reasons the approach using a spring compressed by a brushless DC motor driving a ball screw was finally chosen for prototyping.

In the prototype design shown in Fig. 27.19 (an engineering drawing on the left panel and the prototype on the right), the motor is used to translate a plate (P1) through a ball-screw transmission. P1 is mounted on two linear bearings that slide along shafts parallel to the ball screw to allow smooth translation along the length of the travel of the ball screw. Two electro-magnets (e-magnets), with the magnets on when not powered, are mounted on P1 and used to attach and detach from another plate (P2) that compresses two linear springs. Each of these linear helically wound springs compress and release along the length of a second set of two shafts. As P2 moves along its pair of shafts, it compresses and releases the springs. The projectile is mounted on a mandrel on the outward side of P2 and is released when P2 hits the stop at the end of its travel.

The controller set-up for this device used components similar to the boom deployment controller. A RoboClaw motor controller connected to a laptop through a USB port was used to control the two motors and read the position of the ball screw through the encoder feedback from the motor. In addition, a USB relay controller was used for controlling the e-magnets and turning the motor brake on and off. A 24 V power supply was used for power. A GUI and software written in Python were used to operate the system. The software begins by initializing communication with

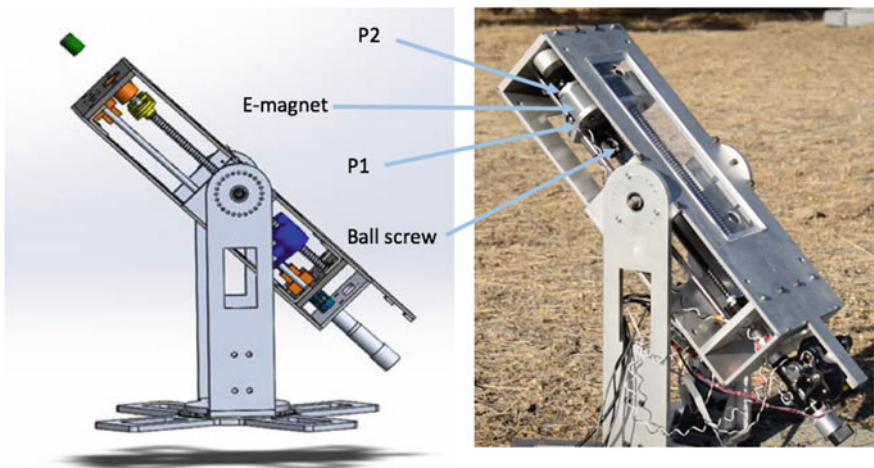


Fig. 27.19 Projectile launcher sampling system prototype

the devices, effectively shaking hands. It then splits the work load among four software threads running simultaneously. The first thread sends and receives data from the RoboClaw, the second thread sends and receives data from the relay controller, the third thread updates and monitors the GUI for button presses, and the fourth thread coordinates the actions of the other threads to produce a more complicated or coordinated response. The software then monitors the current and duty cycle of both motors, and the position and velocity of the brushed motor while awaiting a button press. When a button is pressed, a function is called which tells the necessary threads what to do. It is important to not share control of a device between threads as multiple commands cannot be given or received simultaneously.

The sequence of operations to launch a projectile are: (1) drive P1 out until the e-magnets automatically attach to corresponding steel disks on P2, (2) P1, with P2 attached to it, is then driven in to compress the springs, (3) at the chosen spring compression distance, the motor brake is turned on then the e-magnets are momentarily powered to turn off the magnetic field and release P2, (4) the expanding springs accelerate P2 outward until it hits the mechanical stop at the end of the fully extended length of the springs, (5) due to the momentum it has gained, the projectile continues to travel out of the launcher pulling a tether that is simultaneously released from a spool.

In this concept, the sampling system (described in the next section) falls to the ground, rights itself and performs the sampling procedure to collect samples that are cached on the sampler. The sampler is then retrieved back into the launcher. The procedure to reel in the projectile is to drive a motor attached to the spool to pull in the tether. The tether is wound back onto the spool and the projectile re-seats itself on the mandrel in the launcher and is ready to be launched again. The targeting of specific sampling locations is accomplished by tilting the launcher angle, by rotating the launcher on a turntable and controlling the compression of the springs.

27.4.4 Ice Gripper Sampling System Concept

A concept of a sampling system to be deployed from a lander to operate on ocean worlds was designed and prototyped. This sampling concept, shown in Fig. 27.20, is compatible with either the deployable boom or the projectile launcher concepts described in the previous two sub-sections. The concept scenario is to place the sampler while in a stowed tetrahedral configuration to the desired sampling site either by being lowered to the ground from the boom or by being launched from the projectile launcher. It self-rights itself as it deploys its three petals from a triangular base-plate to lie on the ground (see Fig. 27.21). The three petals are actuated simultaneously by one DC motor rotating a spool of three cables. Each cable, routed around a pulley at the base of each petal, pulls against a constant-force spring to rotate each petal about a hinge at its base.

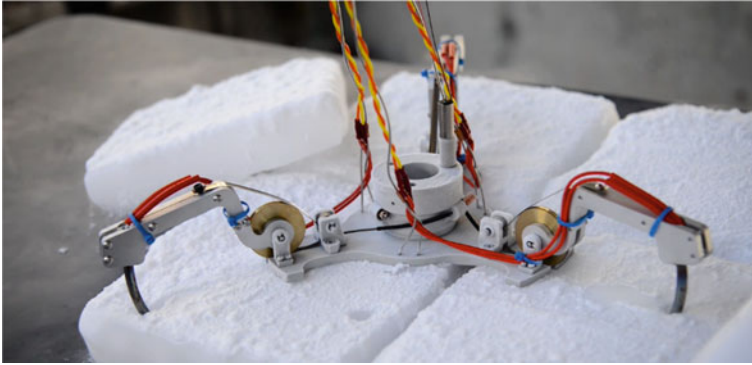


Fig. 27.20 Ice gripper prototype testing in dry ice

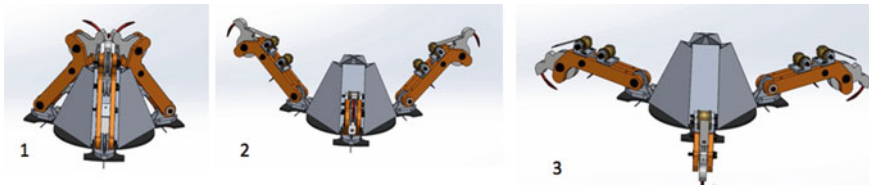


Fig. 27.21 Ice dripper deployment sequence

For the case where the ground is hard ice, the sampling system would have the capability to anchor itself in order to apply the necessary weight-on-bit to drill into the surface. This is done using a heated, curved prong on each petal. As the prongs contact the ice in the cryogenic and vacuum conditions on the surface of an ocean-world body, the ice sublimates and the prong penetrates the ice along a curved path to grip the ice. At the end of its travel, the heaters are turned off and the prongs lock in place to grip the ice. If the surface material is loose regolith, the weight of the sampling system will provide adequate weight-on-bit for sampling. The heating power needed for anchoring is lower than on Earth because sublimation of ice in vacuum occurs at a lower temperature. Initial heat-exchange analyses followed by experiments in cryogenic temperatures in a vacuum chamber were conducted to verify that sublimation occurs within the constraints of the power available on a lander.

Having anchored on the surface, a drill with a sample capture sheath would be used to collect samples from the specified depths below the surface. The samples are cached on the sampling system. The process to un-anchor is accomplished by reversing the anchoring process. The prongs are heated then the petal drive-cables are released. The constant-force springs retract the petals to stow the sampling system back into its tetrahedral configuration. The tether is then pulled either from the boom or the projectile launcher to retrieve the sampler. In the current version of the prototype, actuation of the petals and heating the prongs are performed manually.

Implementing these capabilities under computer control should be trivial using a control system similar to the systems used for the boom and projectile launcher prototypes.

27.5 Summary and Conclusions

The prototype vehicle developed in this investigation was tested with physical mock-ups of the analytical terrains and on analog terrain to simulate the conditions that might be found on ocean-world bodies both for the expected topology and for soft regolith. Early results have shown that it is able to surmount obstacles more easily when in the bogie-facing configuration.

The prototype of the boom concept has met its design objectives and demonstrated that it is possible to implement significantly longer reach for planetary sampling operations from a lander. The projectile launch approach has been shown to be a feasible approach for lander-based sampling with the ability to reach much greater distances. However, this concept has a lower level of maturity. A robust design of a spooling system is needed to reel out the tether during launch and repeatedly re-spool during the projectile retrieval procedure. The novel ice gripper concept has also been shown to be feasible for sampling on ocean worlds missions.

As missions to ocean worlds become an increasing focus for NASA, technology to meet the science goals while satisfying mission and environmental constraints will be needed. The concept technologies we have developed and demonstrated in simulation and experiments will be capable of long-range mobility and sampling over varied terrain.

Acknowledgements Research reported in this chapter was conducted at the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, under a contract with National Aeronautics Space Administration (NASA).

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

References

- Bickler, Donald. 1998. Roving over mars. *Mechanical Engineering* 120 (04): 74–77.
- Brown, H. Robert, R.N. Clark, B.J. Buratti, D.P. Cruikshank, J.W. Barnes, R.M.E. Mastrapa and J. Bauer et al. 2006. Composition and physical properties of Enceladus' surface. *Science* 311(5766): 1425–1428.
- Buratti, Bonnie J., Robert M. Nelson, and Arthur L. Lane. 1988. Surficial textures of the Galilean satellites. *Nature* 333 (6169): 148–151.
- De Fries, P.J. 1967. Lunar surface locomotion. In *Advances in Space Science and Technology*, vol. 9, pp. 275–453. Elsevier.

- DellaCorte, C. 2010. *Nickel-Titanium Alloys: Corrosion “Proof” Alloys for Space Bearing, Components and Mechanism Applications*. National Aeronautics and Space Administration, Glenn Research Center.
- Donahue, B., D. Burks, and D. Cooper. 2016. Exploration opportunities enabled by the space launch system. In *2016 IEEE Aerospace Conference*, pp. 1–11. IEEE.
- Franck, S., A. Block, W. von Bloh, C. Bounama, H.-J. Schellnhuber, and Y. Svirezhev. 2000. Habitable zone for Earth-like planets in the solar system. *Planetary and Space Science* 48(11): 1099–1105.
- Hobley, D.E.J., J.M. Moore, and A.D. Howard. 2013. How rough is the surface of Europa at lander scale? In *Lunar and Planetary Science Conference*, no. 1719, p. 2432.
- Ion Motion Control, Website: <http://www.ionmc.com/>.
- Johnson, A.E., Y. Cheng, J. Montgomery, N. Trawny, B.E. Tweddle, and J. Zheng. 2016. Design and analysis of map relative localization for access to hazardous landing sites on Mars. In *AIAA Guidance, Navigation, and Control Conference*, p. 0379.
- Kattenhorn, S.A., and T. Hurford. 2009. Tectonics of Europa. In *Europa*, pp. 199–236. Tucson: University of Arizona Press.
- Kivelson, Margaret G., Krishan K. Khurana, Christopher T. Russell, Martin Volwerk, Raymond J. Walker, and Christophe Zimmer. 2000. Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289 (5483): 1340–1343.
- Lee, D., Y. Cheng, and H. Nayar. 2016. Indirect and direct planetary illumination modelling for robotic surface exploration sensing. In *AIAA SPACE 2016*, p. 5445.
- Meyers, M.A., and K.K. Chawla. 2008. *Mechanical behavior of materials*. Cambridge University Press.
- Mojarradi, M., P.B. Abel, T.R. Tyler, and G. Levanas. 2011. Electro-mechanical systems for extreme space environments. *NASA Langley Research Center*.
- Moore, J.M., G. Black, B. Buratti, C.B. Phillips, J. Spencer, and R. Sullivan. 2009. Surface properties, regolith, and landscape degradation. In *Europa*, pp. 329–349. Tucson: University of Arizona Press.
- Mukherjee, R., S. Myint, J. Chang, I. Kim, J. Craft, M. Pomerantz, J. Kim, and L. Peterson. 2014. M3Tk: A robot mobility and manipulation modeling toolkit. In *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection.
- Nelson, R.M., M.D. Boryta, B.W. Hapke, K.S. Manatt, A. Nebedum, D. Kroner, Y. Shkuratov, V. Psarev, and W.D. Smythe. 2015. Jupiter’s Satellite Europa: Evidence for an extremely fine-grained, high porosity surface. In *AAS/Division for Planetary Sciences Meeting Abstracts# 47*, pp. 405–02. 2015.
- Paranicas, C., J.F. Cooper, H.B. Garrett, R.E. Johnson, and S.J. Sturmer. 2009. Europa’s radiation environment and its effects on the surface. In *Europa*, pp. 529–544. Tucson: University of Arizona Press.
- Patthoff, D.A., R.T. Pappalardo, H. Chilton, and P.C. Thomas. 2013. Using limb profiles and stereo imagery for ridge comparisons on icy satellites. In *AGU Fall Meeting Abstracts*, vol. 2013, pp. P23E–1835.
- Porco, C.C., P. Helfenstein, P.C. Thomas, A.P. Ingersoll, J. Wisdom, R. West, and G. Neukum et al. 2006. Cassini observes the active south pole of Enceladus. *Science* 311(5766): 1393–1401.
- Roberts, S.N. 2014. *Developing and Characterizing Bulk Metallic Glasses for Extreme Applications*. PhD diss., California Institute of Technology.
- Sparks, W.B., K.P. Hand, M.A. McGrath, E. Bergeron, M. Cracraft, and S.E. Deustua. 2016. Probing for evidence of plumes on Europa with HST/STIS. *The Astrophysical Journal* 829(2): 121.

- Spencer, John R., Leslie K. Tamppari, Terry Z. Martin, and Larry D. Travis. 1999. Temperatures on Europa from Galileo photopolarimeter-radiometer: Nighttime thermal anomalies. *Science* 284 (5419): 1514–1516.
- Sykes, M.V. 2002. *The Future of Solar System Exploration (2003–2013)– Community Contributions to the NRC Solar System Exploration Decadal Survey*. Vol. 272.
- Wu, Xianqian, and Vikas Prakash. 2015. Dynamic compressive behavior of ice at cryogenic temperatures. *Cold Regions Science and Technology* 118: 1–13.

Chapter 28

Communication and Obstacles Detection Using Piezoelectric Transducers in a Penetrator Melting Deep Ice on Ocean Worlds



**Yoseph Bar-Cohen, Xiaoqi Bao, Hyeong Jae Lee, Benjamin Hockman,
Mircea Badescu, Stewart Sherrit, and Shyh-Shiuh Lih**

Abstract In the search for life, one of NASA's priorities is in-situ exploration of ocean worlds in the solar system where there might be life under the ice shell. This requires the ocean below the ice shell to be reached by traversing through great ice depths that are extremely cold. Europa, one of Jupiter's moons, is one such challenging planetary body; its ice shell is estimated to be 40 km deep. An approach has been conceived to reach the ocean using a Cryobot, which is a melting probe with a lander as the platform for its deployment. This ice-penetrating vehicle concept consists of a cylindrical, narrow-bodied probe that encases a radioisotope heat/power source to melt through the icy shell. The probe would include a suite of scientific instruments to analyze both the ice during descent and the ocean water

Y. Bar-Cohen (✉)

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
e-mail: yoseph.bar-cohen@jpl.nasa.gov

X. Bao

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
e-mail: baoxq@yahoo.com

H. J. Lee · S.-S. Lih

Jet Propulsion Laboratory, MS 67-119, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA
e-mail: hjlee@jpl.nasa.gov

B. Hockman

Surface Mobility Group Jet Propulsion Laboratory, Robotics Technologist, California Institute of Technology Mail Stop, 198-219, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA
e-mail: Benjamin.j.hockman@jpl.nasa.gov

M. Badescu

Payload and Small Spacecraft Mechanical Engineering Section, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS 67-119, Pasadena, CA 91109-8099, USA
e-mail: Mircea.Badescu@jpl.nasa.gov

S. Sherrit

Jet Propulsion Laboratory, 4800 Oak Grove Dr. MS 67-119, Pasadena, CA 91109, USA
e-mail: ssherrit@jpl.nasa.gov

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

V. Badescu et al. (eds.), *Handbook of Space Resources*,
https://doi.org/10.1007/978-3-030-97913-3_28

underneath. For communication, a set of fiber optic wire and wireless radio frequency (RF) in the very cold porous top layer is assumed, and then acoustic modules are used for communication in the warmer denser ice over a distance of 25 km between modules. In addition to the acoustic communication modules, a sonar is used to avoid obstacles and determine the distance from the ice–water interface region. The focus of this chapter is the use of elastic waves for wireless communication and sonar for operation through great ice depths.

28.1 Introduction

Cold ocean worlds in the solar system (e.g., Europa, Enceladus) have been identified as high-priority targets of the National Research Council's (NRC) Planetary Science Decadal Survey (National Research Council Space Studies Board 2012, 2018; Pappalardo et al. 2016). Research and development efforts to develop new robotic system capabilities have significantly increased, including ice descent probes to access the oceans of these icy moons (Zimmerman et al. 2001a, b; Lishman et al. 2013; Hand et al. 2017; Bar-Cohen and Zacny 2020). One such capability under consideration is a Cryobot that would penetrate and test samples of the thick ice shells on the way down to the ice–ocean interface. The Cryobot was designed to deliver autonomous undersea explorations payloads. Prior studies (Di Pippo et al. 1999; Zimmerman et al. 1998, 2001a, b) and more recent investigations (Cwik et al. 2018) have been aimed towards operational and technological aspects to accelerate the landing and deployment capabilities of the Cryobot. These efforts highlight the need to develop a comprehensive set of end-to-end mission architectures.

There are three major themes within ocean-world scientific exploration objectives: (1) geodynamics—determining what is the structure and dynamic state of the icy crust and ocean interface; (2) habitability—determination of the ocean world's past or present state which may provide the necessary environment to support life; and (3) life detection—to detect whether life has emerged on one of these ocean worlds, and whether it persists today. Many key questions related to these objectives are best answered through in-situ analysis of the ice-shell interior and ocean of these worlds. Therefore, the planetary science community is investigating various technical solutions that may become sufficiently mature to support exploration missions in the next 15–20 years. One such solution under consideration is a highly autonomous Cryobot vehicle, a long-range underwater explorer capable of rapidly penetrating while testing samples in the thick ice shells down to the ice–ocean interface. This concept assumes a cylindrical narrow-bodied probe containing a radioisotope heat/power source that is used to melt through Europa's icy crust, and a suite of scientific instruments for analyzing the ice samples during descent and the liquid ocean beneath. The concept of accessing subglacial oceans with a melt probe is not new, and many competing technologies have been explored. The concept study completed recently is called PRIME (Probe using Radioisotopes for Icy Moons Exploration) (Fleurial et al. 2019).

28.2 Cryobot Architecture and System Integration

A Cryobot is a complex system with various subsystems that have critical interdependencies (Zimmerman et al. 2001b). Thus, it is crucial that the development of a communication (comm) system considers the resources that are available from the Cryobot and the impacts that it may have on other subsystems. For example, power provided by the main probe radioisotope thermoelectric generator (RTG) and the thermal requirements to accommodate additional aft volume are two key system-level interplays that must be considered in development of a comm module that is compatible with a Cryobot.

The concept described here is based on the results of the recent JPL PRIME study, in which a notional Cryobot system architecture was designed to penetrate Europa's ice shell (Fig. 28.1). The comm module is in the aft section behind the rotating section (see Fig. 28.2) and is tapered in such a way as to minimize side-wall heating requirements (i.e., naturally conforming to the refreezing melt plume). Each deployable consists of ice anchors for bracing in the borehole, a mini RTG for power, spooled bays of optical fiber, electronics for driving the transceiver, and a low-frequency ring-type acoustic transducer. The first relay (R1) also has an RF patch antenna for transmitting to the surface. Each relay module may further be separated (as shown in the middle section of Fig. 28.2) so that any liquid melt pore induced by the waste heat of the mini RTG is sufficiently far from the frozen-in transducer.

28.3 Ice Sonar for Obstacle Detection

Sonars are sound navigation and ranging techniques that employ acoustic waves (Benjamin, 2008). They are widely used for obstacle detection in water and particularly underwater, as in submarine navigation. They are also used for communication under the surface of the water. There are two types of sonars:

- Passive—these sonars are used as listening devices for detecting and determine the source of sound that is emitted by vessels.

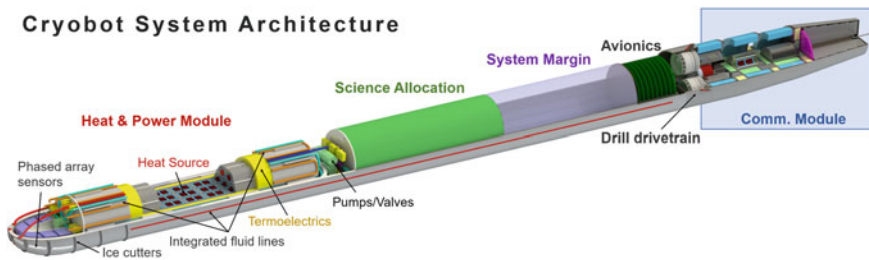


Fig. 28.1 Cryobot System Architecture

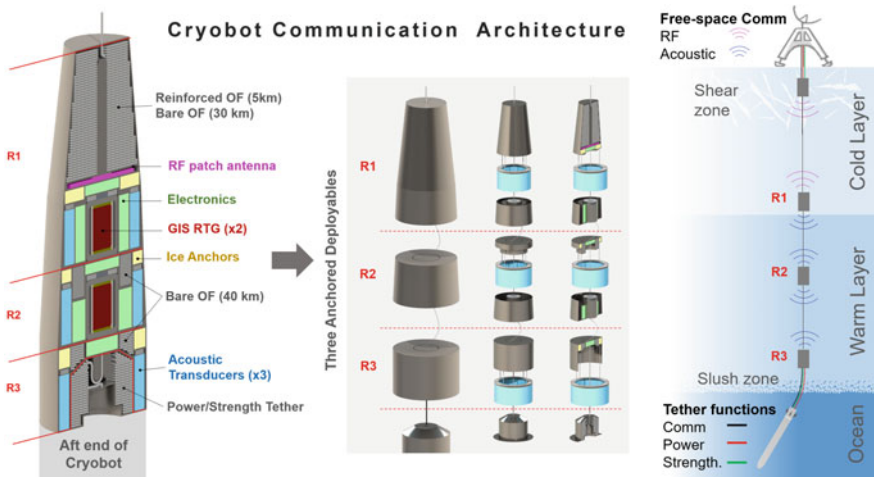


Fig. 28.2 Conceptual design of a Cryobot communication system

- Active—these are sonars that transmit acoustic pulses and search for reflections.

Sonar can be devices that are simply used to determine the location of sound sources as well as characterizing detected objects in water. The frequencies used in sonar systems range from as low as infrasonic to as high as ultrasound. The use of passive sonars started during World War I as a means of detecting submarines. Active sonars are used by vessels (e.g., ships and submarines) and they employ either a single transducer or an array that allows an image of the water media being insonified to be made. The low cost of sonars with imaging capability has led to commercial application for such devices as the fish finder (Fig. 28.3).

Limited studies using sonar have been conducted. However, its application in ice is increasing; the most important application for planetary usage is for the detection of obstacles in potential future Cryobot systems. A recent study by Kowalski et al. (2016) exploring the use of ultrasonic waves as a sonar mechanism was directed towards ice conditions on the Saturnian moon Enceladus. Their concept was based on a melting probe called the IceMole, which is able to travel along curved trajectories as well as upwards. Using its steering capability, it is able to avoid obstacles and selected targets.

In the PRIME study, a sonar was conceived to detect obstacles so that the steering mechanism could use information about their distance and size to avoid collisions. The sonar would be able to detect the ice/water interface at the end of the ice shell as well as determine the distance from it.

Fig. 28.3 Cabin display of a fish-finder sonar. Public domain photo provided from the U.S. *National Oceanic and Atmospheric Administration*. (<https://commons.wikimedia.org/wiki/File:Fishfinder.jpg>)

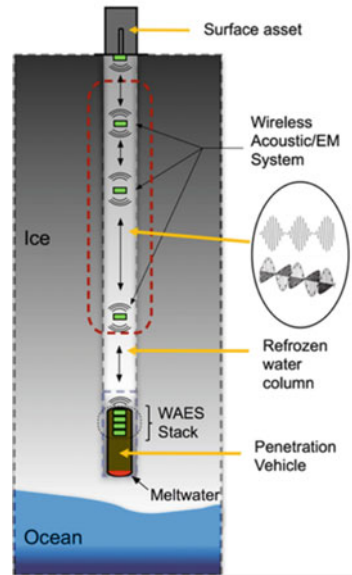


28.4 Acoustic Communication

A communication link is required from the probe to the surface through the ice crust and then to Earth. A proposed solution for the probe–surface communication section currently in development is a wireless communication link through the ice consisting of a series of transceiver modules. These modules include the use of radio-frequency (RF) electromagnetic communications; however, RF signals are highly attenuated in warm, salty or wet ice. In ice warmer than 260 K, ice acoustic communication is used. The communication capability (baud rate) of the acoustic transceiver relative to the available power and spacing has been studied. Jupiter’s moon Europa was chosen as the target body but the results are applicable to other icy planetary bodies. Our current understanding of the properties of the ice crust was used to guide the study.

The potential wireless communication system being considered for a descending ocean access probe is illustrated in Fig. 28.4. The proposed system includes multiple pucks that are released from the probe as it descends. The pucks consist of electromagnetic or acoustic transceivers or both to relay the communication signals between the probe and the surface asset (lander, rover, etc.). The electromagnetic transceivers are mainly used to cover the upper low-temperature zone in ice and the acoustic transceivers for the lower high-temperature zone. The units are powered by miniature RTGs with limited electric power.

Fig. 28.4 Proposed wireless communication system



28.4.1 Acoustic Transmission in Ice

Acoustic communication over long distances is widely used in marine systems (Sherman and Butler 2007). The major components of the acoustic modem and a general block diagram of acoustic communication are shown in Fig. 28.5. In essence, there is little difference in the process of communication between underwater and ice environments. The principle of acoustic communication operation is as follows: an onboard computer control system is responsible for converting the logical information (0 and 1) into analog signal, then amplifying and modulating it and sending it to the transmitter. The modulated signal is then converted to an acoustic pressure wave via a transduction mechanism, and propagates through the ice. When the acoustic pressure wave reaches the deployed acoustic modem, the pressure wave is converted back into the modulated analog signal, and is processed onboard in real time, converting it back into the original digital information.

On Earth, today’s acoustic modems typically achieve reliable acoustic communications beyond kilometers, or even hundreds of kilometers, at a low data rate in water. The commercial acoustic modem developed by LinkQuest (27–45 kHz), for example, offers a transmission range of 1.5 km with transmit power of 4 W and receive power of 0.8 W at a data rate of 9600 bps, while the modem by Teledyne Benthos (16–21 kHz) has a transmission range of 2–6 km with transmit power of 12 W and receive power of 0.4 W at a data rate of 2400 bps (Stojanovic and Preisig

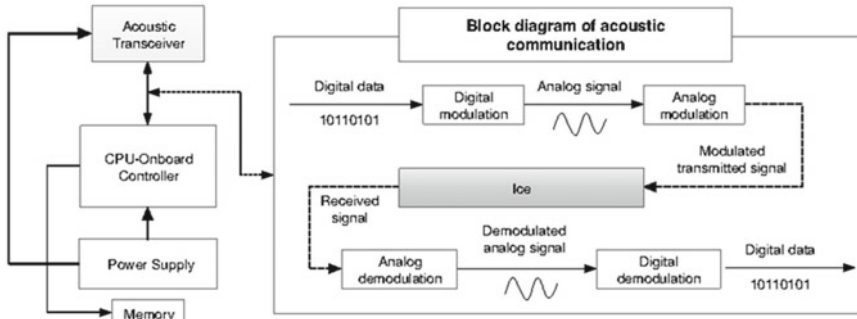


Fig. 28.5 Internal architecture (left) and the block diagram (right) of an acoustic modem system

2009; Benson et. al. 2010). Other commercial acoustic modems are providing in-water rates as high as 15,360 bps (e.g., <http://www.teledynemarine.com/920-series-atm-925?ProductLineID=8>).

Transmitting and receiving acoustic waves through ice in Europa, however, is a challenging task as the acoustic medium of ice is significantly different from water, possessing higher acoustic impedance and higher transmission loss. In addition, there are many unknown effects that impede acoustic wave propagation in ice, such as multipath propagation and volume inhomogeneities in ice (Lee et al. 2003, 2005). Note that the shear wave has a speed lower than the pressure wave. In addition, the shear wave is a wave mode that is not supported in fluids. Therefore, at a distance in ice, the transducer may receive two signals with different time delays or phases that may interfere with the communication. For these reasons, the accurate representation of such a complex and dynamic acoustic wave propagation model is very difficult, and many assumptions are required to estimate the acoustic communication capability.

As an analog for acoustic channel characteristics of Europa ice, we used South Pole data as a baseline. Note that Europa ice and South Pole ice may have significant differences due to the method of formation and the trapping of air in the South Pole ice; thus, we need to apply conservative margins to estimate the capability of acoustic communication technology in Europa. The important parameters for acoustic channel characteristics are the attenuation and the background noise. The transmission loss (TL) in ice can be calculated using the equation $TL = 20\log(x) + \alpha x$, which accounts for the effect of spreading and absorption loss. Recent measurements of the attenuation of acoustic signals in the South Pole ice (Abbasi et al. 2011) at depths between 190 and 500 m suggest attenuation length $\lambda = 1/\alpha$ of 0.3 km, where α is the attenuation coefficient, with 20% uncertainty for signals in the 10–30 kHz range corresponding to an amplitude attenuation coefficient of $\alpha = 3.20 \pm 0.57 \text{ km}^{-1}$ which was larger than previous theoretical estimates (Vandenbroucke et al. 2009). These data suggest an attenuation is about 28 dB/km in the frequency range studied. Note that this attenuation is significantly higher than that of water, which is less than 5 dB/km at 10–20 kHz (Stojanovic and Preisig 2009).

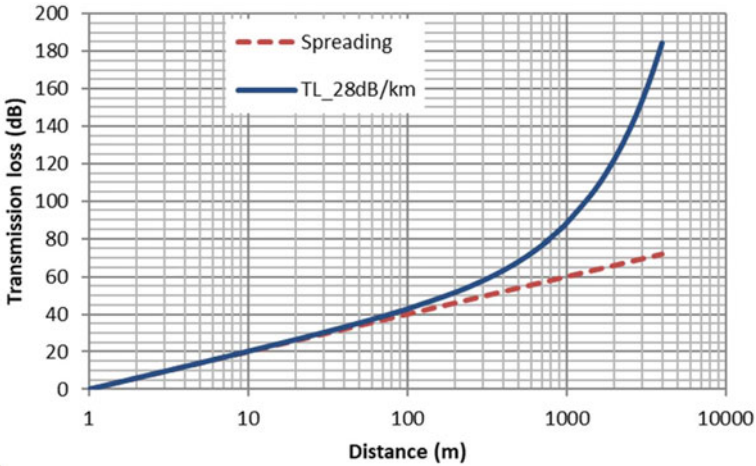


Fig. 28.6 Estimated transmission loss for 20 kHz acoustic wave in ice

According to the South Pole data (Abbasi et al. 2011), the total noise level (in the 10–50 kHz frequency range) was measured to be 20 mPa above 200 m depth and 14 mPa in deep ice. Thus, if we use the noise level measured in South Pole ice (20 mPa over 10–50 kHz) to estimate the noise level (NL), $NL = 20\log(20 \text{ mPa}/1 \mu \text{ Pa}) = 86 \text{ dB}$. Then, by applying available measured attenuation data of South Pole ice (28 dB/km), the sum of transmission loss (TL) and beam spreading in dB versus distance are shown in Fig. 28.6.

28.4.2 Communication Capacity of Acoustic Channel

In order to determine the baud rate, we applied the Shannon–Hartley theorem (Shannon 1949) to calculate the communication capacity, C , of the acoustic channels from one puck to the next.

$$C = B\log_2(1 + s/n),$$

where B is the bandwidth and s/n is the ratio of signal to noise.

We utilized the sonar equations (Urick 1975) to calculate the ratio of signal to noise. The signal level, S , in dB referring to the intensity of a plane P-wave is expressed as

$$S = SL - TL,$$

where SL is the source level and TL is the transmission loss as described in previous section. The source level is written as

$$SL = SL_0 + 20\log(P) + DI,$$

where P is the acoustic power of the source, DI is the directivity index of the transmitting transducer, and SL_0 is the sound level at 1 m distance of an omnidirectional acoustic source of 1 W in the acoustic medium. For a P-wave in ice of 3900 m/s in sound speed and 917 kg/m³ in density, SL_0 is equal to 174.5 dB.

The directivity index of a circular piston transducer is estimated as

$$DI = 20\log(\pi d/\lambda),$$

where d is the diameter of the transducer and λ is the wavelength. The diameter of the transducers is limited by the diameter of the probe. For a circular piston transducer 22 cm in diameter the directivity index DI is 11.0 dB at 20 kHz.

The signal-to-noise ratio level in dB, SNR , is written as

$$SNR = S - NL,$$

where NL is the noise level received by the receiving transducer and is calculated as

$$NL = NS + 10\log(B) - DI,$$

where NS is ambient noise spectrum level and considered being omnidirectional. The use of $10\log(B)$ is because the noise is random. The DI is the directivity index of the receiving transducer, which is equal to the directivity index of the transmitting transducer if both have the same structure and dimensions.

A study (Panning et al. 2018) suggests that the expected ambient background noise in ice shell of Europa in frequency range of 0.005–2 Hz caused by tide-induced ice cracking and turbulent motion in the oceans may be much lower than the noise level in Earth. Therefore, we still use South Pole data and point out that the data might be higher than Europa's. Future studies may provide more meaningful predictions. In this study we used the noise level measured in South Pole ice (14 mPa over 10–50 kHz), i.e., corresponding noise spectrum level $NS = 39.636.9$ dB, to estimate the noise level (NL) in Europa ice, where NS is the noise spectrum level = $20\log(X) - 10\log(B)$, and B is the bandwidth.

Applying the equations listed above and assuming frequency of 20 kHz and channel bandwidth of 5 kHz, we estimated the signal-to-noise ratio and acoustic communication capacity as the function of the distance (spacing of the communication pucks) for electric supply powers of 0.1, 1.0, 10 and 100 watts, respectively. The results are presented in Figs. 28.7 and 28.8. The electroacoustic efficiency of the transducer was set as 0.5.

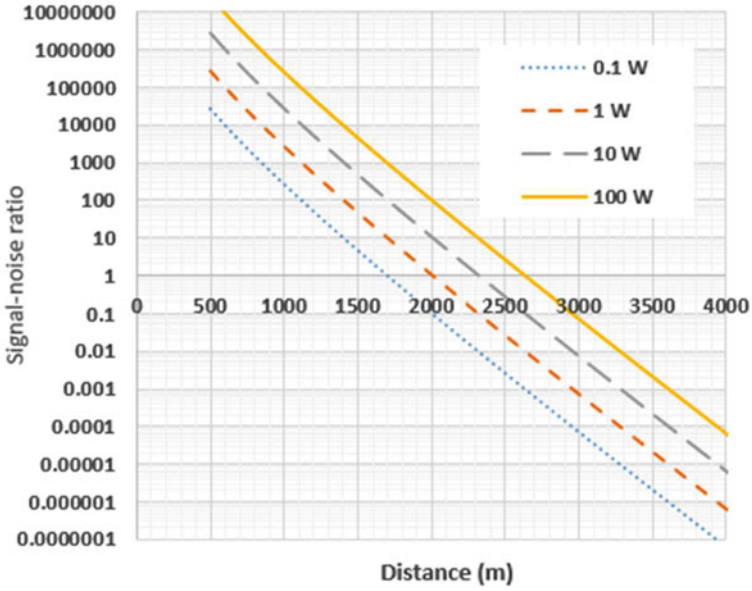


Fig. 28.7 Signal-to-noise ratio versus distance for various transmission powers

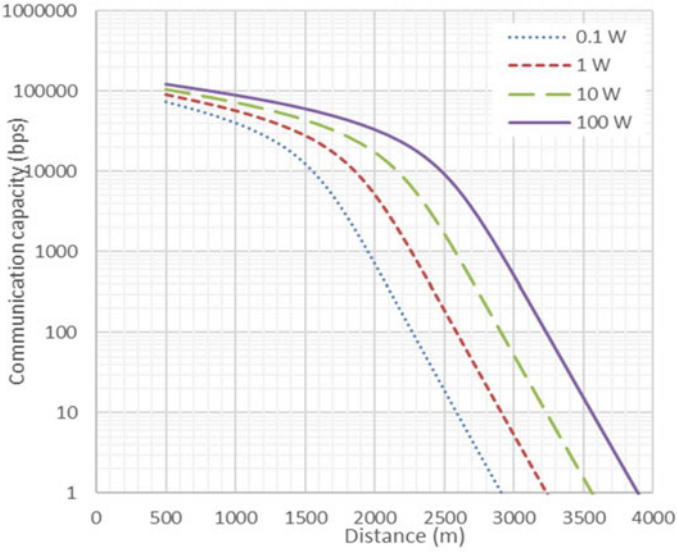


Fig. 28.8 Communication capacity versus distance for different transmission powers

28.4.3 *Communication Capacity Under Limited Power Supply*

The electric power supply available for the wireless communication pucks will be limited. We assume the source of electricity would be a miniature RTG capable of providing 120 mW electric power (Whalen 2008). This power level would not be enough to operate the communication relay puck continuously: a rechargeable battery or super-capacitor would be required to store electricity to allow for a higher power transmission burst. Using the charged battery or capacitor we can operate the modem with different transmission power levels. A higher power operation is able to get higher communication speed, but also means a shorter operating time. A trade-off is needed for performance optimization.

Taking a commercially available acoustic modem as a reference (Teledyne Marine 2019), we propose a reasonable set of potential power parameters for the communication puck as.

- (1) Power generated by RTG: 0.12 W.
- (2) Power in idle mode: 0.01 W.
- (3) In active mode when relaying signals, power needed for receiving, control and other support circuits except the power output to transmitting transducer: 0.6 W.
- (4) Efficiency (charge/discharge, etc.): 0.9.

The maximum time ratio of active mode to total time depends on the selection of transmission power. The calculated ratios and the energy efficiencies (transmission energy/total energy) for transmission power of 0.1, 1, 10 and 100 W are listed in Table 28.1. The energy efficiency is lower in the lower transmission power operation, especially for the 0.1 W.

The achievable data rates averaged in total time are shown in Fig. 28.5 as a function of distance for the four operation power choices. The results can be used to help determine puck deployment spacing and operating power selection. In general, for best performance with limited available power supply, when a high data rate is required, we should operate the acoustic modem at low transmission power mode and shorten the spacing. When large spacing is needed we should operate at high power mode. Currently we think an average data rate of 1000 bps, i.e., 95 G bits data in 3 years, might be comfortably sufficient for a potential ocean-accessing probe. Selecting 1 W transmission power mode can achieve this rate as well as covering a distance of 1700 m. In this operation mode the acoustic transceiver relays the data at 16,000 bps in active time with 1 W transmission power (Fig. 28.9) and the active

Table 28.1 Active mode ratios and energy efficiency

Transmission power (W)	0.1	1	10	100
Active mode total (W)	0.7	1.6	10.6	100.6
Active time/total time	14%	6.125%	0.9245%	0.0974%
Transmission energy/total energy	0.117	0.510	0.770	0.812

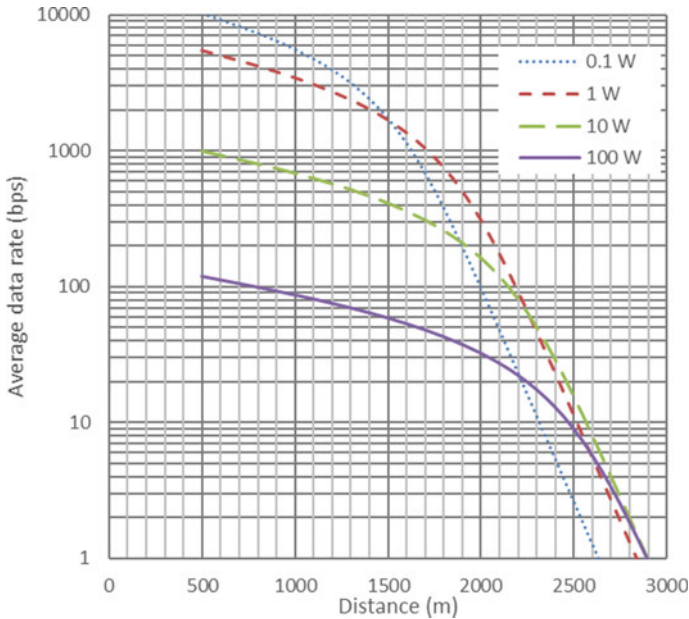


Fig. 28.9 The all-time average communication data rate with 120 mW available electric power versus distance for different transmission power selections

time is limited to 6.125% of the total time with a 120 mW RTG (Table 28.1). When an average rate of 100 bps (9.5 G bits per 3 years) is acceptable, the spacing of the transceivers could be extended to 2200 m.

28.4.4 Communication Summary

Based on the acoustic attenuation and noise level data measured in the South Pole ice (Abbasi et al. 2011), the communication capability of an acoustic transceiver operating at 20 kHz was analyzed as a potential design example for applications to ocean-accessing probes in a cold ocean world. The results suggest that with a limited electric power supply of 120 mW, the transceiver may achieve an average data transmission rate of 1000 bps to a transceiver puck spacing of 1700 m, or communicate at an average rate of 100 bps to 2200 m.

For longer-distance communication, a lower operating frequency could be considered. The attenuation coefficient is expected to decrease with the frequency decrease but the environment noise power spectrum density may increase. It is difficult to estimate the noise level in Europa ice, so a reasonable assumption has been made

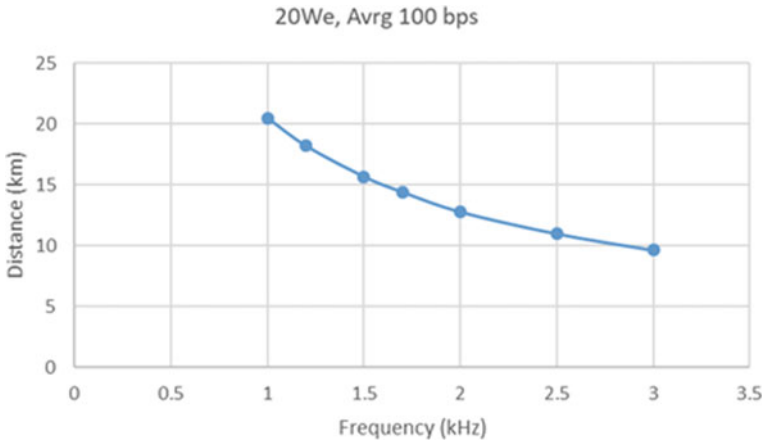


Fig. 28.10 Communication distance versus frequency with 20 We for data rate of 100 bps

from the best available information. We assume the attenuation coefficient is proportional to the frequency. And we assume the noise power spectrum density proportional to the inverse of the frequency that makes the assumed noise level at ~ 2 Hz roughly matching the estimated background noise in Europa's ice shell (Panning et al. 2018). As shown in Fig. 28.8, the communication distance is greatly increased at low frequency e.g., ~ 20 km with 20 We (Watts electric) at 1 kHz for 100 bps (Fig. 28.10) in comparison with ~ 3 km at 20 kHz (see Fig. 28.8). Although the estimate is very rough with very high uncertainties, the results of operating the modem at low frequency with relatively high power are still encouraging for long-distance communication. (The high power supply of 200 W or above is assumed to be available to the deepest puck directly from the ice descent probe.)

28.5 Acoustic Transceiver

28.5.1 Piezoelectric Transducer

Various transduction approaches for generating and sensing acoustic waves are widely used, including electromagnetic, magnetostrictive, electrostrictive, and piezoelectric transduction. Common cryogenic transduction techniques are summarized in Table 28.2 (Bar-Cohen 2016). Piezoelectric transduction, which involves shape changes on exposure to external electric fields, has found numerous application areas including sensors, actuators, sonars, and transducers that operate at a broad range of temperatures and environments. Immunity to magnetic fields, low power consumption, efficient conversion of electricity to mechanical energy or vice versa, and low heat generation make piezoelectric devices superior compared with other

Table 28.2 Comparison of common cryogenic transduction techniques

Cryogenic techniques	Typical strain level (%)	Advantages	Disadvantages
Magnetostriction (Dooley et al. 1999)	1	High strain level Better performance at low temperatures	Need for superconducting coil High materials cost Bulky system
Electrostriction (Mulvihill et al. 2002)	0.01	Broad material choices Low power consumption Immunity to magnetic field	Low strain level High electric voltage
Piezoelectricity (Jiang and Rehrig 2005)	0.2	Immunity to magnetic field Low power consumption High efficiency	Lower performance at low temperature

techniques in many applications. Moreover, piezoelectric materials can sustain exposure to high temperatures, making them compatible with methods of addressing planetary protection requirements.

The critical factor that determines the performance of a piezoelectric transducer is the transmitting acoustic pressure level (or acoustic power). For a plane wave propagating in the *z* direction, the amplitude of radiation pressure is equal to the product of the characteristic acoustical impedance of the propagating medium (*Z*) and vibration velocity (*u*₀) (*p*₀ = *Z*·*u*₀). The time average of radiated acoustic power (*P*_a) is equal to 1/2*R*_r*u*₀², where *R*_r is the resistance associated with transfer of power from the transducer by a surface, *A*, (i.e., *R*_r ~ *Z*·*A*). Thus, the transmitted acoustic power is proportional to the square of the vibration velocity of the transducer material in a medium, which is a function of piezoelectric material properties, specifically the product of piezoelectric strain *d* coefficient and mechanical quality factor *Q*_m, i.e., *u*₀ ~ ω(*d*₃₃·*Q*_m)*V*, where *V* is the input voltage. Therefore, both high *d*₃₃ and *Q*_m of piezoelectric materials allow the delivery of acoustic power levels at cryogenic temperatures.

For cyrogenic applications, it is important to understand the influence of temperature on piezoelectric responses. Currently, the majority of piezoelectric materials are ferroelectric materials, which exhibit a spontaneous polarization in the absence of an electric field. The piezoelectric response contains not only the intrinsic contribution, but also an extrinsic contribution caused by movement of non-180° ferroelectric domain walls. For instance, PZTs are generally tailored with the use of off-valence dopant ions, which impede or facilitate domain wall movements. Note that for PZT ceramics, more than 50% of the net piezoelectric coefficient arises from the motion of non-180° ferroelectric domain walls. When PZT materials are used at cryogenic temperatures, the extrinsic contributions (i.e., ferroelectric domain wall

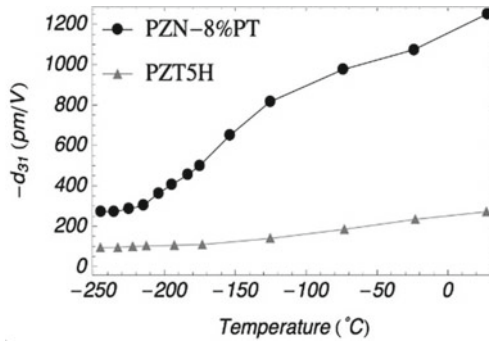


Fig. 28.11 Temperature-dependent piezoelectric transverse coefficients (d_{33}) for PZN-8%PT single crystals compared with those of PZT5H ceramics (Jiang and Rehrig 2005)

motion) are frozen out, and consequently the piezoelectric materials lose their piezoelectric performance. For example, the piezoelectric d_{33} coefficient was reported to be decreased from 760 pC/N to 220 pC/N when the operating temperature was decreased from 300 to 30 K (Jiang and Rehrig 2005). This points to the need for appropriate piezoelectric materials for use in communication through ice at cryogenic temperatures.

Studies of relaxor-ferroelectric-based single crystals have revealed their extremely high piezoelectric responses, strain over 1.7%, and it has been reported that they offer significantly higher piezoelectric performance at both room temperature and cryogenic temperatures, e.g., d_{33} of single-crystal piezoelectrics (PMN-PT or PZN-PT) at 30 K is about equal to the d_{33} of PZT-5A at room temperature, indicating promise for cryogenic applications. Figure 28.11 shows a comparison between PZT5H ceramics and PZN-8PT single crystals in terms of the piezoelectric coefficients as a function of temperature. The results suggest that, although both polycrystalline PZT ceramics and single-crystal PZN-PT materials show decreased piezoelectric coefficients with decreasing temperature, single-crystal PZN-PT materials possess much higher values, exhibiting the room temperature piezoelectric strain coefficient of soft PZT materials even at -200 °C. This implies that the similar acoustic source level of transducer can be achieved by replacing conventional PZT material with relaxor-ferroelectric single crystal at cryogenic temperatures.

28.5.2 Acoustic Transceiver Design

Working conditions and operating requirements for acoustic transceivers are listed in Table 28.3. Several types of underwater piezoelectric transducer, such as the piston type, ring-type, flexural disks and flextensional-type transducers, can be candidates for transmitting and receiving acoustic signals in the desired frequency range

Table 28.3 The working conditions and requirements for the acoustic transceivers

	Range	Remark
Environment temperature	110 K near ice surface and 273 K in the ocean	The transceivers will need to survive and be operational at low temperatures
Transceiver temperature	Higher than the environment temperature	Will depend on the RTG power, thermal control design and the heat converted from the electric power to the transceiver
Frequency	10 s kHz	Current best estimation from trading wave attenuation and communication data transfer rate
Power	Watts to 10 s of watts	Current best estimation from trading RTG power and communication distance
Size	Diameter < 10–30 cm	Depends on the melting probe diameter and the potential RTG size

and dimensions. A thorough description of the physics and characteristics of these transducer designs is given elsewhere (Sherman and Butler 2007).

Several types of underwater piezoelectric transducer can be candidates for transmitting and receiving acoustic signals in the desired frequency range and dimensions in ice. Transmitting through ice involves acoustic impedance that is higher than for water and the generation of shear waves. The shear wave has a speed lower than the pressure wave, which is the wave mode that is supported in fluids. Therefore, at a distance in ice, the transducer may receive two signals with different time delays or phases that may interfere with the communication. To address the issues related to operation in ice, numerical models must be established for various potential transducer types, and simulations performed to predict and optimize the selected type and its performance in ice.

For acoustic communication transducers, the cymbal, which is a miniaturized Class V flexensional transducer, can be a candidate because of its small size, low cost, and low weight (Tressler et al. 2006). This design can generate lower-frequency (1–10 kHz) sound waves at the power levels required. In addition, its performance has been measured and modeled extensively and proven to be effective for underwater sonar transducers (Dogan et al. 1997; Sherman and Butler 2007; Zhang et al. 2016), showing great potential for acoustic communication in ice. A schematic of the cross-section of a cymbal-type flexensional transducer is shown in Fig. 28.12. As depicted, the cymbal transducer consists of a thin piezoelectric disk sandwiched between and bonded to two cymbal-shaped metal and end caps. The operating principle of this cymbal-type transducer is the coupling of radial vibrations of the piezoelectric disk to the metallic flexensional frame (cymbal). This results in a mechanical amplification

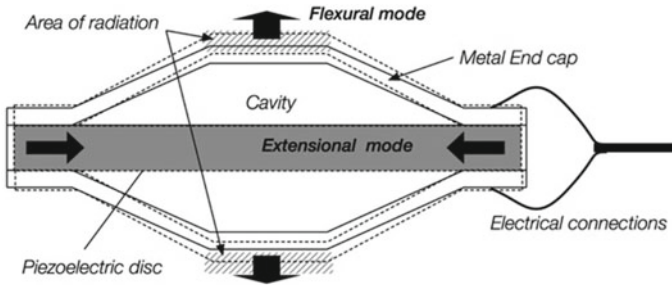


Fig. 28.12 Schematic diagram of the cymbal flextensional transducer. Arrows show the displacement direction under positive bias

of the radial displacement into a much larger axial displacement, thereby enhancing the vibration velocity.

Even though the cymbal flextensional transducer has the advantages of a thin profile, low cost and wider bandwidth, the drawback of this transducer design is a low radiating surface area (i.e., the diameter of the cymbal is on the order of half-inch) compared to the wavelength in water around resonance; thus, it has a very low radiation resistance and a relatively high radiation reactance, which means that the transfer of radiated acoustic power from the surface of the radiator to the water is very inefficient. To overcome this issue, cymbal array configurations have been studied and a configuration proposed (Tressler and Newnham 1999) that can greatly improve transmitting efficiency. Another drawback of the flextensional transducer is that it can only operate at shallow depths as the air-backed acoustic flexures undergo degradation in performance when exposed to elevated hydrostatic pressure, which can be a key limiting factor to implementing acoustic communication in ice.

Unlike flextensional-based transducer designs, a ring-type transducer encounters no performance changes with hydrostatic pressure (Fig. 28.13). They can also perform well in both transmitting and receiving, and can be manufactured in large dimensions, so a low-frequency operation (<1 kHz) can be achieved by adjusting ring diameter and length. One drawback is a poor radial directivity, which is handled by an omnidirectional beam (toroidal shape in the horizontal plane).

28.6 Conclusions

One of NASA's priorities is to explore ocean worlds in the solar system where there is a potential for life under the ice shell. This requires reaching the ocean below through great depths of ice that are extremely cold. Europa (one of Jupiter's moons) is estimated to have a 40 km ice shell. An approach using a melting probe called Cryobot to reach the ocean has been conceived for a potential future mission. This European melt probe was named PRIME (Probe using Radioisotopes for Icy Moons

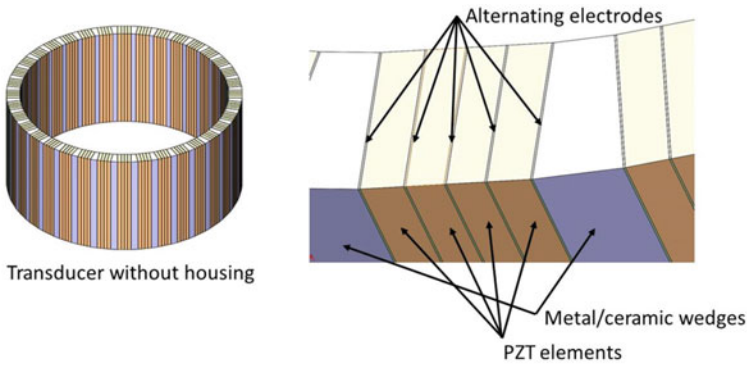


Fig. 28.13 Transducer cylindrical configuration

Exploration). A lander was assumed to be the platform from which the Cryobot would be deployed. The ice-penetrating vehicle concept consists of a cylindrical, narrow-body probe that encases a radioisotope heat/power source that would be used to melt through the icy crust. The baseline design of the probe includes a suite of scientific instruments to analyze the ice during descent and the liquid ocean underneath. For communication, a set of fiber optic wire as well as wireless RF in the very cold porous top layer is assumed, then acoustic modems are used for communication in warmer, denser ice over distance that can potentially reach a distance of 20 km level between the modems with 20 W power. Assuming 100 bps is a low requirement of the PRIME probe, greater power (200 W or above) is available for the deepest pack from the probe for up-stream data transmission while down-stream transmission may be at a lower rate.

The reachable distance is strongly dependent to the properties of the ice and the noise level in the ice shell. In an effort to prevent potential collision with solid obstacles along the path as well as finding the depth of the water interface, a forward-facing sonar has been considered. The focus of this chapter was on the use of elastic waves and piezoelectric transducers for communication and sonar for the avoidance of obstacles.

Acknowledgements Research reported in this abstract was conducted at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with National Aeronautics Space Administration (NASA). This chapter also presents the latest results from an ongoing JPL formulation and technology development effort for the European melt probe “PRIME—Probe using Radioisotopes for Icy Moons Exploration”, which was funded by an internal award. The authors would like to thank the PRIME team members who contributed to the conception of the Cryobot configuration that is reported in this chapter. They include Scott H. Bryant, Jean-Pierre Fleurial, Jeremiah B. Gayle, Terry J. Hendricks, Samuel M. Howell, Emily F. Klonicki, Juergen Mueller, Knut I. Oxnevad, Adarsh Rajguru, Miles Smith and David F. Woerner, Jet Propulsion Laboratory/California Institute of technology.

References

- Abbasi, R., Y. Abdou, T. Abu-Zayyad, J. Adams, J.A. Aguilar, M. Ahlers, K. Andeen, et al. 2011. Measurement of acoustic attenuation in South Pole ice. *Astroparticle Physics* 34 (6): 382–393.
- Bar-Cohen, Y. (Ed.). 2016. *Low Temperature Materials and Mechanisms*, ISBN-10: 1498700381, ISBN-13: 978–1498700382, CRC Press, Taylor & Francis Group, Boca Raton, FL, 500 pages.
- Bar-Cohen, Y., and K. Zaczny (Eds). 2020. *Advances in Terrestrial and Extraterrestrial Drilling (Design, Modeling, and Specialized Techniques for Extreme Environments)*, Vol. 2: Advances in Extraterrestrial Drilling, ISBN-10: 0367653478; ISBN-13: 978–0367653477; CRC Press/Taylor & Francis Group LLC, Boca Raton, Florida, 364 pages.
- Benjamin, K. 2008. Transducers for sonar systems. In *Handbook of Signal Processing in Acoustics*, 1783–1819. New York: Springer.
- Benson, B., Y. Li, B. Faunce, K. Domond, D. Kimball, C. Schurgers, and R. Kastner. 2010. Design of a low-cost underwater acoustic modem. *IEEE Embedded Systems Letters* 2 (3): 58–61.
- Cwik, T. et al. 2018. A technology architecture for accessing the oceans of icy worlds. *69th International Astronautical Congress (IAC)*, IAC-18-A7.3.4, Bremen, Germany, 1–5 October 2018.
- Di Pippo, S., R. Mugnuolo, P. Vielmo, and W. Prendin. 1999. The exploitation of Europa ice and water basins. *Planetary and Space Science* 47: 921–933.
- Dogan, A., K. Uchino, and R.E. Newnham. 1997. Composite piezoelectric transducer with truncated conical endcaps ‘Cymbal.’ *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 44 (3): 597–605.
- Dooley, J.A., C. Lindensmith, R. Chave and N. Good. 1999. Magnetostriction of single crystal and polycrystalline Tb_{0.60}Dy_{0.40} at cryogenic temperatures. *Journal of Applied Physics* 85(8), 6256–6258. <https://doi.org/10.1063/1.370128>
- Flaurial, J.-P., S. Howell, D. Woerner, D. Landau, J. Gayle, M. Smith, J. Mueller, T. Hendricks, S. Roberts, B. Nesmith, S. Bryant, Y. Bar-Cohen, E. Klonicki, W. West, B. Hockman and B. Clement. 2019. Notional Concept of Operations and System Capability Definition for Enabling Scientific Ocean Access Missions on Icy Worlds. paper No. EPSC-DPS2019–1268, OPS3/EXO16—Ocean Worlds and Icy Moons—EPSC-DPS Joint Meeting.
- Hand, K.P., and the Project Engineering Team. 2017. *Report of the Europa Lander Science Definition Team*, Tech. Rep. JPL D-97667, NASA.
- Jiang, X., and P.W. Rehrig. 2005. Advanced piezoelectric single crystal based actuators. presented at the *Proc. of SPIE*.
- Kowalski, J., P. Linder, S. Zierke, B. von Wulfen, J. Clemens, K. Konstantinidis, G. Ameres, R. Hoffmann, J. Mikucki S.Tulaczyk O. Funke, D. Blandfort, C. Espe, M. Feldmann, G. Francke, S. Hiecker, E. Plescher, S. Schöngarth and K. Helbing. 2016. Navigation technology for exploration of glacier ice with maneuverable melting probes. *Cold Regions Science and Technology*, Elsevier 123, 53–70 <https://doi.org/10.1016/j.coldregions.2015.11.006>
- Lee, S., M. Zanolin, A.M. Thode, R.T. Pappalardo, and N.C. Makris. 2003. Probing Europa’s interior with natural sound sources. *Icarus* 165 (1): 144–167.
- Lee, S., R.T. Pappalardo, and N.C. Makris. 2005. Mechanics of tidally driven fractures in Europa’s ice shell. *Icarus* 177 (2): 367–379.
- Lishman, B., J. Wadham, B. Drinkwater, J. Kendall, S. Burrow, G. Hilton, and I. Craddock. 2013. Assessing the utility of acoustic communication for wireless sensors deployed beneath ice sheets. *Annals of Glaciology* 54 (64): 124–134. <https://doi.org/10.3189/2013AoG64A022>
- Mulvihill, M.L., R.J. Shawgo, R.B. Bagwell, et al. 2002. Cryogenic cofired multilayer actuator development for a deformable mirror in the next generation space telescope. *Journal of Electroceramics* 8: 121–128. <https://doi.org/10.1023/A:1020551812502>
- National Research Council Space Studies Board. 2012. *Vision and voyages for planetary science in the decade 2013–2022*. National Academies Press.
- National Research Council Space Studies Board. 2018. *Astrobiology Science Strategy for the Search for Life in the Universe*. National Academies Press.

- Panning, M.P., et al. 2018. Expected seismicity and the seismic noise environment of Europa. *Journal of Geophysical Research—Planets* 123 (1), 163–179.
- Pappalardo, R.T., L.M. Prockter, D.A. Senske, R. Klima, S.F. Vance, K. Craft, and the Europa Science Team. 2016. Science objectives and capabilities of the NASA Europa mission. *47th Lunar and Planetary Science Conference* #3058.
- Shannon, C.E. 1949. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers* 37 (1): 10–21.
- Sherman, C.H., and J.L. Butler. 2007. *Transducers and Arrays for Underwater Sound*. New York: Springer.
- Stojanovic, M., and J. Preisig. 2009. Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Communications Magazine* 47 (1): 84–89.
- Teledyne Marine. 2019. Acoustic modems 960 Series ATM-966. <http://www.teledynemarine.com/960-series-atm-966?ProductLineID=8> (Oct. 23 2019).
- Tressler, J., and R. Newnham. 1999. Capped ceramic underwater sound projector: The “cymbal” transducer. *The Journal of the Acoustical Society of America*.
- Tressler, J.F., T.R. Howarth, and D. Huang 2006. A comparison of the underwater acoustic performance of single crystal versus piezoelectric ceramic-based “cymbal” projectors. *The Journal of the Acoustical Society of America*.
- Urick R.J. 1975. *Principle of underwater sound*, 16–28, 2nd edition. McGraw-Hill Books.
- Vandenbroucke, J., and IceCube Collaboration. 2009. Measurement of acoustic properties of South Pole ice for neutrino astronomy. *Nuclear Instruments and Methods in Physics Research Section a: Accelerators, Spectrometers, Detectors and Associated Equipment* 604 (1): S164–S170.
- Whalen S.A., A.A. Christopher, and L.A. Terrence. 2008. Improving power density and efficiency of miniature radioisotopic thermoelectric generators. *Journal of Power Sources* 180 (1): 657–663.
- Zhang, S., H.J. Lee, and S. Sherrit. 2016. Low temperature transducing materials. Chapter 7 in Bar-Cohen Y. (Ed.), (2016), *Low Temperature Materials and Mechanisms*, ISBN-10: 1498700381, ISBN-13: 978-1498700382, CRC Press, Taylor & Francis Group, Boca Raton, FL, pp. 183–212.
- Zimmerman W. et al. 1998. Europa Cryo-Hydro Integrated Robotic Penetrator System (CHIRPS) Feasibility Study, Jet Propulsion Laboratory/ California Institute of Technology, July 1998.
- Zimmerman W., S. Bryant, J. Zitzelberger and B. Nesmith. 2001a. Radioisotope powered Cryobot for penetrating the European ice shell. *American Institute of Physics Conference Proceedings* 552, 707.
- Zimmerman, W., R. Bonitz, and J. Feldman. 2001b. Cryobot: An ice penetrating robotic vehicle for Mars and Europa. *IEEE Aerospace Conference Proceedings* 1: 1311–1323.

Chapter 29

Ice Melting Probes



**Bernd Dachwald, Stephan Ulamec, Julia Kowalski, Marc S. Boxberg,
Fabian Baader, Jens Biele, and Norbert Kömle**

Abstract The exploration of icy environments in the solar system, such as the poles of Mars and the icy moons (a.k.a. ocean worlds), is a key aspect for understanding their astrobiological potential as well as for extraterrestrial resource inspection. On these worlds, ice melting probes are considered to be well suited for the robotic clean execution of such missions. In this chapter, we describe ice melting probes and their applications, the physics of ice melting and how the melting behavior can be modeled and simulated numerically, the challenges for ice melting, and the required

B. Dachwald (✉) · F. Baader
Faculty of Aerospace Engineering, FH Aachen University of Applied Sciences,
Hohenstaufenallee 6, 52064 Aachen, Germany
e-mail: dachwald@fh-aachen.de

F. Baader
e-mail: fabian.baader@fh-aachen.de

S. Ulamec · J. Biele
German Aerospace Center (DLR), Space Operations and Astronaut Training – MUSC, Linder
Höhe, 51147 Cologne, Germany
e-mail: stephan.ulamec@dlr.de

J. Biele
e-mail: jens.biele@dlr.de

J. Kowalski
Georg-August University of Göttingen, Computational Geoscience, Goldschmidtstr. 1, 37077
Göttingen, Germany
e-mail: julia.kowalski@geo.uni-goettingen.de

M. S. Boxberg
RWTH Aachen University, AICES and Faculty of Georesources and Materials Engineering,
Schinkelstraße 2a, 52062 Aachen, Germany
e-mail: boxberg@aices.rwth-aachen.de

N. Kömle
Space Research Institute, Austrian Academy of Sciences, Schmiedlstrasse 6, 8042 Graz, Austria
e-mail: norbert.koemle@oeaw.ac.at

key technologies to deal with those challenges. We also give an overview of existing ice melting probes and report some results and lessons learned from laboratory and field tests.

29.1 Introduction

Although the icy bodies in the solar system are currently not considered as resource providers for the near or medium-term future, it is helpful to discuss the technologies required for their later exploration to guide early economic and political decisions. Technologies for penetrating thick ice layers can help in identifying methods for future resource prospection.

A probe can penetrate ice via three principal methods: melting (thermal), cutting/drilling (mechanical), or mixed cutting and melting (hybrid). Melting requires a considerable amount of energy, but the removal of ice to the rear end of the probe as meltwater happens automatically with no action. Cutting requires less energy but faces the problem of removing the unmelted ice chips to the rear end of the probe without clogging. Hybrid probes aim to combine the advantages of cutting and melting (Zacny et al. 2018). At the same time, the Cryobot avoids a major flaw of earlier ice melting probe designs, namely a highly stressed conductor wire for the probe's power supply over distances of several kilometers, which are initially stored on compact spools (French et al. 2001).

In this chapter, we focus on ice melting probes, which are sometimes also referred to as thermal drills. Section 29.2 gives a brief overview of potential applications of ice melting probes in the solar system and on Earth. After this overview, Sect. 29.3 describes the physical theory behind ice melting probes and methods to model and simulate their melting behavior and performance. Section 29.4 then discusses the challenges for ice melting probes and the key technologies to deal with them. Some designs of existing ice melting probes which have been developed by various research institutions, together with some of their achievements, are described in Sect. 29.5. This section also presents some key insights, field test experience and experimental results related to the terrestrial and extraterrestrial application of ice melting probes. Finally, we summarize the main points and draw some conclusions in Sect. 29.6.

29.2 Applications for Ice Melting Probes

Besides their usefulness for the exploration of resources, the main driver for ice melting probes currently lies in the field of (astro)biology and glaciology. Because life can thrive in icy environments and subglacial lakes on Earth (Bidle et al. 2007; D'Elia et al. 2008; Deming and Eicken 2007; Priscu and Christner 2004), it is now widely debated whether environmental niches can exist on other solar system bodies, either within the ice or in aquatic environments beneath the ice, that may harbor

Table 29.1 Relevant parameters for ice melting probes on Earth, Mars, Europa, and Enceladus

Parameter	Earth	Mars	Europa	Enceladus
Mass in kg ^a	5.972×10^{24}	6.417×10^{23}	4.800×10^{22}	1.080×10^{24}
Mean radius in km ^a	6371.0	3389.5	1560.8	252.1
Surface gravity in $\frac{m}{s^2}$	9.81	3.72	1.31	0.113
Surface temperature in K ^h	220–270 ^b	205–210 ^b	50–125 ^c	33–145 ^c
Surface radiation in $\frac{Gy}{d}$	≈ 0	$\approx 0.03^d$	$\approx 10^{5e}$	$\approx 10^{2e}$
Atm. pressure in mbar ⁱ	≈ 1000	4–9 ^f	Effectively 0	
Ice composition ^c	H ₂ O, (some air)	H ₂ O, CO ₂ , dust	H ₂ O, (possibly salts, dust, CO ₂ , CH ₄)	

^aNASA JPL (2021) ^bUlamc et al. (2007)

^cDachwald et al. (2013) ^dSimonsen and Nealy (1993)

^eDachwald et al. (2020) ^fNASA NSSDCA (2020)

^gcalculated from mass and radius ^hpolar values for Earth and Mars

ⁱexact value depends on altitude and atmospheric conditions

extraterrestrial life. The ones believed to have the highest potential are Mars, Jupiter’s moon Europa, and Saturn’s moon Enceladus. The most relevant parameters for ice melting probes on Earth, Mars, Europa, and Enceladus are summarized in Table 29.1.

For extraterrestrial applications, drill rigs appear to be far too heavy and complex to be implemented in currently considered lander missions. Ice melting probes may therefore be more feasible and also the cleaner option to comply with planetary protection rules. They are much less complicated and simpler to operate, and can be made autonomous and steerable (Kowalski et al. 2016), although there remain significant technological challenges, especially for long-range ice melting in the order of kilometers (see Sect. 29.4). Short-range ice melting in the order of meters, however, seems to be a first relatively low-cost and low-mass option to reveal at least some of the information that is concealed in the ice and underneath it (Di Pippo et al. 1999; Ulamec et al. 2005). For such an ice melting probe, a Li-ion battery may already be sufficient to reach a depth of several meters (Biele et al., 2011).

29.2.1 Mars

For future exploration missions, Mars is certainly one of the prime targets, and at some point, resource utilization of material from Mars (and its polar caps) will become relevant. In contrast to missions to the icy moons of Jupiter and Saturn, missions to Mars impose significantly smaller propulsive requirements (astrodynamically speaking, they have a much lower ΔV), and therefore they are much easier to perform. Conse-

quently, Mars may be the first target for the in situ exploration of extraterrestrial ices. So far, no clear evidence for extant or extinct life has been found on Mars. However, the most promising places where extremophilic life could still exist, assuming there had been life in the past, have not yet been explored. The average temperature on the surface is roughly -60°C , with a maximum temperature of about 20°C at the equator during summer, and a minimum temperature of about -140°C at the poles in winter. However, places like the polar caps and the recently discovered subglacial lakes beneath them (Orosei et al. 2018; Lauro et al. 2020) contain liquid water and would shelter life forms from the harsh and hostile Martian surface environment. Nevertheless, their suitability for life, which depends on factors such as pH, water activity, salinity, nutrients, and environmental stability, is debated (see, e.g., Cockell et al. 2011; Stevenson et al. 2015; Checinska Sielaff and Smith 2019; Tarnas et al. 2021). The polar ice caps of Mars are so large that they are already visible through an ordinary telescope. They cover an area of approximately $1 \times 10^6 \text{ km}^2$, with a thickness as great as 3 – 4 km (Clifford et al. 2000). With about $2.6 \times 10^6 \text{ km}^3$, their volume is similar to that of the Greenland ice sheet (Byrne 2009). Mirroring the recent atmospheric and climatic history of Mars, both polar caps have a complex stratigraphy and geomorphology with similar features but also differences in size, thickness, and composition. The polar deposits consist mainly of water ice with layers of dust and seasonal CO_2 -ice on the surface (Byrne 2009; Orosei et al. 2015). The polar ice caps of Mars are a prime target for in situ subsurface exploration with an ice melting probe. Besides searching for life in the ice or in the subglacial lakes, even short-range melting probes would allow us to unravel the recent climate history of Mars.

29.2.2 *Icy Moons*

Two moons in the solar system are expected to be of very high astrobiological relevance, because there is evidence from various measurements that they have global water oceans in contact with the rocky core: Jupiter's moon Europa and Saturn's moon Enceladus (Nimmo and Pappalardo 2016).

Europa's surface shows almost no craters, which implies that it is covered by a geologically very young layer of water ice. Furthermore, the surface structures indicate geologically recent movements of the ice crust and a global subsurface ocean (Greenberg and Geissler 2002) whose volume is about twice that of Earth's oceans (Chyba and Phillips 2007) and that is warmed by tidal flexing. The surface temperature on Europa ranges between 86 and 132 K (Prockter and Pappalardo 2014). For the thickness of the ice crust, after much scientific debate, there is still no generally accepted model available. Values vary between a few to a few tens of kilometers (Nimmo and Pappalardo 2016). For reference calculations in Sect. 29.4, a "typical" value of 22 km (from Howell 2020) is assumed, although local thicknesses may vary considerably.

For Enceladus, the lower and upper limits of the average ice shell thickness are 18 km and 44 km, respectively (Hemingway et al. 2018). Although it is now generally agreed that the ice shell is thinnest at the poles and thickest at the equator, model predictions for the south pole still vary between 2 and 20 km. Recent investigations of microwave emission radiometry at the south polar terrain (SPT) by Le Gall et al. (2017), however, are only in agreement with an ice shell thickness of not more than 5 km. Cracks at the SPT appear to cut through the entire ice shell and tap into the global ocean underneath (Kite and Rubin 2016; Spencer et al. 2018). The water surface is expected to be situated much closer to the surface than the general ice–ocean interface (at about 90 % of the distance from the ocean to the surface), so that it seems well possible that liquid water can be encountered at a depth of only a few hundred meters. The width of the water-filled part of the cracks is unknown but is estimated in the order of 1 m (Kite and Rubin 2016; Spencer et al. 2018). A broad range of surface temperatures are observed on Enceladus. They vary from 33 to 85 K depending on insolation, are 120 to 160 K at hot spots, and peak at about 200 K in the so-called tiger-stripe regions (Dalton 2010; Goguen et al. 2013). On the bottom of Europa’s and Enceladus’ oceans, environments similar to the hydrothermal vent environments at the bottom of the terrestrial oceans can be seriously considered (Sekine et al. 2015), although the ability to support life, of course, does not guarantee its origin or presence (Gaidos et al. 1999; Cockell 2014).

Answering the question of whether life has arisen on Europa and/or Enceladus will require further exploration. However, an intrinsic problem in the exploration of their ice crusts and the oceans beneath is the penetration of the ice with a device suitable for a planetary exploration mission. While short-range ice melting probes would allow a first exploration of near-surface ice, long-range ice melting probes may someday be used to penetrate more deeply and release autonomous underwater vehicles to explore the vast oceans underneath (Dachwald et al. 2020).

29.2.3 *Earth*

The diversity of life found by exploring extreme environments on Earth can guide our imagination of what we may expect in extraterrestrial ice sheets and oceans. So far, more than 400 lakes have been detected deep underneath the Antarctic ice sheet (Siegert et al. 2016). They have been isolated from the atmosphere for up to millions of years (Priscu et al. 1999). These lakes are of tremendous scientific interest as a unique habitat for extremophilic life that has persisted and evolved in such cold and secluded environments, whose conditions are considered similar to the conditions that may exist in extraterrestrial cases (Marion et al. 2002).

Any investigation of subglacial lakes needs to guarantee minimal contamination with microbial life from the surface and any potentially toxic material. Thus, conventional drilling appears to be problematic. For their clean investigation, it would be necessary to use methods that satisfy the strict requirements of planetary protection,

as applied, for example, for the exploration of Mars, and foreseen for any future mission to the icy moons (see Sect. 29.4.8) by the Committee on Planetary Protection Standards for Icy Bodies in the Outer Solar System (2012). A sterilized ice melting probe may be the optimal solution for the clean in situ investigation of Antarctica's subglacial lakes with minimal contamination risk. In 2014, a sterilized ice melting probe was used to take a clean sample from an englacial water conduit at Blood Falls, McMurdo Dry Valleys, Antarctica, an outflow of a subglacial brine pool of unknown depth, trapped approximately 400 m underneath Taylor Glacier, providing unique access to a subglacial ecosystem (Mikucki et al. 2009; Mikucki and Priscu 2007; Campen et al. 2019; German et al. 2019; Dachwald et al. 2014; Kowalski et al. 2016). Such terrestrial deployments of ice melting probes also provides an excellent opportunity for analogue planetary research and for testing the technologies required for extraterrestrial applications.

29.2.4 *Other Solar System Targets for Ice Melting Probes*

Although Europa and Enceladus (together with Mars) appear to be the most promising candidates for applying an ice melting probe for the subsurface exploration of ice layers and the oceans beneath, there are other icy moons where melting technology may be considered, especially for the exploration of resources. Ganymede and Callisto in the Jovian system as well as Titan, the largest moon of Saturn, are expected to have subglacial global oceans, similar to Europa and Enceladus, but sandwiched between a much thicker upper ice layer and a lower high-pressure ice layer (Nimmo and Pappalardo 2016). Titan may be a special case for applying melting probe technology, where—in addition to the surface lakes—*solid* hydrocarbons have been identified on the surface (Clark et al. 2010), which, in principle, could be evaporated and penetrated by melting. Another application for ice melting probes could include small bodies and, in principle, even comets. However, this would be associated with unique challenges like very “dirty” and porous “ice” and negligible gravity. Consequently, a melting probe which needs a certain contact-force-inducing weight to operate efficiently would most probably have to be combined with an actuator system that guarantees this contact force independently of its weight. One example would be to use an ice screw, as demonstrated for the IceMole (Dachwald et al. 2014), to secure contact to the ice under low gravity or a combined melting/hammering mechanism to pass through layers with a high concentration of solids.

29.3 Ice Melting Probe Theory

An idealized engineering model for thermal ice melting probes is derived from the energy balance and reads

$$P_h = \underbrace{VA\rho_i L}_{\dot{Q}_{\text{latent}}} + \underbrace{VA\rho_i C_{p,i}(T_m - T_i)}_{\dot{Q}_{\text{sensible}}}, \quad (29.1)$$

in which P denotes the heating power required to operate a probe of cross-section A at constant melting velocity V . It is given by the sum of heat flow rate \dot{Q}_{latent} that accounts for melting the ice in the vicinity of the probe and $\dot{Q}_{\text{sensible}}$ that increases the ice's temperature towards the melting point. Furthermore, L denotes the latent heat of melting, ρ_i stands for the ice density, $C_{p,i}$ is the ice's heat capacity and T_i denotes the ice's far-field temperature. Note that this consideration assumes a physical regime that sustains melting rather than sublimation. This may be justified as it is believed that the pressure within the melt channel will be above the triple point for most of the probe's transit through the icy shell. The probe's dynamic in a sublimation regime is still not fully theoretically understood to date, although some valuable considerations can be found in Li et al. (2020). In the following, we will restrict ourselves to the melting regime with a pressure above the triple point of water. In that case, we can use the expression in Eq. (29.1) to approximate the melting velocity based on the surface heat flux $\dot{q} = P_h/A$:

$$V = \frac{\dot{q}}{\rho_i \underbrace{(L + C_{p,i}(T_m - T_i))}_{L^*}}, \quad (29.2)$$

in which L^* refers to the *reduced* latent heat of melting. Note that the surface heat flux scales inversely with the cross-sectional area A of the probe, and so does melting velocity.

The power requirement resulting from the idealized engineering model (29.1) is to be understood as its necessary lower limit P_{min} . Additional losses P_{loss} are to be expected for specific probe designs and in any realistic deployment scenario, for example, due to electrical energy conversion or because not all the heat provided can be leveraged for forward motion. A first theoretical analysis of the performance of cylindrical melting probes is credited to Aamot and dates back to 1967 (Aamot 1967a). More recently, differently flavored modeling and simulation efforts have started to account for the probe's complex geometry (see, e.g., Talalay et al. 2019), and/or for the extreme ambient conditions (low temperature, pressure, and gravitational acceleration) the probe will face on extraterrestrial targets such as Mars, the icy moons, or comets (see, e.g., Schüller and Kowalski 2019). In general, two types of modeling approaches can be distinguished:

1. **Efficiency, velocity, and trajectory models** that approximate the melting probe's response to its input power for a given ambient ice condition. These models are formulated in a reduced or mixed dimensional setting and are hence computationally efficient and robust. They can be used for offline strategic mission preparation (many simulation runs required) or online operational decision-making conducted on board (limited computational resources available).
2. **High-fidelity models** aimed at providing a holistic picture of the physical processes around the ice melting probe, for example, phase change, melt water flow, and heat transport, as well as their interplay with the probe's forward motion. This requires advanced numerical techniques to solve large nonlinear partial differential equation systems on moving meshes at high spatiotemporal resolution, resulting in significant model development lead time and computationally resource-intensive codes. Once set up, they can be used for design optimization, for example, regarding the probe's geometry or sensor implementation, in order to improve its performance or to maximize the mission's knowledge return.

In this chapter, we will restrict ourselves to theoretical efficiency, velocity and trajectory models.

29.3.1 Efficiency Models

Efficiency models quantify dominant losses $P_{\text{loss}} = P_h - P_{\text{min}}$ of the melting probe system. In general, these losses depend on the specific probe design (geometry, structure, surface materials) and on the ambient physical regime and ice conditions (ice temperature and composition, ambient pressure, or local gravitational acceleration). Knowing the losses, we can compute the efficiency η according to

$$\eta = \frac{P_{\text{min}}}{P_{\text{min}} + P_{\text{loss}}} = \frac{P_{\text{min}}}{P_h}, \quad (29.3)$$

in which P_{min} stands for the minimum required power if the probe is operated at 100% efficiency as given by (29.1). The ratio hence denotes how much of the overall heating power P_h is leveraged for the probe's forward motion.

Losses due to the **lateral conduction of heat** into the ice were first assessed in Aamot's original performance analysis (Aamot 1967b) for cylindrical melting probes and carried over to extraterrestrial conditions by Ulamec et al. (2007). These models rely on the idealized assumption of a direct thermal coupling of the melting probe into the ice. Figure 29.1 displays corresponding efficiencies and total power requirements for terrestrial conditions found at Dome C in Antarctica, and for three extraterrestrial conditions: Mars polar, Enceladus close to its south pole, and Europa. A summary of site-specific conditions is provided in Table 29.2.

Ice density $\rho_i(T)$ and specific heat capacity of the ice $C_{p,i}(T)$ are parameterized as functions of temperature according to Ulamec et al. (2007). In Fig. 29.1, both

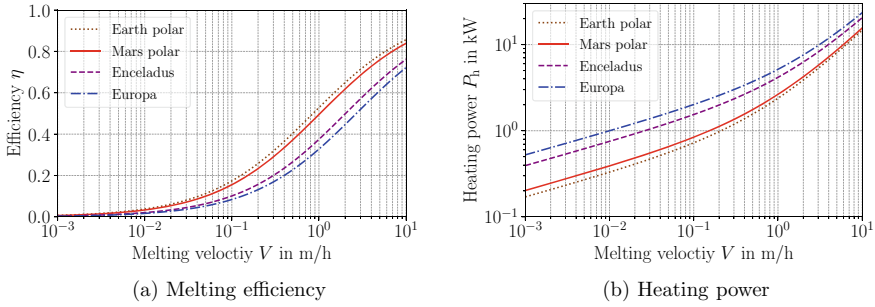


Fig. 29.1 **a** Melting efficiency η and **b** required heating power P_h as a function of nominal melting velocity following Ulamec et al. (2007). Site and material characteristics are summarized in Table 29.2

Table 29.2 Ice material properties and model parameters as used for the performance and trajectory models shown in Figs. 29.1, 29.2, and 29.3. Note that T_s stands for a characteristic ice surface temperature

Parameter	Earth	Mars	Europa	Enceladus
g in $\frac{m}{s^2}$	9.81	3.72	1.31	0.113
T_s in K	220 ^b	210 ^c	109 ^d	150 ^e
T_m in K	273.15 ^f			
L in $\frac{J}{kg}$	333430 ^f			
$\kappa_w(T_m)$ in $\frac{W}{mK}$	0.55567 ^g			
$\rho_w(T_m)$ in $\frac{kg}{m^3}$	999.84 ^g			
$C_{p,w}(T_m)$ in $\frac{J}{kgK}$	4219.4 ^g			
$\mu_w(T_m)$ in Pa s	0.00177 ^h			

^aFrom Table 29.1 ^bPolar, see Clarke et al. (2013)
^cPolar, see Ulamec et al. (2007) ^dDalton (2010)
^eKonstantinidis et al. (2015) ^fFeistel and Wagner (2006)
^gHaynes et al. (2017) ^hCalculated based on parametrization provided in Ulamec et al. (2007)

have been evaluated at the mean temperature $(T_m + T_s) / 2$. Note that efficiencies are plotted over *nominal velocities*. For a decreasing ice temperature, the melting velocity hence reduces due to a higher nominal power demand and decreasing efficiency. For a 2 kW melting probe of 0.06 m radius, the melting velocity, for example, results in approximately 0.96 $\frac{m}{h}$ for Earth polar conditions ($\eta = 61\%$) and 0.54 $\frac{m}{h}$ on Europa ($\eta = 38\%$), with site and material properties again taken from Table 29.2.

A slightly different approach is chosen when accounting for the fact that thermal coupling into the ice is in fact indirect, as the melting probe is surrounded by melt water. The melt water conducts a certain portion of the heat through the micro-scale melt channel (forward motion), while another part is **lost due to melt water convection around the probe**. See Schüller et al. (2016) and Schüller and Kowalski (2017) for a detailed investigation. Because of different assumptions (direct versus indirect thermal coupling into the ice), the two approaches cannot be readily com-

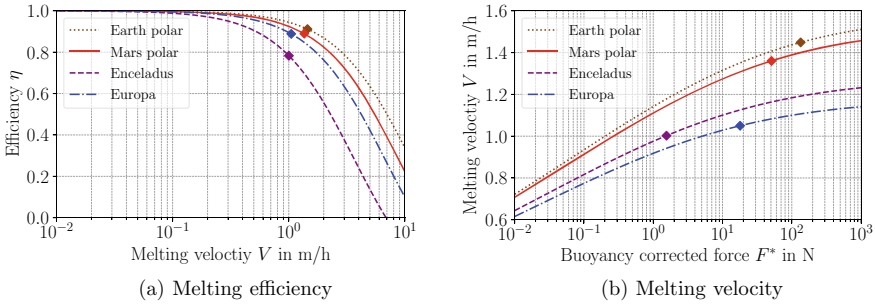


Fig. 29.2 **a** Melting efficiency η as a function of melting velocity V considering convective losses following (Schüller and Kowalski 2019) plotted for different ice environments and **b** melting velocity as a function of the buoyancy-corrected force for probe of 1 m length, 0.06 m radius and weight of 25 kg with heating power of 2 kW. The melting velocity is plotted for different ice environments. The diamonds mark the velocity and buoyancy-corrected force at the target. Site and material properties are summarized in Table 29.2

bined, although research is underway to integrate both approaches in a combined theory. Still it is quite informative to look at the idealized situation of convective losses only as displayed in Fig. 29.2.

In contrast to conductive losses being lowest at high velocities (Fig. 29.1a), we now see that convective losses in the melt film are lowest at very low melting velocities (Fig. 29.2a). Furthermore, an interesting observation is that convective losses in the melt film are sensitive to a change in the gravitational acceleration, as the contact force of the probe determines how strong the melt water is squeezed out of the melt film. In Fig. 29.2b, we see that, due to a higher gravitational acceleration, the expected melting velocity in colder ice on Europa is higher than on warmer ice expected at Enceladus’ south pole.

29.3.2 Critical Refreezing Length

Assuming a perfect lateral isolation of the melting probe, we can use the previously summarized models to assess the extent to which convective losses by means of heat losses in the melt film can be leveraged as deliberate conductive heat losses at the side of the probe to prevent lateral refreezing. This gives rise to the notion of the so-called critical refreezing length, first introduced by Schüller and Kowalski (2019) for cylindrical melting probes. It denotes the position along the melting probe’s side at which the initially warm melting water has reached the melting temperature. This position corresponds to the minimum distance from the melting head at which side wall heaters would have to be implemented to avoid a refreezing-induced stalling of the probe. The impact of various parameters, such as probe diameter, length, and penetration length, on the critical refreezing length can be found in Li et al. (2020).

29.3.3 Velocity Models

Efficiency models can be reformulated as velocity models for a certain melting probe design and a specific ambient ice environment.

$$V = V(g, T_i, \rho_i(T_i), \dots). \quad (29.4)$$

Following Schüller and Kowalski (2019), the competing effects of gravitational acceleration and ice temperatures on the melting velocity subject to convective performance losses for a standardized melting probe design with 2 kW heating power are displayed in Fig. 29.2b. Site specifics and parameter values are chosen as before, see Table 29.2 and the previous parameterization. The competing effect between gravitational acceleration and temperature becomes evident when comparing expected velocities on Enceladus' south pole and on Europa. Though melting into colder ice on Europa, the higher gravitational acceleration compensates for convective losses, resulting in a higher effective velocity. Note that current-day computational models are restricted to idealized geometric designs (cylindrical probes).

29.3.4 Trajectory Models

Knowing the probe's melting velocity allows us to integrate the local velocity model into a global trajectory model according to

$$x(t) = \int_0^t V(\tau, T_i(x(\tau)), C_{p,i}(T(x(\tau))), \dots) d\tau. \quad (29.5)$$

This enables us to track the melting velocity along a vertical temperature gradient throughout the ice layer, as shown in Fig. 29.3 for three melting probes at 1, 2, and 5 kW for conditions on Earth (Dome C, Antarctica) and on Europa. Here, we assume an idealized linear temperature gradient. In all six trajectories, but most obviously for the 5 kW probe on Europa, we can observe how the melting velocity increases when approaching its lower boundary due to the higher background temperature. Trajectory models can be evaluated into global mission indicators such as transit time as well as total and average energy demand for a given probe design and ice thickness, as summarized in Table 29.3.

29.4 Challenges and Key Technologies

In Sect. 29.2, various applications for ice melting probes were identified and described. Missions to icy moons clearly pose the greatest challenges for ice melting probes, whereas, although still difficult, melting on Mars is far less demanding.

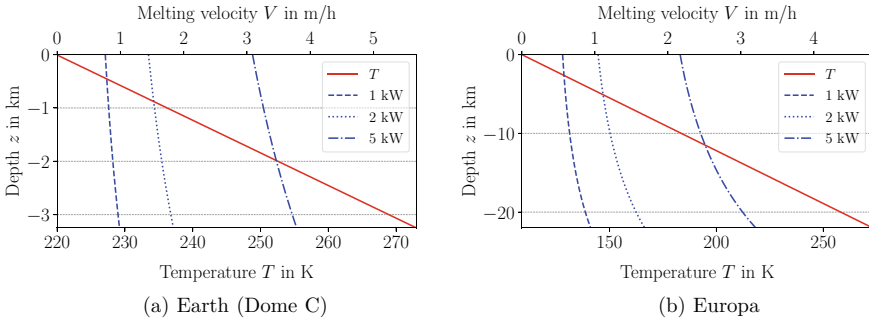


Fig. 29.3 Trajectory curves for two different ice environments and heat flow rate at the melting head of 1 kW, 2 kW, and 5 kW. The linear temperature profile and the melting velocities for the different heat flow rates are shown as a function of depth

Table 29.3 Results of trajectory calculations for the Earth and Europa scenarios shown in Fig. 29.3

Heat flow rate in kW	Earth (Dome C)			Europa		
	1	2	5	1	2	5
Ice thickness in m	3250 ^a			22000 ^b		
Transit time in d	161.9	84.1	40.7	1351.9	733.3	363.8
Average energy demand in $\frac{\text{kWh}}{\text{m}}$	1.197	1.244	1.505	1.475	1.600	1.984
Total energy demand in MW h	3.886	4.039	4.889	32.444	35.200	43.653

^aClarke et al. (2013) ^bHowell (2020)

Depending on the target body and the objective of a mission, some of the challenges mentioned in this section may not be as severe as on icy moons or may not even exist. A unique challenge for ice melting probes is given on comets: melting in very “dirty” ice that contains a high fraction of dust particles and under almost zero gravity. Space exploration missions with ice melting probes require some general goal-independent and design-independent key technologies. Depending on the desired ice penetration depth, considerable up to immense amounts of energy are required (see Table 29.3). But missions to the outer solar system, and, to a lesser degree, already to Mars, suffer from a scarcity of solar power. Furthermore, when a probe is in the ice, commands and data must be exchanged with the surface station to be able to control the probe and to obtain data. The structure of the ice is unknown and may include solids and cracks. Communicating through dozens of kilometers

of such ice, as would be required for the penetration of the thick ice shell of an icy moon, is a considerable challenge. Another inherent challenge for ice melting probe application is the refreezing of the melting channel due to either sublimation of water (or CO_2) vapor or refreezing of melting water. This can have advantages (e.g., increased local pressure in a closed cavity) but also includes the danger of blocking the probe and making any retrieval challenging. Nevertheless, some design proposals include retrieval and upward melting, for example, by using a thermal head at the top end (RECAS, Talalay et al. (2013)) or turning by 180° after sampling of the ocean water (VALKYRIE, Stone et al. (2018)). Yet another challenge is navigation and steering in the unknown ice, given that external references like stars are not available but only the absolute position and orientation of the surface station. This requires a large degree of autonomy. At least on Europa, the radiation at the surface from the high-energy particles trapped in Jupiter's strong magnetic field is immense and requires intense shielding for the electronic components. It also renders solar power systems infeasible, so a nuclear power source is required. Landers and hardware that penetrate the surface require particular methods for cleaning and sterilization so that the surface and the ice are not contaminated by terrestrial life and the planetary protection requirements are fulfilled. The ΔV to land on an icy moon is very large. Therefore, every landed kilogram multiplies the mass required in Earth orbit at departure, depending on the selected propulsion system and transfer trajectory. This fact makes miniaturization especially important.

29.4.1 Unknown Environments and Mission-Critical Hazards

In ice layers, cracks may be present near the surface. If an ice melting probe penetrates the wall of such a crack, it may be canted or even fall uncontrollably into the crack, thereby breaking the tether or cable that is required for power transmission and communication. Also, it may become incapable of re-initiating forward melting. Therefore, an ice melting probe should be able to navigate around cracks, which requires the capability of forefield sensing, navigation (see Sect. 29.4.4), and steerability. Also, meteorites or dust accumulations could be present in the ice, which should be able to be circumvented. Furthermore, closer to the base of the ice layer, there may be liquid-filled pockets and cavities that may set an unexpected end to an ice melting mission. While such obstacles may be rare, they would probably block the motion and end the entire mission prematurely. Therefore, it would be preferable to prepare a multi-billion-dollar mission for such a situation. Even if an ice melting probe is able to avoid driving into obstacles, discontinuities and ice inclusions affect the ability for communication (see Sect. 29.4.3). Even for routine ice melting operations, there are many unknowns in the structure and parameters of extraterrestrial ices. Lacking a dense atmosphere, the near-surface ice layers of icy moons are persistently disrupted by impact gardening due to hypervelocity collisions with minor bodies and meteoroids. Owing to the low surface temperature and low gravity on small icy bodies, the brittle zone is expected to retain substantial porosity in excess

of a value of 0.1 (Lee et al. 2005). This is significant and will strongly affect thermal diffusivity, tensile strength, acoustic and dielectric properties, vapor transport, and other key properties relevant for ice-penetrating probes (Moore et al. 2009b). The mechanical and thermophysical properties of ice are further affected by the incorporation of non-ice contaminants due to the delivery of micrometeoroids, the deposition of salty plume particles (on Enceladus), partial melting, or volatile release and upward percolation of cryomagmas (McCord et al. 1998). Small temperature variations are sufficient to promote clathrate decomposition and subsequent release of trapped gases. The presence of salts or ammonia in subsurface liquid reservoirs could cause significant melting point reduction (Kargel 1998), hence increasing viscosity at the base of the icy shell and shifting the brittle–ductile transition to greater depths (Durham et al. 2010). Figure 29.4 schematically shows the ice variability with depth across Europa’s ice shell.

29.4.2 Power

Ice melting probes have high power demand (see the analysis of efficiency and minimum power required in Sect. 29.3), and the power needs to be available at the tip of the probe, typically far below the surface. Therefore, one of the major challenges for the application of ice melting probes is the provision and distribution of power, especially in the outer solar system, but to a lesser degree already on the ice caps of Mars.

The solar radiation flux at Europa is less than 4 %, and that at Enceladus only about 1 %, relative to the value at Earth. Although spacecraft with large solar arrays are feasible at Jupiter (e.g., JUNO, JUICE, Europa Clipper), it is almost impossible to think of a technical design on the surface to provide sufficient power to operate a lander base station for an ice melting probe. For solar cells at Europa, the high radiation flux at the surface (see Sect. 29.4.7) would also lead to a very rapid degradation. Thus, at least for icy moon missions, solar power cannot be considered as a primary power source. Therefore, radioisotope thermal generators (RTGs) or fission reactors must be considered as primary power sources.

Several types of RTGs have been used for deep space missions (Fraser 2018). A new RTG generation, the multi-mission RTG (MMRTG), produces about 2 kW of thermal power and 120 W of electric power at beginning of life. It contains about 4.8 kg of $^{238}\text{PuO}_2$ and has a mass of 45 kg (Hammel et al. 2009). It is thus hardly conceivable to power a melting probe electrically via a harness, with MMRTGs aboard the surface station. Even using the heat directly generated by the decay of ^{238}Pu will not be sufficient to operate an ice melting probe, keeping in mind a minimum requirement of protective housing for the pellets and the resulting power per volume. Note that a ^{238}Pu heat source placed on the surface of Europa (or Enceladus), purposely or by accident, would not melt into the ice (Lorenz 2012). Therefore, an adapted denser packaging concept for the $^{238}\text{PuO}_2$ would be required for an ice melting probe. A power level of 2 kW thermal and 110 W electric is also assumed for the concepts and

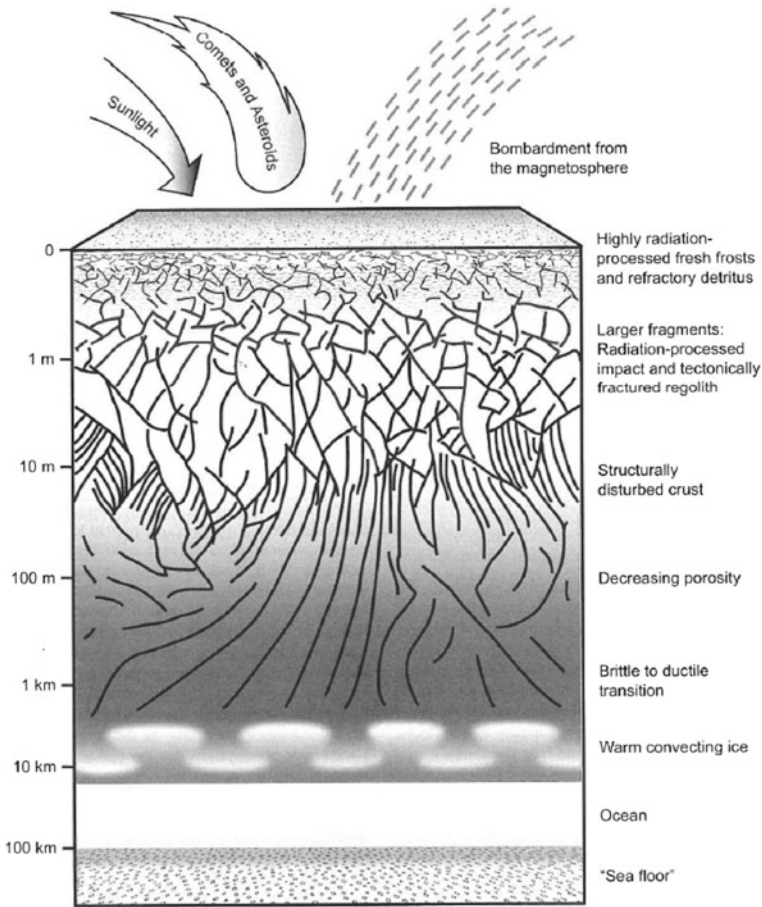


Fig. 29.4 Schematic cross-section of Europa’s outer ice shell with a logarithmic vertical scale (adapted from Moore et al. 2009a)

technologies developed within NASA’s SESAME program (NASA Glenn Research Center 2019). The use of radioisotopes other than plutonium, with higher specific power due to a shorter half-life, may be considered. Table 29.4 gives an overview of some radioactive isotopes that may be considered for space missions (Ulamec et al. 2010).

So far, only ^{210}Po -based (Lunokhod) and ^{238}Pu -based RTGs have spaceflight heritage. ^{90}Sr is used for some terrestrial RTG applications (e.g., for remote light-houses) but requires considerably more shielding (factor 60) than, for example, ^{238}Pu , due to the beta radiation and the resulting bremsstrahlung. Also, the daughter product, ^{90}Y , is a strong beta emitter. ^{90}Sr has been suggested, for example, in the frame of the Very deep Autonomous Laser-powered Kilowatt-class Yo-yoing Robotic Ice Explorer (VALKYRIE) proposal (Stone et al. 2018). ^{210}Po , despite its high specific

Table 29.4 Available radioisotopes for RTGs

Isotope	Half life in years	Radiation	Specific thermal power in $\frac{W}{g}$	
			Begin of life	After 6 yrs
¹²¹ Po	0.37	$\alpha, (\gamma)$	141.0	0.00
²³⁸ Pu	87.7	$\alpha, (\gamma)$	0.56	0.52
⁹⁰ Sr	28.8	β	0.93	0.76
⁸⁵ Kr	10.8	$\beta, (\gamma)$	0.62	0.36
⁶⁰ Co	5.27	β, γ	17.7	5.7
²⁴⁴ Cm	18.1	$\alpha, (\gamma)$	2.83	2.03
²⁴¹ Am	432.2	α, γ	0.11	0.11

power at beginning of life, is not favorable for a mission to the outer solar system, where the transfer alone takes many years, due to its short half-life. Therefore, RTGs are also hardly usable for powering extraterrestrial ice melting probes.

Another solution, which would indeed provide sufficient power, could be to place a nuclear reactor on the surface, close to the melting location. Obviously, the transport and operations of such infrastructure has significant challenges. Reactors do have space heritage, however only in Earth orbit (El-Genk 2009). In the former Soviet Union, “BUK” power systems were used for a total of 31 satellites (RORSATs) from 1970 to 1988, and two TOPAZ reactors were used for two Cosmos missions in 1987. In the USA, the SNAP-10A system was launched in 1965 (El-Genk 2009). The SNAP-10A reactor was intended to produce 500 W of electric power for one year (but had to be shut down after 43 days for reasons not related to the reactor). The TOPAZ system (Fig. 29.5) was designed to provide electric power of about 5.5 kW for a lifetime of up to one year. The mass of the reactor was about 320 kg and the complete system had a mass of about 980 kg. The Soviet YENISEI (TOPAZ II) reactor, designed for longer lifetimes, has never been flown.

The Soviet program for developing reactors for space applications was stopped in 1988; however, there has been renewed interest in this technology in recent years. The European Commission, for example, was funding the DEMOKRITOS project, preparing for a megawatt-class flagship mission with nuclear electric propulsion (Jansen et al. 2016). The Keldysh Research Center in Russia, in cooperation with Roscosmos, Rosatom, and the Russian Academy of Sciences, manages the transportation module as a nuclear power and propulsion system (Koroteev et al. 2013). Activities at NASA to develop a 1 – 10 kW (electric) reactor for space applications are described, for example, by Gibson et al. (2017). Being far from 100 % in efficiency, high-power nuclear reactors in space require large radiators to dissipate waste heat. The regulation of the reactor and the generation of electric power by a fully autonomous system is demanding. Nevertheless, such systems can in principle be realized and will probably become relevant when a permanent lunar base or human missions to Mars are realized. For the concepts and technologies developed within NASA’s SESAME program (NASA Glenn Research Center 2019), it should



Fig. 29.5 Nuclear power plant for space applications “TOPAZ” (from Kraus and Shabalin 2013)

be assumed that NASA provides a fission reactor with 43 kW of thermal power and 420 W of electric power.

29.4.3 *Communication*

Establishing a communication link for telemetry and telecommand between a surface station and Earth is demanding by itself. For an icy moon mission, relay via an orbiter, either around the moon or in the Jovian or Saturnian system, is most likely the preferred solution. Here, we concentrate on the challenges in receiving telemetry from the probe within the ice.

Terrestrial ice melting probes can employ tethers to transmit power and data. Because of losses in the cable and mass/volume issues, such tethered systems are practically limited to relatively shallow exploration depths of the order of a few kilometers. There is also significant danger of tearing of the tether after the refreezing of the ice channel, due to tidal flexing.

For communication between an ice melting probe and a surface station, radio-wave communication has also been considered, for example, by Bryant (2002). The $1/e$ penetration depth of radio waves in pure ice is in the range of 100 m around -20°C and in the range of kilometers at -60°C (there are no measurement points at lower temperatures). However, attenuation is higher for impure ices (containing salts), and for greater distances, effects of cracks and/or dust layers will lead to signal refraction and weakening. Bryant proposed the use of a number (7–35 to reach 10 km depth) of relay transceiver pods placed between the ice probe and the surface

(but did not provide a detailed discussion on key technical challenges like deployment, power supply, and thermal control).

It appears at the moment, despite a number of technical issues (mass and volume, tether storage, deployment, electrical losses, durability, danger of breaking the tether, e.g., due to tidal stresses), that the use of a tether (including a coaxial cable and/or optical fibers) is the most promising solution. Work to assess the usability of tethers for communication and to increase their technological readiness level has already been started by McCarthy et al. (2019).

29.4.4 Navigation and Autonomy

Navigation is required for deep subsurface ice exploration (Kowalski et al. 2016). Navigation in the ice is very challenging, given that no natural external references exist other than the local gravity vector and the magnetic field vector at the surface. An ice melting probe that is connected to the surface via a tether could measure the propagation path, and thus, when combined with a tilt-measuring accelerometer, derive the depth from the surface. An inertial measurement unit (IMU) that measures the orientation of the probe could provide an additional navigation aid within an ice melting probe. If the attitude and the velocity over time are known, the trajectory of the probe can be determined (or, more precisely, estimated). The disadvantage of IMUs, however, is their drift on relatively short timescales, so that regular updates from external references are required. One potential artificial external reference concept that would allow relative navigation between two (or more) elements of an under-ice system is the deployment of an acoustic (or electromagnetic) transponder network that enables the estimation of positions via trilateration. Due to attenuation, however, the range of signals in ice is limited (see Sect. 29.4.3). Attenuation and wave propagation speed are strongly affected by the currently unknown salinity and particle content, and impurities such as cracks and bubbles in the ice are additional factors. Therefore, unless the acoustic system is self-calibrating, navigation must be based on estimated values and is consequently less accurate. The variable magnetic field can only be used as an additional relative external reference with respect to a surface station, if the latter also carries a magnetometer. A key technology for navigation, but also other technical aspects, is autonomy. Autonomy, however, is a broad field on its own, and a sufficiently deep discussion would go beyond the scope of this chapter. Also, in contrast to the other key technologies of this section, there is a substantial terrestrial market for autonomous systems, and space applications will be able to profit from them (Dachwald et al. 2020).

29.4.5 *Miniaturization*

A general concern for space missions is minimizing the mass and volume of systems and instruments, because the transport of every kilogram to an icy moon requires dozens of kilograms of propellant and structural mass. It must also be considered, however, that the laws of physics and the principles of technology set limits for the miniaturization of systems and instruments. An ice melting probe of a given size requires minimal power for efficient melting by minimizing conductive losses into the surrounding ice (i.e., melting a small volume of ice in front of the probe instead of just heating the surrounding large volume of ice). An extremely small “point-size-like” ice melting probe would still face the issue of losses, but requires an unrealistically high power density (note that losses by conduction are proportional to the surface area of the probe, while required power density is proportional to the volume of the heater element). Also, electric power supply via a cable requires a certain thickness to avoid high resistivity losses, so that very thin cables or small tether containers are not advisable. Therefore, ice melting probes with a diameter of less than approximately 5 cm would become very inefficient. For the instrumentation inside a probe, however, the situation is different, and it is indeed advantageous to find ways to integrate small and lightweight devices. The same applies for communications devices or sensors for navigation and attitude control.

29.4.6 *Pressure Resistance*

The subsurface pressure p for a celestial body under the ice or liquid can be easily calculated by applying the formula $p = \rho g d$ (with ρ being the ice density, g the acceleration of gravity and d the depth). In order to obtain an estimate for the design of an ice melting probe, it is sufficiently accurate to assume a constant $\rho \approx 920 \frac{\text{kg}}{\text{m}^3}$. For an exact calculation of the pressure, the effect of density variation due to composition (e.g., salinity or dust inclusions in the ice), crystalline structure of the ice, and compressibility needs to be estimated. The variation in g with depth can be ignored in this context (e.g., for Enceladus, in a depth of 1 km, g is reduced by only about 0.8 %). So, for example, at an ice depth of 1 km, the pressure on Europa is about 1.2 MPa, on Enceladus about 0.1 MPa, and on Earth about 9.0 MPa. An additional aspect that must be considered is the pressure transient when reaching the ice–water interface. Although this pressure jump should not occur if ice and water are both in hydrostatic equilibrium, it was observed on Earth by Tüg (Ulamec et al. 2007). Aamot (1968) also reported measurements of pressure jumps due to freezing or melting of surrounding material with associated volume and pressure changes. There is considerable experience in handling pressures in the range of tens of megapascals applied to submarines or offshore equipment, but generally, the design of probes that can withstand pressures in the range of several megapascals is not trivial and requires considerable effort to stiffen the hull of the probe. The alternative to a thick hull pro-

tecting a “dry” interior is to have the probe internally pressurized, for example, by filling it with melt water or silicone oil. Such a “wet” architecture has been used, for example, in the Philberth and AWI (Alfred Wegener Institute) probes for the tether storage canister section of the probe (Aamot 1967c; Tüg 2003).

29.4.7 Radiation Hardness

Europa is orbiting within Jupiter’s intense radiation belts, thus resulting in a severe radiation environment, depending on the exact landing location. As can be seen in Fig. 29.6, however, radiation rapidly decreases with depth. Nordheim et al. (2018) analyzed the dose for energetic particles for various depths of water and different locations on Europa in the context of potential biosignatures. While the dose at the surface is in the range of $100 \frac{\text{kGy}}{\text{d}}$ (equal to 10000 krad/d), this enormous value decreases by about eight orders of magnitude below 1 m of water/ice. But at least before being covered with ice, any component aboard a Europa spacecraft will be exposed to high levels of radiation, thus requiring extensive radiation hardness, shielding, and fault protection. If an internal radioisotope power source is used, spacecraft com-

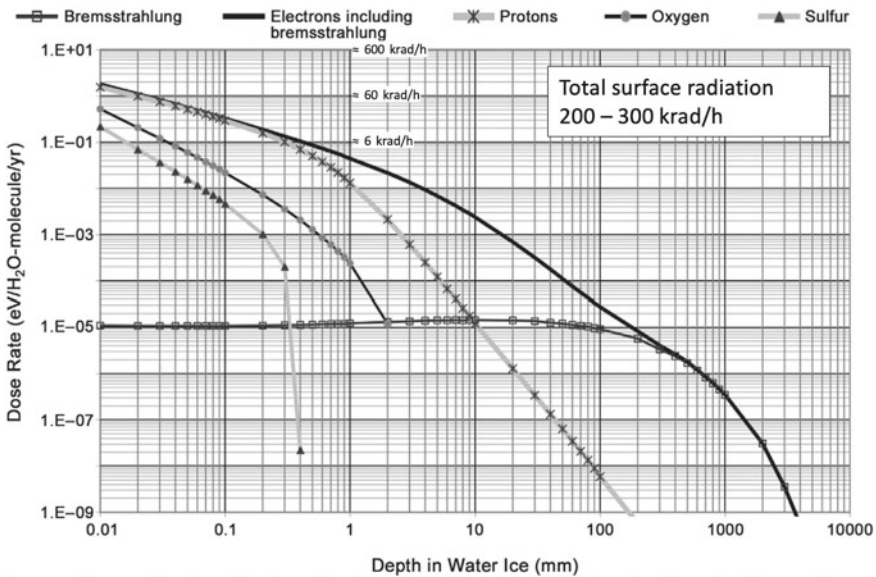


Fig. 29.6 Predicted dose rate versus depth at the apex of Europa’s trailing hemisphere by charged species using an input spectrum for various species and an ice surface. Heavy ions are stopped almost immediately in the water ice. At large depths, the electron dose rate becomes dominated by the contribution from secondaries (bremsstrahlung), adapted from Paranicas et al. (2009). 200 – 300 krad/h equal $2 - 3 \frac{\text{kGy}}{\text{h}}$

ponents also have to be shielded from its radiation. Since the magnetic moment of Saturn is about 60 times smaller than that of Jupiter, radiation at Enceladus is a less severe problem than at Europa. On the surface of Enceladus, the radiation dose is about three orders of magnitude lower than on Europa (derived from Nordheim et al. 2018). While there is some literature on the scientific aspects of the radiation environment at the icy moons, there is considerably less available literature on the engineering implications. The issue of coping with the harsh radiation environment, however, is addressed in relation to the development of NASA's Europa Clipper mission (Phillips and Pappalardo 2014) and ESA's JUICE mission (Grasset et al. 2013), which both plan for multiple Europa flybys, as well as NASA's Europa Lander study (Hand et al. 2017).

29.4.8 Cleanliness and Sterilization

In situ investigation of extraterrestrial ices in order to search for organics, potential biomarkers, or even signs of past or present life requires the cleanliness and sterilization of space hardware in the frame of strict contamination control. The Committee on Space Research (COSPAR) promotes a planetary protection policy (Kminek et al. 2017) as an international standard to guide compliance with the Outer Space Treaty (United Nations Office for Outer Space Affairs 1966). For the exploration of Antarctic subglacial aquatic environments, the Scientific Committee on Antarctic Research (SCAR) has also formulated a code of conduct (Siegert and Kennicutt 2018). On Earth, microorganisms are found almost everywhere, ranging from several kilometers up in the atmosphere to the deep oceanic and continental subsurface. Consequently, they are also present in spacecraft assembly cleanrooms. Therefore, the assembly, integration, and testing of space hardware according to planetary protection requirements must be carefully planned. The space hardware has to be cleaned and sterilized with accepted methods and the bioburden has to be measured with specific assays, for example, according to the standards developed by the European Cooperation for Space Standardization (ECSS). Because the COSPAR regulations do not define certain levels of organic contamination as acceptable, the environmental conditions and the evolution of contamination must be modeled over the whole mission lifetime to determine acceptable limits for contamination at launch.

29.4.9 Instrumentation

For the instrumentation of an ice melting probe, several options can be considered (see, e.g., Dachwald et al. 2020), because the probe's interior can easily be kept in the operational temperature range of standard hardware. Any instrumentation, however, must comply with the small volume and the shape of the probe. Also, constraints concerning mass, power, and data rate apply. Instruments can investigate meltwater

either in situ, by pumping it into the interior, or remotely, by a side-looking window in the hull of the probe. If such a window were in the melting head, it would require excellent thermal conductivity, because otherwise, it would create a cold spot that would prevent melting. For in situ analysis inside the probe, an ice screw, as was used in the IceMole design (Dachwald et al. 2014), can also ingest an unmelted ice sample if it is thermally isolated from the melting head.

29.5 Existing Melting Probe Designs and Tests

Simple ice melting probes were applied already in the 1940s, primarily to investigate glaciers without the need to use heavy equipment or large teams for drill operation. In 1948, for example, an electrically heated probe melted into the Jungfraufirn glacier in the Swiss Alps until it reached solid rock after two weeks of continuous operation at a depth of 137 m (Gerrard et al. 1952). The first ice melting devices mostly acted as a low-tech drill substitute where no suitable mechanical core drilling hardware was available. Consequently, the ice melting probes of a Japanese expedition in Antarctica were described as “far less effective than an electrodrill”, but “more easily made and more stable in operation” (Suzuki 1976). Nevertheless, soon after ice melting probes proved to be a valid alternative to mechanical ice coring drills, the technology was also used to carry onboard scientific payloads to subglacial locations. Given that the deployment of ice melting probes on extraterrestrial ice masses enables exceptional scientific investigation (Ulamec et al. 2007; Sherwood 2016), some notable examples in the development and evolution of ice melting probes are presented in this section. This section, however, can only provide a brief overview on the most important designs and achievements of ice melting probes to date. For a more detailed description, the reader is referred to the book by Talalay (2020). A comparison of some key characteristics of melting probes deployed under varying terrestrial ice conditions is shown in Table 29.5.

29.5.1 *First Ice Melting Probe Designs*

The first investigations concerning ice penetration with heated bodies were initiated by Karl Philberth. In 1962, he presented a concept to investigate the internal temperatures of Greenland’s ice sheet (Philberth 1962), working with his brother Bernhard, who investigated the feasibility of depositing nuclear waste inside the thick ice layers of Greenland or Antarctica. He identified the ice temperature and the thermal budget as important parameters when estimating the enclosure time before the waste reaches a coastline (Philberth 1959, 1961).

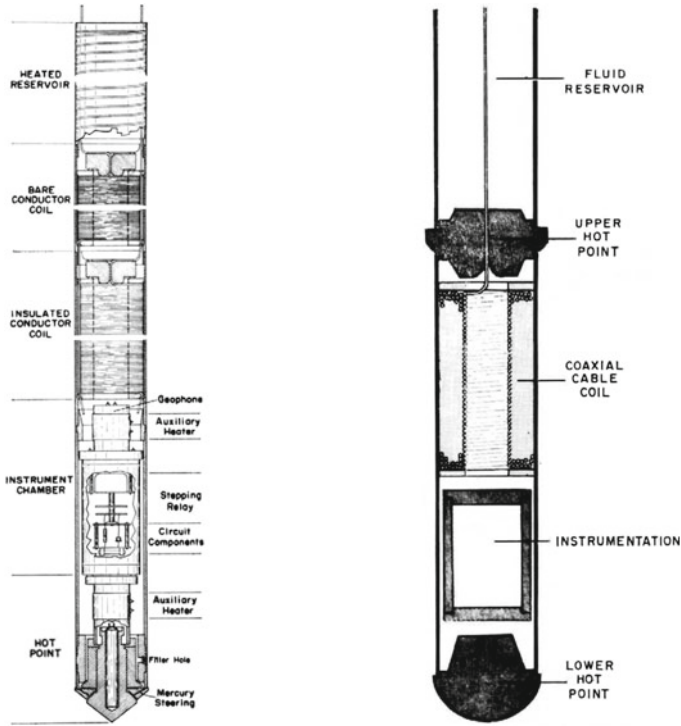
Karl Philbert designed an ice melting probe with an electrically heated melting head and a method for directional stabilization that he called mercury steering (Philberth 1964). In cooperation with the Cold Regions Research and Engineering

Table 29.5 Key characteristics of some ice melting probes deployed so far. The cross-section is given as the diameter for cylindrical probes (marked \varnothing) and as the edge length for probes with a square-shaped cross-section (marked \square). Only results from Earth-like tests are included here for easier comparability

Probe name or operator	Heating power in W	Length in m	Cross-section in mm	Demonstrated velocity in $\frac{m}{h}$	Sections
CRREL	3000–5000	2–5	$\approx \varnothing 110$	2.5	29.5.1
SUSI	3400	2.25	$\varnothing 100$	2.9	29.5.1
IWF I	150	0.12	$\square 40$	0.23	29.5.2.2
IWF II	85	0.125	$\varnothing 60$	0.15	29.5.2.2
EnEx-IceMole	2880	2	$\square 150$	2.2–4	29.5.2.3
Ice Diver	4500	1.5	$\varnothing 85$	6.6	29.5.2.4
VALKYRIE	5000	1.5	$\varnothing 250$	0.9	29.5.2.5

Laboratory (CRREL), multiple similar probes were developed, which all relied on the same basic concept. Figure 29.7a shows the first design of the Philberth probe, including the mercury steering method. While the stabilization method remained a special feature of Philberth’s design, the arrangement of different functional segments was essentially preserved in most subsequent designs—not limited to the CRREL probes. The Philberth probe uses a very intuitive segmentation with, of course, a melting head (or hot point) in front, followed by power electronics for the heater, control circuits, and eventually payload systems. A conductor stored on a spool is released from the rear containment and serves as an interface for power supply and control or data link, respectively. As two conductors are used and only one of them is insulated, it is also possible to estimate the length of cable deployed from the surface and, therefore, the probe’s depth. Like the majority of probes, the Philberth probes relied on an elongated cylindrical hull (109.2 mm in diameter at varying lengths above 2 m) and a conical tip. Most of the volume was necessary to accommodate the two conductor wires, each 3000 m in length. The conductor storage segment was not sealed off from the meltwater but was filled with a nonconductive oil to improve cable insulation. Because of its higher density than that of water, the oil would not be displaced by the surrounding meltwater.

At the same time, CRREL engineer Haldor W.C. Aamot investigated improvements in the probe design, as well as methods to estimate the penetration performance for a given geometry and heating power. Most notably, he proposed a simpler stabilization approach in 1967, called pendulum steering (Aamot 1967b, 1970), which is still used in much more recent probes (e.g., Ice Diver, see Sect. 29.5.2.4; Winebrenner et al. 2013): A heated thickening in the rear area of the probe (above its center of gravity) shifts its main support point away from the front melting head when the surface temperature of the rear heated area is lower than that of the head. Therefore, the probe no longer balances on its tip, but rather hangs from its rear suspension and consequently experiences effective directional stabilization. Further-



(a) General concept, maintaining directional stabilization with mercury steering. (b) Aamot's modified Philberth-like probe, implementing pendulum steering.

Fig. 29.7 Probes from CRREL (Aamot 1968)

more, Aamot derived a solution to estimate thermal losses during the operation of ice melting probes with a circular cross-section (Aamot 1967a). Although he tailored his calculations specifically for the Philberth probes, the same analytical ansatz is still commonly used today for the energy budget estimation (Ulamec et al. 2007). Several probes built at CRREL were deployed during field test campaigns. The first probe was deployed near Camp Century, Greenland, in 1965 and reached a depth of about 90 m, before an insulation failure caused a loss of contact. A second probe reached a depth of about 259 m after four days of continuous operation. Penetration rates of $2.5\text{--}5 \frac{\text{m}}{\text{h}}$ were observed for power levels of 5–15 kW and for ice temperatures of around -20°C (Aamot 1970).

In 1968, Philberth deployed two more probes with a diameter of 108 mm and length of 2.92 m and 2.55 m, respectively, into the 2500 m thick ice shield near Jarl-Joset station in central Greenland. Being stopped by an insulation fault, the first probe reached a depth of 218 m. The second probe arrived at 1005 m and initiated a scheduled measurement stop before it was also deactivated as a consequence of an electrical malfunction. Both probes were operated at a total probe power of 3.7 kW, resulting in a penetration rate of around $2.0 \frac{\text{m}}{\text{h}}$.

During the 1990s, a new series of probes, named Shuttle Under Ice Shelf (SUSI), was developed and tested by the German Alfred Wegener Institute for Polar and Marine Research (AWI). Both the design and performance of the probes were closely aligned with the ice melting probes tested earlier at CRREL. However, an active heater control loop relying on a rope-tension sensor was introduced for directional stabilization. The second prototype, SUSI II, reached a depth of 60 m in the Austrian Rettenbach Glacier in 1990 and 225 m during the 1992/1993 Antarctic field season in the shelf ice near the Neumayer Station. During the first mission, in temperate ice, the melting channel stayed unfrozen long enough to recover the probe. At the Neumayer Station, SUSI II accessed the ocean beneath the ice shelf. An improved design, called SUSI III, was planned to be tested in Antarctica as an addition to a larger hot water drilling campaign (Tibcken and Dimmler 1997). It was intended to melt through 700 m of ice in 9–10 days, followed by a measuring period of at least one year. Measurements were to have been performed using a modified commercial probe and were to include the temperature, pressure, salinity, current velocity, and current direction of the water below the ice shelf (Alfred Wegener Institute for Polar and Marine Research 1995, p. 23).

29.5.2 Notable Ice Melting Probe Designs for Specific Applications

While the early ice melting probe designs aimed at terrestrial use only, this section introduces some remarkable projects with a focus on new technology and performance under extraterrestrial conditions.

29.5.2.1 Cryobot

When first indications of liquid water under Europa’s thick ice crust were found in scientific data of the Galileo probe mission (Carr et al. 1998), the use of ice melting probes for extraterrestrial applications came increasingly into focus. The Cryobot was developed at the NASA Jet Propulsion Laboratory (JPL) in an attempt to benefit from the obvious advantages of ice melting probes over conventional drilling systems for space missions, which are constrained by tight power and mass budgets and by hard operational constraints. At the same time, the Cryobot avoids a major flaw of earlier ice melting probe designs, namely a highly stressed conductor wire for the probe’s power supply over distances of several kilometers, which are initially stored on compact spools (French et al. 2001). The obvious solution to this issue is to completely avoid a conductor connected to the surface. Instead, the Cryobot concept for extraterrestrial use proposes an integrated radioisotope thermoelectric generator (RTG) to generate electric and thermal power directly in the probe. The usually weak electrical efficiency of RTGs is not a disadvantage in this case, since “power losses”

in the form of heat can be directly transferred to the heater surfaces. In addition to this passive heating, the Cryobot possesses an active mode with hot water jets to improve both the ice penetration rate and its dirt layer penetration capabilities. Compared with previous probes, the thermal power (directly used for heating) is at a low 1 kW. Combined with a diameter of 120 mm, the resulting penetration rate is estimated to be $0.3 \frac{\text{m}}{\text{h}}$ in Europa's cold ice (Zimmerman et al. 2001). A data link was intended to be established by small transceiver relays, which were to be deployed from the probe's rear end and left behind in the refreezing melt channel. Some prototypes were built, albeit all relying on traditional tether-based energy supply. Rather conventional setups (in terms of power supply) were tested in ice near the melting point under laboratory conditions and on a glacier in Svalbard: Inside an artificial ice block with a height of approximately 5 m, the probe was operated in both active and passive mode. Heating with 240–536 W, the average penetration rate was $0.345 \frac{\text{m}}{\text{h}}$ in passive operation and $0.6 \frac{\text{m}}{\text{h}}$ in active mode (Zimmerman et al. 2001). In October 2001, the probe descended with $0.6\text{--}0.7 \frac{\text{m}}{\text{h}}$ into one of Svalbard's glaciers. While using 750 W heating power and operating in active mode, the probe successfully penetrated multiple particle-laden ice layers during this test (Bentley et al. 2009, p. 289). In response, ideas were also presented on how a nuclear reactor with an electrical output of several kilowatts could be integrated into the probe (Elliott and Carsey 2004). The proposed probe concept relies on a reactor design being already available at this time. As a consequence, the probe diameter increased to about 0.5 m, resulting in a dramatic increase in power demand for a given penetration rate.

29.5.2.2 Ice Melting Probe Experiments in Simulated Extraterrestrial Environments

During the period 2002–2009, several projects aiming at the development of ice melting probes for extraterrestrial applications were performed at the Space Research Institute (Institut für Weltraumforschung (IWF) in Graz, Austria). A research group around Norbert Kömle, working at IWF's Extraterrestrial Surfaces Laboratory, reasoned that experimental tests in a simulated extraterrestrial environment would be essential for deploying a melting probe on Mars or the icy moons (Kömle et al. 2002; Kaufmann et al. 2009). To investigate the performance of a classic (i.e., Philberth-like) ice melting probe under Mars surface conditions, two series of experiments were performed at IWF in 2017/2018. In the first series (Kömle et al. 2018b), a probe with square cross-section (40 mm × 40 mm) with a conical tip composed of copper was used. For the second test series (Kömle et al. 2018a), a probe with a hemispherical tip made of brass and a cylindrical body (Ø60 mm, length 12 cm) composed of stainless steel was employed. Both probes were refurbished versions from earlier experiments. The first series consisted of four experiments with different environmental pressures while keeping all other parameters the same. All samples consisted of semitransparent compact water ice.

In the first experiment, the surface gas pressure was in the range of 800–900 Pa, which is slightly above the water triple point, while in the second

experiment, the pressure was kept below the triple point, in the range of 500–600 Pa. In the remaining two tests, the gas pressures were much different from Mars conditions: 20 Pa in the third experiment and 10^5 Pa (Earth-normal surface pressure) in the fourth experiment. While the penetration rate under normal Earth surface conditions can be roughly estimated with Aamot's method as described in Aamot (1967a), the nature of the phase transition would have to be known exactly for a reasonable estimation under low ambient pressure. However, this is often not the case, since variations in pressure inside the melting channel, as well as a closing channel, have a notable influence on the phase change process at the melting head. The nature of this process, in turn, has a major effect on the penetration rate. Therefore, as was shown in Ulamec et al. (2007), it makes sense to estimate an effective specific heat L_{eff} , which is typically a value between the respective specific heat of melting and sublimation. Depending on the intensity of the evaporation taking place in the melt cavity, L_{eff} will be larger than the specific heat of melting L , and in the worst case, it can reach a value corresponding to pure sublimation, which is approximately eight times higher than L . Besides the probe geometry and ice porosity, heat capacity, melting temperature, and heat losses due to lateral convection also have an influence on L_{eff} , which makes it difficult to predict. The four melting experiments at IWF offer the possibility to estimate values for L_{eff} under different pressure conditions, also taking into account the measured average penetration rate for each case. The results of the four experiments described above indicate that the penetration rate as a function of ambient pressure can be well fitted using the extended formulae given in Ulamec et al. (2007).

The second series of experiments studied a potential probe deployment on Mars. It was devoted to the influence of embedded mineral layers and layers of more volatile CO_2 ice on the penetration performance of the probe. Moreover, the ice samples used were significantly higher than those in the first series (about 50 cm), which allowed the team to study probe penetration and the cable payout process over many probe lengths. Moreover, these experiments allowed researchers to investigate the evolution of the penetration channel with time and to observe the re-formation of the melt cavity halfway after a cooling phase. Figure 29.8a shows a sample consisting of compact ice layers with embedded mineral layers (water-saturated JSC1A-Mars material) of several centimeters in thickness. On the left side, the fresh sample is shown before starting the experiment, while the right-hand image illustrates the appearance of the sample after the end of the test, with the probe visible through the transparent ice near the bottom. Figure 29.8b shows the dust distribution after the experiment in detail, where the upper left image is a top view after opening of the chamber, the right image shows the distribution of the dust along the melt channel, and the bottom left view is a top view of the dust distribution shown in the right image. The last one only became visible after the outer layers (having poor transparency during the experiment) had melted in a room-temperature environment.

The main difficulty faced during these test series, which were mainly performed under low-pressure conditions, was the poor thermal conductance between the probe's hot tip and the surrounding ice. This often led to weak performance of ice penetration, which became worse, as the lower the environmental gas pressure

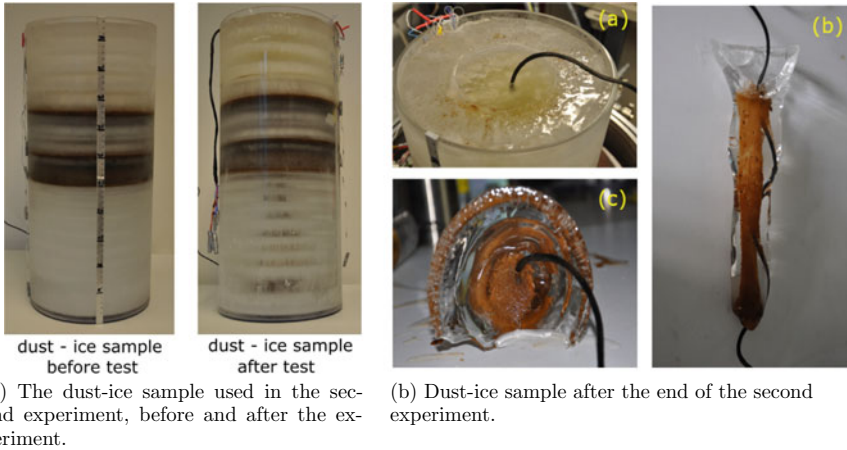


Fig. 29.8 Dust layer penetration experiments from Kömle et al. (2018a)

was chosen. An important conclusion that could be drawn from these tests was, however, that—for a certain cross-section of a probe—there seems to exist a threshold value for the power: If the power is below this threshold, no effective penetration will take place. Instead, the probe will excavate only a shallow crater around its tip and will stop penetration after some time, despite continuous heating. If the power is beyond this threshold value, penetration with a mostly constant velocity will set in and a narrow penetration channel will be created.

The behavior of melting probes in an extraterrestrial environment was also the subject of research at the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt (DLR)) in Cologne. One series of experiments used a simple probe consisting of a hemispherical melting head (\varnothing 115 mm) with 600 W heating power and attached to an exchangeable tube. The probe was first tested in a cold lab at $-20\text{ }^{\circ}\text{C}$ to generate reference data at ambient pressure. In a second step, it was deployed inside a vacuum chamber with a pressure below 1 mbar, which is well within the sublimation regime of water, which is below 6.11 mbar. Using liquid nitrogen as a cooling medium, ice temperatures were lowered to around 100 K. It was found that the observed penetration rates in both ambient and reduced pressure were very well aligned with predictions from a simple energy balance equation, when a 20–25% offset was attributed to an oversized melting channel and conductive heat loss, neither being considered in the calculation (Ulamec et al. 2005). However, the melting channel rapidly begins to close around the probe during the vacuum experiments after an initial phase of sublimation. As a consequence, the probe may advance at a higher speed due to the resulting increase in pressure around the melting head, causing the phase change to take place as melting again. Nevertheless, the observed increased penetration rate was still well below the prediction for pure melting penetration. On the other hand, the channel may also enclose the probe’s hull, causing it to stall completely. Both effects can occur intermittently and are hard to predict, but the latter

could be avoided by sufficient lateral heating power (Treffer et al. 2006). Depending on ice structure and porosity, it is also possible that melting does not take place at all, because the vapor escapes through permeable ice layers and the resulting pressure increase is not sufficiently large (Biele et al. 2011). Porosity also has direct and indirect effects on the required energy to cover a specific distance in the ice. Besides the lower mean density of porous ice, making an ice volume energetically cheaper to melt or sublimate, porosity also affects the excess diameter of the melt channel.

29.5.2.3 IceMole

The IceMole concept and design originated from a student initiative at FH Aachen University of Applied Sciences (FH Aachen) in Germany. It attacked the problem of insufficient contact between the melting head and the ice surface in front of it. Furthermore, through its ice screw and differential heating, the IceMole is a maneuverable probe that can even move against gravity (Dachwald et al. 2014). Following the construction of a prototype in 2009, IceMole 1 was tested as the first complete implementation of the concept. After three major design iterations, the EnEx-IceMole¹ represents its most advanced development dedicated to field usage.

Like conventional melting probes, the IceMole is equipped with an electrically heated melting head, but is also characterized by a motor-driven ice screw at the center of the melting head. The screw geometry is identical to that of commercially available ice screws, as used in alpine sports and ice climbing (ca. Ø 18 mm, length 12 cm). A second remarkable feature derives directly from the use of the motor-driven ice screw. As a circular outer hull would not counter the torque of the screw drive, the IceMole is designed with a square instead of a circular cross-section. All IceMole probes until 2014 relied on the same square shape with an outer edge length of 150 mm. Besides the need for torque compensation, a square shape also simplifies payload accommodation but yields generally reduced melting efficiency. As the system is designed in a modular manner, the length varies with the payload to be used. Typically, the length is 1–2 m, of which approximately 50 cm is required for the obligatory heater control and drive system in the front and a power and communications compartment in the back of the probe. The front heater system is equipped with heating cartridges, providing 2.4 kW (2.88 kW in the EnEx-IceMole). They may be activated separately and segment-wise for each quarter of the melting head with a defined power. Typical penetration rates are around $2.2\text{--}4 \frac{\text{m}}{\text{h}}$. Additionally, all but the first IceMole featured sidewall heaters on all four sides with varying heating power and power distribution to support curve melting. However, the cable used for power and communications is not stored inside of the probe but has to be pushed from the surface. Consequently, great depths or cold ice would be problematic if the melting channel closes behind the probe. To change the melting direction, the heating power of one half (*A*) of the melting head is reduced, while the other half (*B*) stays fully active. As a consequence, the penetration rate on side *B* will be higher,

¹ EnEx: Enceladus Explorer Initiative by the German Aerospace Center, DLR Space Administration.

as the head is firmly pressed onto the ice supported by the screw drive. The curve radius may be further reduced by activating the sidewall heater on side *B* to enable the probe's tail to swing out. For the comparatively long EnEx-IceMole, a curve radius as small as 10 m was demonstrated inside the Swiss Morteratsch Glacier. The ability to penetrate dust layers between the ice was also demonstrated. However, this method becomes increasingly ineffective for very dirty ice (Dachwald et al. 2014).

In combination with a suitable in-ice navigation solution for localization and targeting and a reconnaissance system for obstacle avoidance, a maneuverable IceMole probe could be used to autonomously penetrate a remote ice sheet and to reach a defined location. For the EnEx-IceMole, such a system was developed and successfully tested as part of the Enceladus Explorer (EnEx) initiative, as described in Kowalski et al. (2016). A dead reckoning system based on inertial and magnetic attitude determination is used to determine the probe's location relative to the starting point. It also takes into account the distance traveled, which is derived from the ice screw thread pitch and rotational frequency. An acoustic positioning system provides information on the absolute position inside an ice volume. Based on trilateration, the system is dependent on additional ultrasonic transmitters inside the ice. Experimental forefield exploration is provided by a sonographic acoustic reconnaissance system. To detect and locate potential obstacles on the projected melting path, phased acoustic signals are emitted from arrays inside the melting head. In case of disturbances inside the carrier medium ice (for example, obstacles, impurities, crevasses), an echo signal is received and processed. Finally, the information acquired by those systems is processed by a high-level sensor fusion unit to increase accuracy. While some subsystems performed as expected in a glacier environment, the absolute positioning system was found to be very sensitive to unknown or changing ice properties. Therefore, before deploying the system in an extraterrestrial ice volume, a good understanding of the local ice properties and structure is necessary. In addition, the integration of the phased arrays into the melting head had a very negative impact on the penetration rate and remains an engineering challenge.

The IceMole probes were extensively tested in alpine glacier ice (Switzerland, Iceland, Italy), as well as inside the Antarctic Taylor Glacier. The concept proved to be reliable (in comparison with the field application of earlier probes), and all probes were recovered from the ice. The EnEx-IceMole was also shown to be capable of taking a clean sample from a subglacial water reservoir without prior knowledge about the location of the water–ice interface. For this purpose, the probe descended into a predefined target area and located the subsurface crevasse. It stopped in its very proximity and was decontaminated with hydrogen peroxide, using its clean access and sampling system. The sterilized probe then penetrated the ice wall and sampled several hundred liters of englacial water through a proboscis, which was deployed pneumatically through the hollow ice screw as shown in Fig. 29.9.

In 2015, the development of a new IceMole-like probe was begun, which is dedicated to laboratory use in a simulated extraterrestrial environment. With a square cross-section of 80 mm edge length and a total length (without ice screw) of only 38 cm, it is significantly smaller than other probes (Baader et al. 2016). At the same time, it has higher total heating power (8.22 kW in total, of which 4 kW is avail-

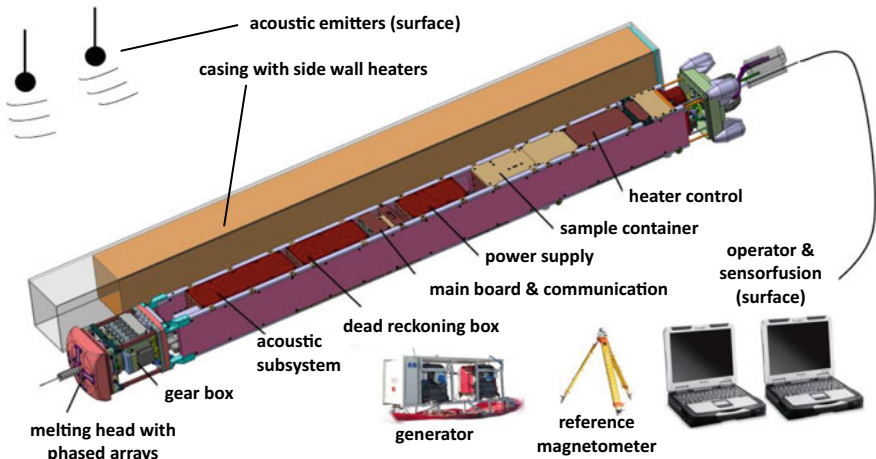


Fig. 29.9 EnEx-IceMole systems overview and setup (from Kowalski et al. (2016))

able in the melting head). Heaters are controlled in 30 separate groups. The power setting of each heater group is autonomously updated once per second to reach its individual target temperature. Lower power constraints can be simulated in software by setting a power limit which may not be exceeded. However, because of its high level of integration, the probe does not provide a standardized payload interface like its predecessors. Instead, it carries a set of integrated sensors to investigate melting efficiency and maneuverability in low ice temperatures (around 90 K) and at low pressure (around 1 mbar). These include tracking of heater activation and power state, ice screw force measurement, pressure sensors along the outer hull, melting distance measurement, an inertial measurement unit (IMU), and cameras for infrared and visible light in the back to observe melting channel variation over time.

29.5.2.4 Ice Diver

Ice Diver is a traditional melting probe developed at the Applied Physics Laboratory of the University of Washington with a focus on the exploration of extraterrestrial ices. The probe, described as a subscale prototype, stands out due to its very slim shape. It also has many features resembling the Philberth probes described in Sect. 29.5.1, including pendulum steering and two conductors, with only one of them being insulated. With a melting head diameter of only 65 mm and an upper hot point diameter of 85 mm, but heating power in the same order as other probes, comparatively high penetration rates are possible. Unlike the high-voltage probes of the 1960s, the Ice Diver probes are reported to operate reliably at 1050 V and at power of 2.15 kW. With this input power and in -15°C cold ice, a probe first descended to a depth of 80 m at $2.4 \frac{\text{m}}{\text{h}}$ and then to 400 m after a restart one year later. The fully operational

probe was then deactivated. Another probe was operated successfully at 2000 V and higher power of 4.5 kW, which allowed it to advance at the impressive speed of $6.6 \frac{\text{m}}{\text{h}}$ (Winebrenner et al. 2016).

29.5.2.5 VALKYRIE

In 2014, an innovative approach was demonstrated by Stone Aerospace for the transmission of large amounts of energy to a subsurface ice melting probe without risking short circuits in a marginally insulated high-voltage conductor. As part of the VALKYRIE project, a cylindrical melting probe was designed to be supplied with heat and electrical power by an optical transmission system (Stone et al. 2018). Using a commercial 5 kW laser system on the ice surface, the laser beam is directed into a wavelength-adjusted optical fiber, which is unwound from a spool inside the probe during probe operation. Inside the probe, the laser beam is utilized to generate both heat and electricity using photovoltaics. As in other probes, the spool diameter constrains the lower limit for the total probe diameter. The core test vehicle in 2014 had a length of 1.5 m, whereas a second version in 2015 including scientific payload reached 2.5 m (Stone et al. 2018). While the heat is used to heat water for five hot water jets (one directed in the frontal and four in the lateral direction), the generated electric power of 160 W is used to power the onboard electronics and pump systems driving both the cooling circulation and water jets. During a field test in 2014 on an Alaskan temperate glacier, the VALKYRIE probe demonstrated a penetration rate between 0.85 and $0.95 \frac{\text{m}}{\text{h}}$. The tested probe is shown in Fig. 29.10. Applying its lateral jets, the team also demonstrated small controlled deviations from the vertical direction.

29.5.2.6 RECAS

The Recoverable Autonomous Sonde (RECAS) is a melting probe which was proposed by Pavel Talalay at the Polar Research Center at Jilin University, China. The ice melting probe is designed for clean access to a subglacial lake covered by ice sheets up to 4 km thick, including sample and probe recovery (Talalay et al. 2013). On both the upper and lower ends, it carries elongated cone-shaped melting heads with a parabolic tip, which are supposed to combine adequate efficiency, stability, and particle tolerance. Both have a diameter of 150 mm and an electric heating power of 5 kW installed. The midsection is a circular tube with a slightly smaller diameter of 140 mm. Lateral heaters with an adjusted power density (decreasing linearly from front to rear) on the side walls prevent the probe from freezing due to refreezing meltwater in cold ice. Both its length and lateral heating power depend strongly on the desired maximum depth and the corresponding coaxial cable length, which is stored on a coil inside the probe. Talalay considers two versions of RECAS with different cable lengths, the first having 1.2 km of cable stored on a coil 1.3–1.5 m in length, and the second with 4 km of cable on a 4.4–5 m long coil. The conductor cable is unwound



Fig. 29.10 The VALKYRIE probe during a field test on Matanuska glacier in Alaska, 2014, from Stone et al. (2018, p. 81)

from the coil with a gear-driven motor, which is also capable of retracting it when recovering the probe. Like earlier probe designs, the high-voltage power supply with a power line as thin as possible was identified as a potential risk for insulation failure when exposed to the high hydrostatic pressure inside a deep borehole (Talalay et al. 2014). Additionally, RECAS carries an integrated payload for meltwater sampling and in situ analysis or storage, as well as sensors to investigate borehole orientation, pressure, temperature, pH value, sound velocity, and conductivity.

29.5.2.7 IceShuttle Teredo

The German Robotics Innovation Center (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)) presented a prototype of a melting probe, the IceShuttle Teredo, as part of the Europa Explorer (EurEx) initiative of the German DLR Space Administration. In a potential icy moon ocean exploration mission, an autonomous underwater vehicle (AUV) would first have to be transported through the thick ice layer overlaying the ocean. The IceShuttle Teredo is designed to carry such a vehicle down to the ice–ocean interface, where it should anchor itself to the walls of the melting channel. With its front end reaching into the open water, the IceShuttle is designed to release an elongated AUV into the water. Also, a set of small underwater gliders is accommodated inside the probe and is aimed to be released to establish a local navigation network. The design is strongly driven by the large primary payload, and

therefore the probe has extraordinary dimensions in both length (6.75 m) and diameter ($\text{\O}280$ mm) (DFKI 2016). However, the hull wall thickness of this prototype can only withstand the pressure of water depths up to 20 m. Hence, a probe built for greater depths would require a thicker wall, resulting in an even larger diameter (Wirtz and Hildebrandt 2016). The melting head is heated by six electric cartridge heaters with 3.6 kW in total and is estimated to enable ice penetration with $0.8 \frac{\text{m}}{\text{h}}$.

29.6 Summary and Conclusions

The experiences from the deployment and testing of ice melting probes indicate several recurring challenges in the design and operations, but also reveal some proven technical solutions. The provision and transmission of power to a probe clearly remains a major challenge, especially for operations in extraterrestrial environments. A classic ice melting probe setup includes a power source on the surface, which provides electrical energy via a tether/cable with conductor wires. However, since the melt channel tends to close again after the probe has passed, the tether/cable has to be stored inside the probe, which takes up considerable space and increases the total power demand. In order to obtain the highest possible packing density (to realize realistic probe dimensions), the insulation layer of conductor wires is often extremely thin and at high risk of being damaged during preparation work or spooling, by a closing melting channel, and by stresses in the ice. To avoid this risk, some concepts propose using optical fibers for power transmission or even to generate heat and electricity inside the probe itself. Some concepts rely on additional hot water jets to assist the passive melting using a heated tip. While it turned out that this technology can be advantageous in compact or particle-laden ice, there are also cases where water jets are very ineffective. Experiences from the Cryobot penetrating Antarctic glacier ice of Ice Stream C showed that passive melting is especially important in porous ice layers (firn, for example). In this case, meltwater does not accumulate in front of the probe but is rinsed down through the pores. In contrast, in compact ice masses, active melting is preferable, because it generally offers better efficiency. Especially at low melting speed and in porous ice, low energy efficiency was described as a major drawback. Consequently, available heating power was identified as an important design challenge in the development of probes for extraterrestrial use (Zimmerman et al. 2001). Experiences concerning planetary protection are positive. The deployment of IceMole in an Antarctic Specially Protected Area (ASPA) was very successful and demonstrated effective in situ decontamination. An experimental investigation of microbiological contamination caused by melting probes supports the implementation of complex decontamination procedures before sampling a secluded water reservoir, as was done with the Antarctic EnEx-IceMole deployment. By using Ice Diver probes in a lab environment, Schuler et al. (2018) showed that bacteria from upper ice layers can be carried down into deeper layers. Accumulations of bacteria were detected in particular when a point of increasing diameter passed the detector (that is, the melting head and upper hot point of the Ice Diver). Effective directional

stabilization can be achieved by a variety of passive and active methods as described in Sect. 29.5. Further, several steering approaches were presented, which may be coupled with sophisticated navigation and localization technology, if desired. During field tests, some methods proved to be sensitive to varying ice properties. This impedes the transfer of laboratory tests to the field (Kowalski et al. 2016). An extraterrestrial application of ice melting probes requires an even more thorough understanding of the prevailing environment to enable a realistic performance estimation during the design phase.

Generally, melting probe technology appears to be an attractive way to penetrate extraterrestrial ice bodies. However, several technological hurdles need to be overcome before this technology is mature enough to be used in a real space mission. This is particularly valid if long-range melting is envisaged, as it is required for reaching the ice–water interface of icy moons. The main areas where future development efforts need to be undertaken are power supply, communications, and navigation. When applying a melting probe at Europa or Enceladus, significant efforts will also be needed in the selection, development, and miniaturization of optimized instrumentation.

29.7 List of Acronyms

ASPA	Antarctic Specially Protected Area
AUV	Autonomous Underwater Vehicle
AWI	Alfred Wegener Institute for Polar and Marine Research
CRREL	Cold Regions Research and Engineering Laboratory
COSPAR	Committee on Space Research
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
DLR	Deutsches Zentrum für Luft- und Raumfahrt
ECSS	European Cooperation for Space Standardization
EnEx	Enceladus Explorer
EurEx	Europa Explorer
FH Aachen	FH Aachen University of Applied Sciences
IMU	Inertial Measurement Unit
IWF	Institut für Weltraumforschung
JPL	NASA Jet Propulsion Laboratory
RECAS	Recoverable Autonomous Sonde
RTG	Radioisotope Thermoelectric Generator
SCAR	Scientific Committee on Antarctic Research
SPT	South Polar Terrain
SUSI	Shuttle Under Ice Shelf
VALKYRIE	Very-deep Autonomous Laser-powered Kilowatt-class Yo-yoing Robotic Ice Explorer

References

- Aamot, H.W.C. 1967a. Heat Transfer and Performance Analysis of a Thermal Probe for Glaciers. In: CRREL Technical Report 194.
- Aamot, H.W.C. 1967b. Pendulum Steering for Thermal Probes in Glaciers. *Journal of Glaciology* 6(48): 935–938. <https://doi.org/10.3189/S0022143000020220>.
- Aamot, H.W.C. 1967c. The Philberth Probe for Investigating Polar Ice Caps.
- Aamot, H.W.C. 1968. Instrumented Probes for Deep Glacial Investigations. *Journal of Glaciology* 7(50): 321–327.
- Aamot, H.W.C. 1970. Self-Contained Thermal Probes for Remote Measurements Within an Ice Sheet. *International Association of Scientific Hydrology* 86: 63–68.
- Alfred Wegener Institute for Polar and Marine Research. 1995. Expeditionsprogramm Nr. 36.
- Baader, F., B. Dachwald, C. Espe, et al. 2016. Testing of a Miniaturized Subsurface Icecraft for an Enceladus Lander Mission Under Enceladus-Like Conditions. In *Conference on Enceladus and the Icy Moons of Saturn. LPI Contribution No. 1927*. Boulder, USA: Lunar and Planetary Institute, Houston, Abstract #3029.
- Bentley, C.R., B.R. Koci, L.J.-M. Augustin, et al. 2009. Ice Drilling and Coring. In *Drilling in Extreme Environments*, ed. Y. Bar-Cohen, K. Zacny, Chap. 4, pp. 221–308. Wiley. <https://doi.org/10.1002/9783527626625.ch4>.
- Bidle, K.D., S. Lee, D.R. Marchant, et al. 2007. Fossil Genes and Microbes in the Oldest Ice on Earth. *Proceedings of the National Academy of Sciences* 104(33): 13455–13460.
- Biele, J., S. Ulamec, M. Hilchenbach, et al. 2011. In Situ Analysis of Europa Ices by Short-Range Melting Probes. *Advances in Space Research* 48(4): 755–763. <https://doi.org/10.1016/j.asr.2010.02.029>.
- Bryant, S. 2002. Ice-Embedded Transceivers for Europa Cryobot Communications. In *Proceedings, IEEE Aerospace Conference*, vol. 1, pp. 1-349–1-356. <https://doi.org/10.1109/AERO.2002.1036854>.
- Byrne, S. 2009. The Polar Deposits of Mars. *Annual Review of Earth and Planetary Sciences* 37: 535–560. <https://doi.org/10.1146/annurev.earth.031208.100101>.
- Campen, R., J. Kowalski, W.B. Lyons, et al. 2019. Microbial Diversity of an Antarctic Subglacial Community and High-Resolution Replicate Sampling Inform Hydrological Connectivity in a Polar Desert. *Environmental Microbiology* 21(7): 2290–2306. <https://doi.org/10.1111/1462-2920.14607>.
- Carr, M.H., M.J.S. Belton, C.R. Chapman, et al. 1998. Evidence for a Subsurface Ocean on Europa. *Nature* 391(6665):363–365. <https://doi.org/10.1038/34857>.
- Checinska Sielaff, A. and S.A. Smith. 2019. Habitability of Mars: How Welcoming Are the Surface and Subsurface to Life on the Red Planet? *Geosciences* 9(9): 361. <https://doi.org/10.3390/geosciences9090361>.
- Chyba, C.F. and C.B. Phillips. 2007. Europa. In *Planets and Life* ed. W.T. Sullivan and J.A. Baross, pp. 388–423. Cambridge University Press.
- Clark, N.R., M.J. Curchin, J.W. Barnes, et al. 2010. Detection and Mapping of Hydrocarbon Deposits on Titan. *Journal of Geophysical Research: Planets* 115(E10). <https://doi.org/10.1029/2009JE003369>.
- Clarke, A., G.J. Morris, F. Fonseca, et al. 2013. A Low Temperature Limit for Life on Earth. *PLoS ONE* 8(6): e66207. <https://doi.org/10.1371/journal.pone.0066207>.
- Clifford, S.M., D. Crisp, D.A. Fisher, et al. 2000. The State and Future of Mars Polar Science and Exploration. *Icarus* 144(2): 210–242. <https://doi.org/10.1006/icar.1999.6290>.
- Cockell, C.S. 2014. Habitable Worlds With No Signs of Life. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372(2014): 20130082. <https://doi.org/10.1098/rsta.2013.0082>.
- Cockell, C.S., E. Bagshaw, M. Balme, et al. 2011. Subglacial Environments and the Search for Life Beyond the Earth. In *Antarctic Subglacial Aquatic Environments*, ed. M.J. Siegert, pp. 129–147. American Geophysical Union. <https://doi.org/10.1029/2010GM000939>.

- Committee on Planetary Protection Standards for Icy Bodies in the Outer Solar System. 2012. *Assessment of Planetary Protection Requirements for Spacecraft Missions to Icy Solar System Bodies*. National Research Council: Technical report.
- D'Elia, T., R. Veerapaneni, and S.O. Rogers. 2008. Isolation of Microbes from Lake Vostok Accretion Ice. *Applied and Environmental Microbiology* 74(15): 4962–4965. <https://doi.org/10.1128/AEM.02501-07>.
- Dachwald, B., J. Mikucki, S. Tulaczyk, et al. 2014. IceMole: A Maneuverable Probe for Clean in Situ Analysis and Sampling of Subsurface Ice and Subglacial Aquatic Ecosystems. *Annals of Glaciology* 55(65): 14–22. <https://doi.org/10.3189/2014AoG65A004>.
- Dachwald, B., S. Ulamec, and J. Biele. 2013. Clean In-Situ Subsurface Exploration of Icy Environments in the Solar System. In *Habitability of Other Planets and Satellites* ed. J.-P. de Vera and J. Seckbach, pp. 367–397. Springer. https://doi.org/10.1007/978-94-007-6546-7_20.
- Dachwald, B., S. Ulamec, F. Postberg, et al. 2020. Key Technologies and Instrumentation for Subsurface Exploration of Ocean Worlds. *Space Science Reviews* 216(5): 1–45. <https://doi.org/10.1007/s11214-020-00707-5>.
- Dalton, J.B. 2010. Spectroscopy of Icy Moon Surface Materials. *Space Science Reviews* 153(1–4): 219–247. <https://doi.org/10.1007/s11214-010-9658-7>.
- Deming, J.W. and H. Eicken. 2007. Life in Ice. In *Planets and Life* ed. W.T. Sullivan and J.A. Baross, pp. 292–312. Cambridge University Press.
- DFKI. 2016. Teredo IceShuttle-Through-Ice-Cap Transfer-Vehicle & Base Station. System Description. Robotics Innovation Center (DFKI GmbH). https://robotik.dfki-bremen.de/uploads/tx_dfkiprojects/Systemblatt_IceShuttleTeredo_en_02.pdf.
- Di Pippo, S., R. Mugnuolo, P. Vielmo, et al. 1999. The Exploitation of Europa Ice and Water Basins: an Assessment on Required Technological Developments, on System Design Approaches and on Relevant Expected Benefits to Space and Earth Based Activities. *Planetary and Space Science* 47(6–7): 921–933. [https://doi.org/10.1016/S0032-0633\(99\)00012-4](https://doi.org/10.1016/S0032-0633(99)00012-4).
- Durham, W.B., O. Prieto-Ballesteros, D.L. Goldsby, et al. 2010. Rheological and Thermal Properties of Icy Materials. *Space Science Reviews* 153(1–4): 273–298. <https://doi.org/10.1007/s11214-009-9619-1>.
- Elliott, J.O., and F.D. Carsey. 2004. Deep Subsurface Exploration of Planetary Ice Enabled by Nuclear Power. *IEEE Aerospace Conference Proceedings* 5: 2978–2986. <https://doi.org/10.1109/AERO.2004.1368103>.
- Feistel, R. and W. Wagner. 2006. A New Equation of State for H₂O Ice Ih. *Journal of Physical and Chemical Reference Data* 35(2): 1021–1047. <https://doi.org/10.1063/1.2183324>.
- Fraser, S.D. 2018. Exploration of the Outer Solar System: Missions and Their Power Systems Exploration Missions. In *Outer Solar System: Prospective Energy and Material Resources* ed. V. Badescu and K. Zacny, pp. 823–836. Springer. https://doi.org/10.1007/978-3-319-73845-1_18.
- French, L., F.S. Anderson, F. Carsey, et al. 2001. Cryobots: An Answer to Subsurface Mobility in Planetary Icy Environments. In *Proceedings of the 6th International Symposium on Artificial Intelligence and Robotics & Automation in Space (i-SAIRAS 2001)*. St-Hubert, Canada.
- Gaidos, E.J., K.H. Nealson, and J.L. Kirschvink. 1999. Life in Ice-Covered Oceans. *Science* 284(5420): 1631–1633. <https://doi.org/10.1126/science.284.5420.1631>.
- El-Genk, M.S. 2009. Deployment History and Design Considerations for Space Reactor Power Systems. *Acta Astronautica* 64(9): 833–849. <https://doi.org/10.1016/j.actaastro.2008.12.016>.
- German, L., J.A. Mikucki, S.A. Welch, et al. 2019. Validation of Sampling Antarctic Subglacial Hypersaline Waters With an Electrothermal Ice Melting Probe (IceMole) for Environmental Analytical Geochemistry. *International Journal of Environmental Analytical Chemistry* 1–14. <https://doi.org/10.1080/03067319.2019.1704750>.
- Gerrard, J.A.F., M.F. Perutz, and A. Roch. 1952. Measurement of the Velocity Distribution Along a Vertical Line Through a Glacier. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 213(1115): 546–558. <https://doi.org/10.1098/rspa.1952.0144>.

- Gibson, M.A., S.R. Oleson, D.I. Poston, et al. 2017. NASA's Kilowatt Reactor Development and the Path to Higher Power Missions. In *2017 IEEE Aerospace Conference*. Big Sky, MT, USA. <https://doi.org/10.1109/AERO.2017.7943946>.
- Goguen, J.D., B.J. Buratti, R.H. Brown, et al. 2013. The Temperature and Width of an Active Fissure on Enceladus Measured with Cassini VIMS During the 14 April 2012 South Pole Flyover. *Icarus* 226(1): 1128–1137. <https://doi.org/10.1016/j.icarus.2013.07.012>.
- Grasset, O., M.K. Dougherty, A. Coustenis, et al. 2013. Jupiter ICy moons Explorer (JUICE): An ESA Mission to Orbit Ganymede and to Characterise the Jupiter System. *Planetary and Space Science* 78: 1–21. <https://doi.org/10.1016/j.pss.2012.12.002>.
- Greenberg, R., and P. Geissler. 2002. Europa's Dynamic Icy Crust. *Meteoritics and Planetary Science* 37: 1685–1711. <https://doi.org/10.1111/j.1945-5100.2002.tb01158.x>.
- Hammel, T., R. Bennett, W. Otting, et al. 2009. Multi-Mission Radioisotope Thermoelectric Generator (MMRTG) and Performance Prediction Model. In *7th International Energy Conversion Engineering Conference. Paper AIAA 2009-4576*. Denver, CO, USA. <https://doi.org/10.2514/6.2009-4576>.
- Hand, K.P., A.E. Murray, J.B. Garvin, et al. 2017. Report of the Europa Lander Science Definition Team. Technical report JPL D-97667. NASA/JPL. <https://europa.nasa.gov/resources/58/europa-lander-study-2016-report/>. (visited on 01/09/2021).
- Haynes, W.M., D.R. Lide, and T.J. Bruno. 2017. *CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*. Boca Raton, FL, USA: CRC Press. ISBN: 9781498754293.
- Hemingway, D., L. Iess, R. Tajeddine, et al. 2018. The Interior of Enceladus. In *Enceladus and the Icy Moons of Saturn*, pp. 57–77. The University of Arizona Press. https://doi.org/10.2458/azu_uapress_9780816537075-ch004.
- Howell, S. 2020. The Likely Thickness of Europa's Icy Shell. In *Europlanet Science Congress 2020*. EPSC Abstracts EPSC2020-173. <https://doi.org/10.5194/epsc2020-173>.
- Jansen, F., W. Bauer, F. Masson, et al. 2016. DEMOCRITOS Demonstrators for Realization of Nuclear Electric Propulsion of the European Roadmaps MEGAHIT & DiPoP. *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan* 14(ists30): Pb_225–Pb_233. https://doi.org/10.2322/tastj.14.Pb_225.
- Kargel, J.S. 1998. The Salt of Europa. *Science* 280(5367): 1211–1212. <https://doi.org/10.1126/science.280.5367.1211>.
- Kaufmann, E., G. Kargl, N.I. Kömle, et al. 2009. Melting and Sublimation of Planetary Ices Under Low Pressure Conditions: Laboratory Experiments with a Melting Probe Prototype. *Earth, Moon, and Planets* 105(1): 11–29. <https://doi.org/10.1007/s11038-009-9296-9>.
- Kite, E.S. and A.M. Rubin (2016). Sustained Eruptions on Enceladus Explained by Turbulent Dissipation in Tiger Stripes. *Proceedings of the National Academy of Sciences* 113(15): 3972–3975. <https://doi.org/10.1073/pnas.1520507113>.
- Kminek, G., C. Conley, V. Hipkin, et al. 2017. COSPAR's Planetary Protection Policy. *Space Research Today* 200: 12–25. <https://doi.org/10.1016/j.srt.2017.11.010>.
- Kömle, N.I., G. Kargl, and M. Steller. 2002. Melting Probes as a Means to Explore Planetary Glaciers and Ice Caps. *Proceedings of the Second European Workshop on Exo/Astrobiology* ed. H. Sawaya-Lacoste. ESA SP-518, pp. 305–308. Graz, Austria. <http://adsabs.harvard.edu/full/2002ESASP.518.305K>.
- Kömle, N.I., P. Tiefenbacher, and A. Kahr. 2018a. Melting Probe Experiments Under Mars Surface Conditions-The Influence of Dust Layers, CO₂-Ice and Porosity. *Icarus* 315: 7–19. <https://doi.org/10.1016/j.icarus.2018.06.012>.
- Kömle, N.I., P. Tiefenbacher, P. Weiss, et al. 2018b. Melting Probes Revisited-Ice Penetration Experiments Under Mars Surface Pressure Conditions. *Icarus* 308: 117–127. <https://doi.org/10.1016/j.icarus.2017.08.006>.
- Konstantinidis, K., C.L. Flores Martinez, B. Dachwald, et al. 2015. A Lander Mission to Probe Subglacial Water on Saturn's Moon Enceladus for Life. *Acta Astronautica* 106: 63–89. <https://doi.org/10.1016/j.actaastro.2014.09.012>.

- Koroteev, A.S., Akimov V.N., and A.V. Semekin. 2013. Nuclear Power as the Way to far Space Exploration-The Russian Experience and the Prospects for the Growth. In: MEGAHIT Workshop. Brussels, Belgium.
- Kowalski, J., P. Linder, S. Zierke, et al. 2016. Navigation Technology for Exploration of Glacier Ice with Maneuverable Melting Probes. *Cold Regions Science and Technology* 123: 53–70. <https://doi.org/10.1016/j.coldregions.2015.11.006>.
- Kraus, E.I. and I.I. Shabalin. 2013. Impact Loading of a Space Nuclear Powerplant. *Frattura ed Integrità Strutturale* 7(24): 138–150. <https://doi.org/10.3221/IGF-ESIS.24.15>.
- Lauro, S.E., E. Pettinelli, G. Caprarelli, et al. 2020. Multiple Subglacial Water Bodies Below the South Pole of Mars Unveiled by New MARSIS Data. *Nature Astronomy* 1–8. <https://doi.org/10.1038/s41550-020-1200-6>.
- Le Gall, A., C. Leyrat, M.A. Janssen, et al. 2017. Thermally Anomalous Features in the Subsurface of Enceladus’s South Polar Terrain. *Nature Astronomy* 1(4): 0063. <https://doi.org/10.1038/s41550-017-0063>.
- Lee, S., R.T. Pappalardo, and N.C. Makris. 2005. Mechanics of Tidally Driven Fractures in Europa’s Ice Shell. *Icarus* 177(2): 367–379. <https://doi.org/10.1016/j.icarus.2005.07.003>.
- Li, Y. et al. 2020. Power Consumption of a Philberth Thermal Probe in Ice Sheet Exploration. *Cold Regions Science and Technology* 177: 103114. <https://doi.org/10.1016/j.coldregions.2020.103114>.
- Lorenz, R.D. 2012. Thermal Drilling in Planetary Ices: An Analytic Solution with Application to Planetary Protection Problems of Radioisotope Power Sources. *Astrobiology* 12(8): 799–802. <https://doi.org/10.1089/ast.2012.0816>.
- Marion, G.M., C.H. Fritsen, H. Eicken, et al. 2002. The Search for Life on Europa: Limiting Environmental Factors, Potential Habitats, and Earth Analogues. *Astrobiology* 3(4): 785–811. <https://doi.org/10.1089/15311070322736105>.
- McCarthy, T.C., K.L. Craft, C.R. German, et al. 2019. Europa STI: Exploring Communication Techniques and Strategies for Sending Signals Through the Ice (STI) for an Ice-Ocean Probe. In: *Oceans Worlds 2019* 6023.
- McCord, T.B., G.B. Hansen, F.P. Fanale, et al. 1998. Salts on Europa’s Surface Detected by Galileo’s Near Infrared Mapping Spectrometer. *Science* 280(5367): 1242–1245. <https://doi.org/10.1126/science.280.5367.1242>.
- Mikucki, J.A., A. Pearson, D.T. Johnston, et al. 2009. A Contemporary Microbially Maintained Subglacial Ferrous “Ocean”. *Science* 324: 397–400. <https://doi.org/10.1126/science.1167350>.
- Mikucki, J.A. and J.C. Priscu. 2007. Bacterial Diversity Associated with Blood Falls, a Subglacial Outflow from the Taylor Glacier, Antarctica. *Applied and Environmental Microbiology* 73(12): 4029–4039. <https://doi.org/10.1128/AEM.01396-06>.
- Moore, J.M., G. Black, B. Buratti, et al. 2009. Surface Properties, Regolith, and Landscape Degradation. In *Europa*, ed. R.T. Pappalardo, W.B. McKinnon, and K.K. Khurana, pp. 329–349. The University of Arizona Press.
- Moore, W.B., H. Hussmann, R.T. Pappalardo, et al. 2009. Thermal Evolution of Europa’s Silicate Interior. In *Europa*, ed. R.T. Pappalardo, W.B. McKinnon, and K.K. Khurana, pp. 369–380. University of Arizona Press.
- NASA Glenn Research Center. 2019. SESAME (Scientific Exploration Subsurface Access Mechanism for Europa). <https://www1.grc.nasa.gov/space/pesto/space-vehicle-technologies-current/scientific-explorationsubsurface-access-mechanism-for-europa-sesame/>. (visited on 06/13/2021).
- NASA JPL. 2021. Solar System Dynamics Website. <https://ssd.jpl.nasa.gov>. (visited on 06/13/2021).
- NASA NSSDCA. 2020. Mars Fact Sheet. <https://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html>. (visited on 06/13/2021).
- Nimmo, F. and R.T. Pappalardo. 2016. Ocean worlds in the outer solar system. *Journal of Geophysical Research: Planets* 121(8): 1378–1399. <https://doi.org/10.1002/2016JE005081>.

- Nordheim, T.A., K.P. Hand, and C. Paranicas. 2018. Preservation of Potential Biosignatures in the Shallow Subsurface of Europa. *Nature Astronomy* 2(8): 673–679. <https://doi.org/10.1038/s41550-018-0499-8>.
- Orosei, R., R.L. Jordan, D.D. Morgan, et al. 2015. Mars Advanced Radar for Subsurface and Ionospheric Sounding (MARSIS) After Nine Years of Operation: A Summary. *Planetary and Space Science* 112: 98–114. <https://doi.org/10.1016/j.pss.2014.07.010>.
- Orosei, R., S.E. Lauro, E. Pettinelli, et al. 2018. Radar Evidence of Subglacial Liquid Water on Mars. *Science* 361(6401): 490–493. <https://doi.org/10.1126/science.aar7268>.
- Paranicas, C., J.F. Cooper, H.B. Garrett, et al. 2009. Europa's Radiation Environment and Its Effects on the Surface. In *Europa*, ed. R.T. Pappalardo, W.B. McKinnon, and K.K. Khurana, pp. 529–544. The University of Arizona Press.
- Philberth, B. 1959. Stockage des déchets atomiques dans les calottes glaciaires de la Terre. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 248(2): 2090–2092.
- Philberth, B. 1961. Beseitigung radioaktiver Abfallsubstanzen in den Eiskappen der Erde. *Schweizerische Zeitschrift für Hydrologie* 23(1): 263–284. <https://doi.org/10.1007/BF02505629>.
- Philberth, K. 1962. Une méthode pour mesurer les températures à l'intérieur d'un Inlandsis. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (Paris)* 254(22): 3881–3883.
- Philberth, K. 1964. Über zwei Elektro-Schmelzsonden mit Vertikal-Stabilisierung. *Polarforschung* 34(1/2): 278–280.
- Phillips, C.B. and R.T. Pappalardo. 2014. Europa Clipper Mission Concept: Exploring Jupiter's Ocean Moon. *Eos, Transactions American Geophysical Union* 95(20): 165–167. <https://doi.org/10.1002/2014EO200002>.
- Priscu, J.C., E.E. Adams, W.B. Lyons, et al. 1999. Geomicrobiology of Subglacial Ice Above Lake Vostok, Antarctica. *Science* 286: 2141–2144. <https://doi.org/10.1126/science.286.5447.2141>.
- Priscu, J.C. and B.C. Christner. 2004. Earth's Icy Biosphere. In *Microbial Diversity and Prospecting* ed. A. Bull, pp. 130–145. ASM Press. <https://doi.org/10.1128/9781555817770.ch13>.
- Prockter, L.M. and R.T. Pappalardo. 2014. Europa. In *Encyclopedia of the Solar System* ed. T. Spohn, D. Breuer, and T.V. Johnson. Third. Elsevier, pp. 793–811.
- Schuler, C.G., D.P. Winebrenner, W.T. Elam, et al. 2018. In Situ Contamination of Melt Probes: Implications for Future Subglacial Microbiological Sampling and Icy Worlds Life Detection Missions. In *67th Annual Southeastern GSA section meeting*, vol. 50, no. 3, Paper No. 23–10. Knoxville, USA: Geological Society of America, Abstracts with Programs. <https://doi.org/10.1130/abs/2018SE-312314>.
- Schüller, K. and J. Kowalski. 2017. Spatially Varying Heat Flux Driven Close-Contact Melting-A Lagrangian Approach. *International Journal of Heat and Mass Transfer* 115: 1276–1287. <https://doi.org/10.1016/j.ijheatmasstransfer.2017.08.092>.
- Schüller, K., and J. Kowalski. 2019. Melting Probe Technology for Subsurface Exploration of Extraterrestrial Ice-Critical Refreezing Length and the Role of Gravity. *Icarus* 317: 1–9. <https://doi.org/10.1016/j.icarus.2018.05.022>.
- Schüller, K., J. Kowalski, and P. Råback. 2016. Curvilinear Melting-A Preliminary Experimental and Numerical Study. *International Journal of Heat and Mass Transfer* 92: 884–892. <https://doi.org/10.1016/j.ijheatmasstransfer.2015.09.046>.
- Sekine, Y., T. Shibuya, F. Postberg, et al. 2015. High-Temperature Water-Rock Interactions and Hydrothermal Environments in the Chondrite-Like Core of Enceladus. *Nature Communications* 6: 8604. <https://doi.org/10.1038/ncomms9604>.
- Sherwood, B. 2016. Strategic Map for Exploring the Ocean-World Enceladus. *Acta Astronautica* 126: 52–58. <https://doi.org/10.1016/j.actaastro.2016.04.013>.
- Siegert, M.J. and M.C. Kennicutt. 2018. Governance of the Exploration of Subglacial Antarctica. *Frontiers in Environmental Science* 6: 103. <https://doi.org/10.3389/fenvs.2018.00103>.
- Siegert, M.J., J.C. Priscu, I.A. Alekhina, et al. 2016. Antarctic Subglacial Lake Exploration: First Results and Future Plans. *Philosophical Transactions of the Royal Society A* 374: 20140466. <https://doi.org/10.1098/rsta.2014.0466>.

- Simonsen, L.C., and J.E. Nealy. 1993. *Mars Surface Radiation Exposure for Solar Maximum Conditions and 1989 Solar Proton Events*. Technical Report 3300
- Spencer, J.R., F. Nimmo, A.P. Ingersoll, et al. 2018. Plume Origins and Plumbing: From Ocean to Surface. In *Enceladus and the Icy Moons of Saturn*. University of Arizona Press, pp. 163–174. https://doi.org/10.2458/azu_uapress_9780816537075-ch008.
- Stevenson, A., J. Burkhardt, C.S. Cockell, et al. 2015. Multiplication of Microbes Below 0.690 Water Activity: Implications for Terrestrial and Extraterrestrial Life. *Environmental Microbiology* 17(2): 257–277. <https://doi.org/10.1111/1462-2920.12598>.
- Stone, W., K. Richmond, C. Flesher, et al. 2018. Sub-ice Autonomous Underwater Vehicle Architectures for Ocean World Exploration and Life Search. In *Outer Solar System: Prospective Energy and Material Resources* ed. V. Badescu and K. Zacny. Springer, pp. 429–541. https://doi.org/10.1007/978-3-319-73845-1_10.
- Suzuki, Y. 1976. Deep Core Drilling by Japanese Antarctic Research Expeditions. In *Ice-core Drilling: Proceedings of a Symposium* ed. J.F. Splettstoesser, pp. 155–166. Lincoln, NE, USA: University of Nebraska Press.
- Talalay, P.G. 2020. *Thermal Ice Drilling Technology*. Springer.
- Talalay, P.G., Y. Li, M.A. Sysoev, et al. 2019. Thermal Tips for Ice Hot-Point Drilling: Experiments and Preliminary Thermal Modeling. *Cold Regions Science and Technology* 160: 97–109. <https://doi.org/10.1016/j.coldregions.2019.01.015>.
- Talalay, P.G., A. Markov, and M. Sysoev. 2013. New Frontiers of Antarctic Subglacial Lakes Exploration. *Geography, Environment, Sustainability* 6(1): 14–28. <https://doi.org/10.24057/2071-9388-2013-6-1-52-59>.
- Talalay, P.G., V.S. Zagorodnov, A.N. Markov, et al. 2014. Recoverable Autonomous Sonde (RECAS) for Environmental Exploration of Antarctic Subglacial Lakes: General Concept. *Annals of Glaciology* 55(65): 23–30. <https://doi.org/10.3189/2014AoG65A003>.
- Tarnas, J.D., J.F. Mustard, B.S. Lollar, et al. 2021. Earth-like Habitable Environments in the Sub-surface of Mars. *Astrobiology* 21(7): 16. <https://doi.org/10.1089/ast.2020.2386>.
- Tibcken, M. and W. Dimmler. 1997. Einsatz einer Durchschmelzsonde (SUSI) zum Transport einer kommerziellen CTD-Sonde unter das Schelfeis. Die Expedition ANTARKTIS-XII mit FS "Polarstern" 1995. In: Bericht vom Fahrabschnitt ANT-XII/3, Berichte zur Polarforschung 219, pp. 106–112. https://doi.org/10.2312/BzP_0219_1997.
- Treffer, M., N.I. Kömle, G. Kargl, et al. 2006. Preliminary Studies Concerning Subsurface Probes for the Exploration of Icy Planetary Bodies. *Planetary and Space Science* 54(6): 621–634. <https://doi.org/10.1016/j.pss.2006.02.001>.
- Tüg, H. 2002. Private Communication.
- Tüg, H. 2003. Rechnergesteuerte Schmelzsonde zur Ermittlung unterschiedlicher Messparameter im Eisbereich. *Patentschrift DE* 101(64): 648.
- Ulamec, S., J. Biele, J. Drescher, et al. 2005. A Melting Probe with Applications on Mars, Europa and Antarctica. In *56th International Astronautical Congress*. IAC-A1.7.08. Fukuoka, Japan. <https://doi.org/10.2514/6.IAC-05-A1.7.08>.
- Ulamec, S., J. Biele, O. Funke, et al. 2007. Access to Glacial and Subglacial Environments in the Solar System by Melting Probe Technology. *Reviews in Environmental Science and Biotechnology* 6(1): 71–94. <https://doi.org/10.1007/s11157-006-9108-x>.
- Ulamec, S., J. Biele, and E. Trollope. 2010. How to Survive a Lunar Night. *Planetary and Space Science* 58(14): 1985–1995. <https://doi.org/10.1016/j.pss.2010.09.024>.
- United Nations Office for Outer Space Affairs. 1966. Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies. <http://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/introouterspacetreaty.html>. (visited on 01/09/2021).
- Winebrenner, D.P., W.T. Elam, P.M.S. Kintner, et al. 2016. Clean, Logistically Light Access to Explore the Closest Places on Earth to Europa and Enceladus. In *AGU Fall Meeting Abstracts*, vol. 2016, C51E-08. <https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/156280>. (visited on 01/09/2021).

- Winebrenner, D.P., W.T. Elam, V. Miller, et al. 2013. A Thermal Ice-Melt Probe for Exploration of Earth-Analogs to Mars, Europa and Enceladus. In *44th Lunar and Planetary Science Conference*, Abstract #2986. <https://doi.org/10.1029/2011GL048846>.
- Wirtz, M., and M. Hildebrandt. 2016. *IceShuttle Teredo: An Ice-Penetrating Robotic System to Transport an Exploration AUV into the Ocean of Jupiter's Moon Europa*. Mexico: Guadalajara.
- Zacny, K., J. Mueller, T. Costa, et al. 2018. SLUSH: Europa Hybrid Deep Drill. In: *2018 IEEE Aerospace Conference*, pp. 1–14. <https://doi.org/10.1109/AERO.2018.8396596>.
- Zimmerman, W., R. Bonitz, and J. Feldman. 2001. Cryobot: An Ice Penetrating Robotic Vehicle for Mars and Europa. In *IEEE Aerospace Conference Proceedings*, pp. 1311–1323. <https://doi.org/10.1109/aero.2001.931722>.

Part VII
Economics and Policies

Chapter 30

Lunar Ore Reserves Standards 101 (LORS-101)



**Carlos Daniel Espejel, Sophia Casanova, Serkan Saydam,
and Julien-Alexandre Lamamy**

Abstract An important topic to be addressed prior to and during space-resource utilization activities is the creation, establishment, use and updating of standards for the estimation and public reporting of exploration results (resources oriented), space-resources evaluations, and space-reserves estimation. This work introduces the current state of development of the Lunar Ore Reserves Standards (LORS-101). These standards aim to provide consistent frameworks, definitions, and guidelines for the estimation and reporting of space resources, to any commercial and non-commercial entity interested in the investment, exploration, extraction, commercial transaction, and use (commercial and operational) of space resources (mineral, volatiles, etc.) from different celestial bodies (moons, planets, asteroids, comets, etc.). The LORS-101 classification frameworks consider geological certainty, technical certainty, and technology maturity, together with socio-political, governance, environmental and economic viability. LORS-101 is based on existing and very mature terrestrial standards for the mining and oil and gas industries, as well as the United Nations resource classification system. LORS-101 is intended to become a collaborative, inclusive and iterative effort which involves a number of expert individuals, organizations, and research institutions for the iterative development of this important and pivotal document.

C. D. Espejel (✉) · J.-A. Lamamy
ispace Europe SA, Rue de l'Industrie 5, Paul Wurth InCub, 1811 Luxembourg, Luxembourg
e-mail: c-espejel@ispace-inc.com

J.-A. Lamamy
e-mail: j-lamamy@ispace-inc.com

S. Casanova · S. Saydam
The School of Minerals and Energy Resources Engineering, UNSW Sydney, Kensington,
NSW 2052, Australia
e-mail: s.casanova@unsw.edu.au

S. Saydam
e-mail: s.saydam@unsw.edu.au

30.1 Terrestrial Standard Codes for the Resource Extractive Industries

Background

Standard codes have been developed setting the minimum standards for public reporting and classification of exploration results, resources and reserves for the mining, and oil and gas industries, which hold some of the most important economic natural resources (extractive) on Earth. The classification systems under these codes consider the different possible levels of confidence in geological knowledge, technical parameters, economic parameters, and commerciality (AusIMM 2012; PRMS 2018).

These terrestrial codes were mainly developed to achieve consistency in a number of areas—including, terminology, definitions, guidelines, and resource classification—for investors and their advisors in the mining, and oil and gas sectors (AusIMM 2012).

A list of definitions was first agreed, which would become the foundation language for reporting codes and their classification systems (Casanova and Espejel 2019).

The main standard codes on which the Lunar Ore Reserve Standards (LORS-101) are heavily based, with their associated institutions, are as follows:

- Joint Ore Reserves Committee (JORC)—AusIMM (Australian Institute of Mining and Metallurgy)
- National Instrument (NI) 43–101—CIM (The Canadian Institute of Mining, Metallurgy and Petroleum), OSC (Ontario Securities Commission)
- International Reporting Template—Committee for Mineral Reserves International Reporting Standards (CRIRSCO)
- Petroleum Resources Management System (PRMS)—Society of Petroleum Engineers
- United Nations Framework Classification for Fossil Energy and Mineral Reserves and Resources 2009 (UNFC-2009)—United Nations (UN)

Standard codes are designed to address specific types of resources, specific geological formations, specific types of extraction and beneficiation methodologies, and are therefore for specific resources industries. For example, JORC, NI 43–101, and CRISCO were developed specifically for hard rock (metalliferous) mining projects; PRMS, on the other hand, was developed specifically for oil and gas projects.

UNFC-2009, in contrast to the other codes, was developed by the UN to provide nations with a tool to report their total national resources and reserves assets, rather than a tool orientated for investors (UNECE 2009).

30.2 Lunar Ore Reserve Standards (LORS-101)

Motivation

Similar to the development of standard codes for the terrestrial resources industries, the space resources industry (SRI) needs to develop its own standard codes to address the uniqueness and complexity of this new industry with unique off-earth (space) geological formations, extreme space environment challenges, space-qualified technologies for extraction and beneficiation, international space resources regulations, and the very rapidly emerging space investment opportunities (market).

Another important factor is the current need for a fundamental tool to provide future space resources (SR) investors with a minimum set of standards and regulations that will allow them to properly evaluate SR projects with consistency, clarity, and transparency in the data, results, and most importantly, reporting. This will reduce risk and create trust and confidence in their decisions and investments.

The Lunar Ore Reserve Standards (LORS) were presented for the first time on 9 October 2019, during the Space Resources Week at the Space Mining Summit in Luxembourg. LORS is a joint collaborative project funded by the FNR (Luxembourg Research Fund) and ispace, with additional support from the University of Luxembourg and the University of New South Wales (UNSW Sydney), and developed by Dr Carlos Espejel Garcia.

Intended beneficiaries

LORS is a collaborative, iterative, inclusive, and joint effort involving input from experts in a number of areas, as well as support and input from a number of institutions, and organizations.

Intended use

LORS is intended to be an instrument providing guidance for consistent reporting through a set of classification systems and guidelines (that are not prescriptive), for any investor, space agency, government, commercial entity, or any other entity wishing to explore, extract and utilize space resources. Although the name LORS-101 refers to the Earth's natural satellite, the code aims eventually to include such other space bodies as asteroids, comets, planets, moons and other resources available in space (e.g., solar wind, light, etc.).

LORS will be reassessed and updated periodically in accordance with the LORS Committee, which is made up of expert individuals and representatives from space agencies, private industry (mining, and oil and gas), research institutions, and other relevant entities.

30.2.1 LORS-101 Structure

LORS is in three parts: Classification Systems, Definitions, and Guidelines. The LORS Classification Systems are frameworks designed to evaluate the geological, technological, technical and economic, and commercial certainty of a space resources utilization (SRU) project. The LORS Definitions is a glossary of the main vocabulary and definitions used throughout the LORS Classification Systems and the LORS Guidelines. The glossary is a live document which will be updated as the LORS Classification Systems and Guidelines are updated in agreement with the LORS Committee. The LORS Guidelines will be a major component of LORS, containing detailed instructions on estimating and reporting SR exploration results, space resources, and reserves, for a number of different types of resources (e.g., water ice, H, O, Ti, Fe, Al, REE, etc.) located on different celestial bodies (moons, planets, asteroids, etc.). This major component will be slowly built up through a number of research projects already completed and in progress worldwide.

30.2.2 LORS Classification Systems

The three main SR classification systems of LORS, which cover most aspects of the reporting of space resources exploration results, and space resources for commercial and non-commercial purposes, are (Fig. 30.1):

1. Geological resources (GR-CS1)
2. Extractable resources or GTX resources (GTX-CS2)
3. Economically extractable reserves or GTE resources (GTE-CS3).

30.2.2.1 Geological Resources Classification System (GR-CS1)

This classification system is designed to measure geological certainty only, and to be used as a guide for the reporting of space geological resources throughout the entire exploration process (see Fig. 30.1).

This system is mainly used for scientific, geoscience, and surveying purposes, where an entity is defining the existence, geometry and distribution of a specific type of resource (e.g., water ice) throughout exploration activities (prospecting and ground truth exploration). *Geological certainty* provides the confidence or certainty level in the knowledge of several geological features in a specific area of exploration. This knowledge is gathered through the acquisition of surface and in-situ exploration data (see Fig. 30.1). Geological features include geological structures, mineral/elemental composition, mineral/elemental distribution, material density, and material hardness.

Through this system (GR-CS1) an entity will be able to measure and report the level of certainty of a space resources exploration project, as well as to define how much further exploration work is needed to achieve a desired level of geological

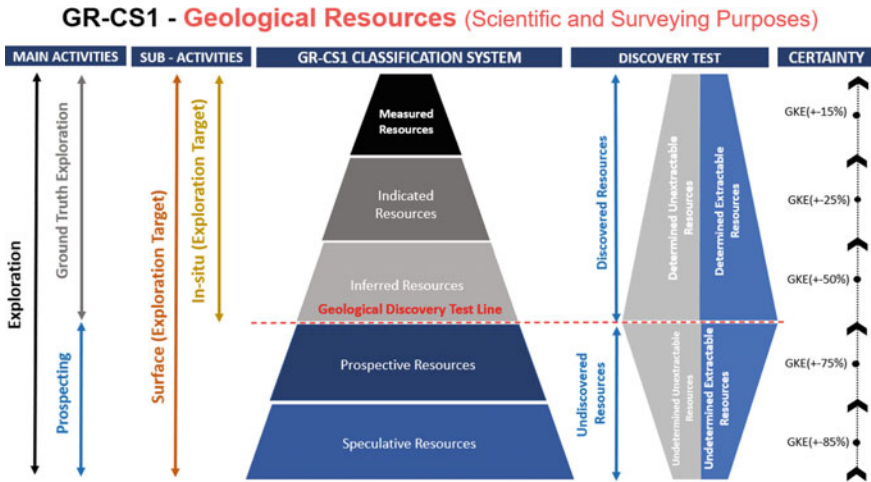


Fig. 30.1 Geological Resources Classification System (GR-CS1). For Scientific, Geosciences, and Surveying Purposes

Table 30.1 Geological Knowledge Error (suggested) in GR-CS1 (Geological Resources)

LORS classification	Geological Knowledge Error (GKE)		
	Min (%)	Mid (%)	Max (%)
Measured Resources	0	15	19
Indicated Resources	20	25	29
Inferred Resources	30	50	69
Prospective Areas	70	75	79
Speculative Regions	80	85	90

certainty. Table 30.1 shows the Geological Knowledge Error (GKE) ranges for every classification level in the Geological Resources (GR-CS1) classification system.

30.2.2.2 Extractable Resources Classification System (GTX-CS2)

This classification system is designed to measure geological, technical and technological certainty only, and to be used as a guide for the reporting of geologically, technically and technologically extractable (GTX) resources during an exploration campaign and during technical evaluations (see Fig. 30.2).

This system is used to report non-sales usable resources for entities wishing to use space resources for operational purposes only, not for commercial or sales activities, such as a space agency wishing to use lunar water to refuel rockets, at space stations, or at permanent bases on the Moon, for operational and survival purposes, but not for commercial activities. In this example, a space agency is extracting

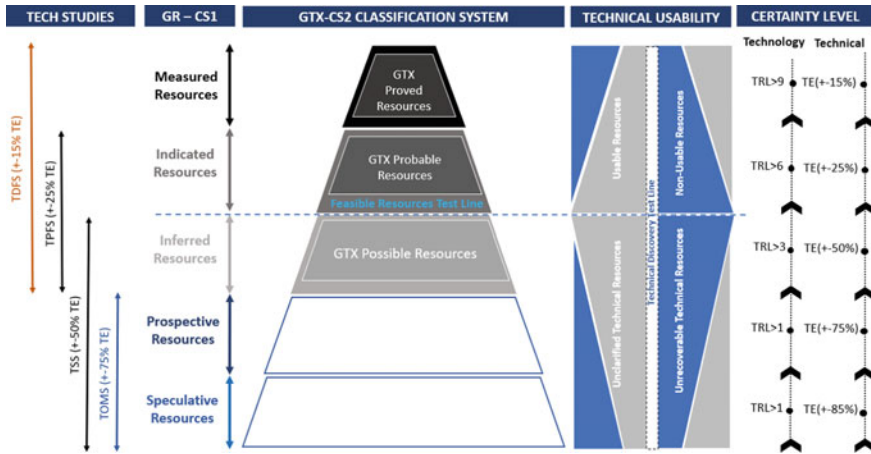


Fig. 30.2 Extractable (GTX) Resources Classification System (GTX-CS 2)

and using resources through their own means, not procuring them from a private company. Usable resources are owned and used for operational purposes, and not for commercial purposes.

This classification system measures the *technological certainty* of the technology to be employed for extraction, processing, and refining of usable resources, and the supply of the final product through the technology readiness level (TRL) system developed by NASA (NASA 2012) (see Fig. 30.2).

Through this system (GTX-CS2), an entity will be able to measure and report the level of geological, technological, and technical certainty of space resources projects, as well as how much further technical work, technological development, and resources exploration needs to be completed to achieve the desired or expected level of overall project certainty. Table 30.2 shows the Technical Error (TE) and TRL ranges for every classification level in the Extractable Resources (GTX-CS2) Classification System.

Please refer to the LORS Definitions section for detailed explanation of the vocabulary used in this classification system.

Table 30.2 (GTX-CS2) Technical Error and TRL Matrix

GTX classification system	Technical Error (TE)			Technology — TRL		
	Min (%)	Mid (%)	Max (%)	Min	Mid	Max
GTX Proved Resources	0	15	19	8	9	9
GTX Probable Resources	20	25	29	5	6	7
GTX Possible Resources	30	50	69	2	3	4

Technical certainty is the ability and confidence of a space resources project to supply a specific amount of resources, at a specific quality, rate, and efficiency, with assumed technologies and methodologies for extraction, processing, refining and supply.

Technical certainty is gained through Technical Studies (TDSS, TPFS, TSS, and TOMS), whose objective is to mature the technical feasibility of an SR project.

Technical Studies allow the estimation of size and quality of usable resources, enabling an entity to identify and decide if further investment should go into a space resources project, or if it should go into operation or be stopped, as well as whether a further resource exploration should be planned, and/or how big it should be. Technical Studies allow the selection of optimum technologies and methodologies for extraction, processing, and refining, to supply a specific required amount of resources and/or products. The result of a technical study might suggest different technologies, different methodologies, different capacities, different capabilities, and further resource exploration programs. These studies require optimization models of the entire value chain to be created, as well as 3D models of the geological resources.

30.2.2.3 Economically Extractable Reserves Classification System (GTE-CS3)

This classification system measures geological, technical, technological, and economic certainty, and is used as a guide for the reporting of geologically, technically, technologically, and economically extractable (GTE) resources during resource exploration programs and technical and economic evaluations (see Fig. 30.3).

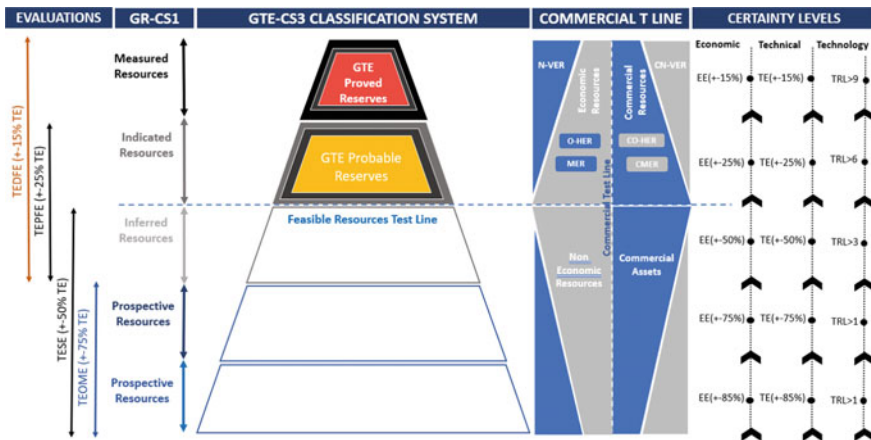


Fig. 30.3 Economically Extractable (GTE) Resources Classification System (GTE-CS3)

Table 30.3 (GTE-CS3) Economic Error, Technical Error and TRL Matrix

GTE classification system	Economic Error (EE)			Technical Error (TE)			Technology — TRL		
	Min (%)	Mid (%)	Max (%)	Min (%)	Mid (%)	Max (%)	Min	Mid	Max
GTE Proved Reserves	0	15	19	0	15	19	8	9	9
GTE Probable Reserves	20	25	29	20	25	29	5	6	7

This system is used to report economic and commercial (sales) resources by entities using space resources purely for commercial purposes. An example of commercial activity is a private company wishing to extract lunar water to sell as fuel for rockets, or to space stations for their operational and survival purposes.

This classification system measures the geological certainty, technical certainty and economic certainty of a space resources project. It also measures the maturity (TRL) of the technology (technological certainty) to be employed for extraction, processing, refining, and supply of the final product (see Fig. 30.3).

Economic certainty, measured through the Economic Error (EE), is the ability and confidence of a space resources project to be economic at a specific Net Present Value (NPV), Internal Rate of Return (IRR), Return of Investment (ROI), or any other economic indicator (see Fig. 30.3 and Table 30.3).

Economic certainty is gained through technical and economic evaluations (TEDFE, TEPFE, TESE, and TEOME), the objectives of which are to assess and mature the technical and economic feasibility of a space resources project.

Technical and Economic Evaluations (TEEs) allow the total economic value of a space resources project, and the size and quality of the economic and/or commercial resources, to be estimated. TEEs enable an entity to identify and decide if a project is economically feasible, economically attractive, and whether further investment should go into the selected, it should be put into operation, or stopped.

TEEs estimate whether a further resource exploration program should be carried out, and/or how big it should be. TEEs also estimate the total operational costs (OPEX), capital expenditure (CAPEX), fixed costs (FC), and economic benefits of a space resources project (NPV, IRR, ROI, etc.). The optimum technologies, and methodologies for extraction, processing, refining, and supplying a required and specific amount of resources and/or products can also be selected.

The result of a TEE might suggest different economic alternatives (sequences), different technologies, different methodologies, different capacities, different capabilities, and further resource exploration programs. These studies require the creation of optimization models of the entire value chain, complex economic and cost models, as well as 3D models of the geological resources.

Table 30.3 shows the Economic Error (TE), Technical Error (TE) and TRL ranges for every classification level in the GTE Classification System 3.

Please refer to the LORS Definitions section for a detailed explanation of the vocabulary used in this classification system.

30.2.3 LORS Definitions

Intended use

The *LORS Definition* is a Space Resources (SR) glossary containing the main vocabulary and definitions used throughout the LORS Classification Systems and the LORS Guidelines. It is a live document which will be updated as the LORS Classification Systems and Guidelines are updated in agreement with the LORS Committee.

Standard codes for major terrestrial resources begin with an agreement on the definitions forming the basis of the resource and reserves classification systems, as well as for the resulting reporting guidelines (Casanova and Espejel 2019). Space agencies, mining industry, the oil and gas industry, research institutions and other key contributing entities are encouraged to be involved, in addition to expert individuals.

Motivation

There is a current need within the scientific community, industry, government, and other entities currently working in the field of ISRU, SRU, and SR, to use the same agreed vocabulary for consistency, transparency and reportability purposes.

Intended beneficiaries

The LORS Definitions are intended to be used by any entity that would like to estimate and report exploration results, resources and reserves estimations. This will allow reporting entities to use agreed vocabulary for transparency and consistency purposes.

The entities that will benefit from this are those that make decisions (technical, financial, etc.) on exploration results, resources and reserves reports. Decision-making entities (e.g., investors, stakeholders, etc.) need consistency and transparency in reports for clear and fair comparison.

30.2.3.1 Geological Resources Classification System (GR-CS1) Definitions

Exploration

Exploration in the context of LORS is the name given to all data- and sample-gathering activities performed with the intention of obtaining geological knowledge (resources oriented) of an identified area/body. Through exploration activities, the geological certainty of a resource project will be raised from *Speculative Resources* (lowest geological knowledge level) to *Measured Resources* (highest geological knowledge level); see Fig. 30.1. Exploration is carried out through two main activities: *Prospecting* and *Ground Truth Exploration*. Geological knowledge is commenced and increased through the collection of individual *Surface Exploration Targets* (SETs) and *In-Situ Exploration Targets* (IETs).

Discovered Resources

According to LORS, all *Inferred Resources*, *Indicated Resources* and *Measured Resources* with a Geological Knowledge Error (GKE) of $\pm 50\%$ and below, are classified as *Discovered Resources*. These resources fall above the *Geological Discovery Test Line*.

Geological Certainty

Geological Certainty is the confidence or certainty level in the knowledge of several geological features in a specific area of exploration. This knowledge is gathered through the acquisition of surface and in-situ exploration data (see Fig. 30.1). Geological features may include geological structures, mineral/elemental composition, mineral/elemental distribution, material density, and material hardness.

Geological Discovery Test Line

According to LORS, the *Geological Discovery Test Line (GDTL)* is a hard boundary that divides from undiscovered resources. *Prospective Areas* and *Speculative Regions* resources with a Geological Knowledge Error $\pm 75\%$ are considered below the *GDTL*.

Ground Truth Exploration

Ground Truth Exploration in the context of LORS is the name given to one of the main exploration activities which focuses on the gathering of geological knowledge (resource oriented), through the collection—by robotic systems or by human specialists—of physical samples, including surface and subsurface (in-situ) samples.

These samples are collected through *Surface Exploration Targets (SETs)* and *in-situ exploration targets (IETs)*. Ground truth exploration is the activity used to raise geological knowledge/certainty (resource oriented) of an SR project from *Inferred Resources* to *Measured Resources* (see Fig. 30.1).

Indicated Resources

In LORS, *Indicated Resources* within the Geological Resources (SR-CF1) Classification System is the second-highest classification level in terms of geological knowledge (certainty) that a project can have, with an error range of $\pm 20\%$ to $\pm 29\%$.

Surface Exploration Targets (SETs) and *In-situ Exploration Targets (IETs)* are employed in ground truth exploration campaigns to obtain a higher level of geological certainty for *Inferred Resources* and above. A large amount of IETs will be needed for the estimation of the in-situ geological continuity and grade continuity.

Indicated Resources have medium to medium–high knowledge for the estimation of mineral content, grades (quality), quantity, economic characteristics, including physical parameters (e.g., density, hardness, etc.). The geological evidence leads to verified geological and/or grade (quality) continuity (AusIMM 2012) at medium or medium–high level of confidence.

Inferred Resources

In LORS, Inferred Resources is the third-lowest classification level within the Geological Resources (SR-CF1) Classification System, in terms of geological knowledge (certainty) that a project can have, with an error range of $\pm 30\%$ to $\pm 69\%$.

Surface exploration targets (SETs) and *in-situ exploration targets* (IETs) are employed in ground truth exploration campaigns to reach the geological certainty of *inferred resources* and above. The quantity and/or quality of SETs and IETs may be low.

Inferred Resources have enough knowledge for the estimation of mineral content, grades (quality) and quantity, with low geological confidence. The geological evidence leads to inferred and assumed but not verified geological and/or grade (quality) continuity (AusIMM 2012).

In-situ Exploration Target

An *In-Situ Exploration target (ISET)* is the name given to the exact location (area or coordinate) where an in-situ (subsurface) sample or in-situ remote sensing reading will be taken.

ISETs can either be a physical sample (e.g., drill core), or a remote sensing reading (e.g., downhole probe) that has similar precision to one taken in a physical sample (e.g., physical assay). ISETs are used for ground truth exploration only.

ISETs are used to define the in-situ (subsurface) geological structure and/or distribution (elemental, mineralogical) of a specific resource, in a specific area and depth.

Measured Resources

In LORS, *Measured Resources* within the Geological Resources (SR-CF1) Classification System is the highest classification level in terms of geological knowledge (certainty) that a project can have, with an error range of 0% to $\pm 19\%$.

SETs and ISETs are employed in ground truth exploration campaigns to achieve the highest level of geological certainty for *measured resources*. A large number of ISETs is needed for the estimation of the in-situ geological continuity and grade continuity.

Measured Resources have high knowledge for the estimation of mineral content, grades (quality), quantity, economic characteristics, including physical parameters (e.g., density, hardness, etc.). The geological evidence leads to highly verified geological and/or grade (quality) continuity (AusIMM 2012) at the highest level of confidence.

Prospecting

Prospecting, in the context of LORS, is the name given to one of the main exploration activities, which focuses on the gathering of geological knowledge (resource oriented), only through the collection of remote sensing data and not including physical samples. These data are only collected through surface exploration targets (SETs). *Prospecting* raises the geological knowledge/certainty (resource oriented)

of a space resources project from *Speculative Regions* to *prospective areas* (see Fig. 30.1).

Prospective Areas

In LORS, *Prospective areas* within the Geological Resources (SR-CF1) Classification System is the second-lowest classification level in terms of geological knowledge (certainty) that a project can have, with an error range of $\pm 70\%$ to $\pm 79\%$.

Prospective Areas are strong indicators where interpretative geology is present, using surface physical samples or remote sensing data (surface only). They provide the technical confidence for a viable *in-situ exploration target* (IET). *Prospective areas* have higher spatial resolution than speculative resources.

Speculative Regions

In LORS, *Speculative Regions* within the Geological Resources (SR-CF1) Classification System, is the lowest classification level in terms of geological knowledge (certainty) that a project can have, with an error range of $\pm 80\%$ to $\pm 90\%$.

Speculative regions are indicators (of the presence of anomalies) showing that an element of interest has potential presence, but that geological confidence is poor (low spatial resolution in usually very large areas).

Surface Exploration Target

A *Surface Exploration Target (SET)* is the name given to the exact location (area or coordinate) where a surface (only) sample, or surface remote sensing reading will be taken.

Surface exploration targets can either be a physical sample (e.g., sample grab), or a remote sensing reading (e.g., neutron spectrometer). Geological certainty may differ depending on the technique or instrument used on a SET. SETs are used for *prospecting* and *ground truth exploration*.

SETs are used to define surface (only) geological features such as structure and/or distribution (elemental, mineralogical) of a potential resource, in a specific surface area.

Undiscovered Resources

According to the LORS, all *speculative regions* and *prospective areas* with a geological knowledge error (GKE) of $\pm 75\%$ and above are considered and classified as *Undiscovered Resources*. These resources are below the *Geological Discovery Test Line* and further exploration is needed if the intention is to report *Discovered Resources*.

Undetermined Extractable Resources

According to the LORS, these are *undiscovered resources* that are considered to be extractable with the data, technology, and techniques and methodologies for extraction and processing that are currently available.

Undetermined Unextractable Resources

According to the LORS, these are *undiscovered resources* that are considered to be non-extractable with the data, technology, and techniques and methodologies for extraction and processing that are currently available.

Determined Extractable Resources

According to the LORS, these are *discovered resources* that are considered to be extractable with the data, technology, and techniques and methodologies for extraction and processing that are currently available.

Determined Unextractable Resources

According to the LORS, these are *discovered resources* that are considered to be non-extractable with the data, technology, and techniques and methodologies for extraction and processing that are currently available.

30.2.3.2 Extractable Resources Classification System (GTX-CS2) Definitions

Feasible Resources Test Line

According to the LORS, the *Feasible Resources Test Line (FRTL)* is a hard boundary dividing usable resources and non-usable resources from unclarified technical resources and unrecoverable technical resources (see Fig. 30.2). The FRTL is a line that divides geological resources with high geological (*indicated* and *measured resources*), technological (>TRL3), and technical (TE > ± 19%) certainty (see Fig. 30.2), from resources with lower certainty.

Resources above the FRTL can be regarded as *GTE reserves* (see Fig. 30.3).

GTX Possible Resources

In the LORS, *GTX Possible Resources* within the Extractable Resources (GTX-CF1) Classification System is the third-highest classification level in terms of geological knowledge, technical certainty, and technology maturity, that a resources extraction project can have, with a geological and technical error range of 30% to ±69%, and a TRL equal and above 2.

Technical Scoping Studies (TSS), *Technology Development programs*, and initial (small) resource exploration programs are used to reach the third-highest level of technical, technological and geological certainty (*GTX Possible Resources*). Studies, initial engineering, initial technology development, and relatively small resource exploration campaigns are needed to reach this level of certainty.

GTX Possible Resources are made up only of *inferred resources* (GR-CS1). *GTX possible resources* are considered and can only be reported as non-usable resources due to their low geological, technical and technological certainty.

GTX Probable Resources

In the LORS, *GTX Probable Resources* within the Extractable Resources (GTX-CF1) Classification System is the second-highest classification level in terms of geological knowledge, technical certainty, and technology maturity, that a *usable resources* project can have, with a geological and technical error range of 20% to $\pm 29\%$, and a TRL equal and above 5.

Technical Pre-Feasibility Studies (TDFS), technology development programs, and very large resource exploration programs are used to reach the second-highest level of technical, technological and geological certainty (*GTX Probable Resources*). A number of studies, engineering, technology development, and large resource exploration campaigns are needed to reach this level of certainty.

GTX probable resources are made up only of *indicated resources* (GR-CS1). *GTX probable resources* are considered and can be reported as usable resources due to their high geological, technical and technological certainty.

GTX Proved Resources

In the LORS, *GTX Proved Resources* within the Extractable Resources (GTX-CF1) Classification System is the highest classification level in terms of geological knowledge, technical certainty, and technology maturity, that a *usable resources* project can have, with a geological and technical error range of 0% to $\pm 19\%$, and a TRL equal and above 8.

Technical Detailed Feasibility Studies (TDFS), technology development programs, and detailed resource exploration programs are used to reach the highest level of technical, technological and geological certainty (*GTX proved resources*). A large number of studies, engineering, technology development, and resource exploration campaigns are needed to reach this level of certainty.

GTX proved resources are made only of *measured resources* (GR-CS1). *GTX proved resources* are considered and can be reported as usable resources due to their very high geological, technical and technological certainty.

Non-Usable Resources

According to the LORS, *Non-Usable Resources (N-UR)* are those resources that are not technically, technologically, and geologically extractable, with very high certainty levels.

Indicated Resources and *Measured Resources* with a technical error equal and below $\pm 29\%$, and a TRL equal and higher than 5, are considered and classified as *non-usable resources*. These resources can be reported as *non-usable resources*.

Since these resources fall below the TDTL, further resource exploration, technical studies, technology development, and assessments are needed if the intention is to report *usable resources*.

Technical Certainty

Technical Certainty refers to the ability and confidence of a space resources project to supply a specific amount of resources, at a specific quality, rate, and efficiency,

with assumed technologies and methodologies for extraction, processing, refining and supply.

Technical certainty is gained through technical studies (TDSS, TPFS, TSS, and TOMS), the objective of which is to mature the technical feasibility of a space resources project.

Technical Studies allow the estimation of size and quality of *usable resources*. Technical studies allow an entity to identify and decide if further investment should go into a space resources project, or if the project should go into operation, or be stopped. Technical studies also estimate if further resource exploration should be done, and/or how big a resources exploration program should be. Technical studies allow the selection of optimum technologies, and methodologies for extraction, processing, and refining to supply a required and specific amount of resources and/or products.

The result of a technical study might suggest different technologies, different methodologies, different capacities, different capabilities, and further resources exploration programs. These studies require the creation of optimization models of the entire value chain, as well as 3D models of the geological resources.

Technological Certainty

Technological Certainty measures the maturity of technology to be used in the extraction, processing, refining and supply of usable resources, through the technology readiness level (TRL) system developed by NASA (2012).

Technical Detailed Feasibility Study

A *Technical Detailed Feasibility Study (TDFS)* is a technical evaluation of a proposed space resources project, to assess and determine if the evaluated resource can be technically extracted, processed, refined, and delivered.

TDFSs are the most detailed type of assessment and primarily aim to determine whether to proceed to operation, re-assess, or to stop a space resources project.

A TDFS considers in the highest detail geology, engineering (mainly mining, mechanical, aerospace, space, chemical, and processing), business (finance, and economic), legal, and all other relevant disciplines involved in the successful completion of a space resources project. a detailed and final space resources operations design has been completed at TDFS level.

TDFS considers mainly *measured resources*, *indicated resources*, and *inferred resources*. TDFS considers only technologies with TRL equal and above 8.

TDFS require a large team, large amount of engineering work, and very high geological certainty. The level of accuracy of a TDFS is within an error range of $\pm 15\%$.

Technical Discovery Test Line

According to the LORS, the *Technical Discovery Test Line (TDTL)* is a hard boundary that divides *usable resources* and *unclarified technical resources* from *non-usable resources* and *unrecoverable technical resources* (see Fig. 30.2). The TDTL is a line that separates the resources considered geologically, technically, and technologically extractable (usable), or potentially extractable (potentially usable),

from the resources that are declared non-extractable (non-usable), or potentially non-extractable (potentially non-usable).

Technical Order of Magnitude Study

A *Technical Order of Magnitude Study (TOMS)* is a very low-certainty technical evaluation of a proposed space resources project, to assess and determine if the evaluated resource can be technically extracted, processed, refined, and delivered. TOMS is the lowest level of technical studies according to the LORS, and less detailed than a technical scoping study.

A TOMS is the very first technical study for resources with very low geological certainty, including *prospective areas* and *speculative regions*.

TOMS considers mainly *areas*, and *speculative regions*. TOMS studies consider technologies with TRL equal and above 1.

TOMS is used to determine whether to proceed with a space resources prospecting and/or exploration program, whether to start or continue initial engineering work, and whether or not to go ahead with a technical scoping study.

TOMS considers the lowest detail and very gross estimations of the following disciplines involved in the successful completion of a space resources project at TOMS level: geology, engineering (mainly mining, mechanical, aerospace, space, chemical, and processing), business (finance, and economic), and legal. Plans, figures, costings, and factors from existing projects are used to complete a TOMS. In the absence of relevant information to support the study, rules of thumb and rough estimates can be used for the purposes of the study. A very high-level estimate of potential available space resources is completed at TOMS level.

TOMS requires a very small multi-disciplinary team, or even only one individual. The level of accuracy of a TOMS is within an error range of $\pm 75\%$.

Technical Pre-feasibility Study

A *Technical Pre-Feasibility Study (TPFS)* is a medium-certainty technical evaluation of a proposed space resources project, to assess and determine if the evaluated resource can be technically extracted, processed, refined, and delivered. A TPFS is more detailed than a technical scoping study, but less detailed than a technical detailed feasibility study.

A TPFS is used for due diligence purposes and also to determine whether or not to go ahead with a detailed technical feasibility study. A TPFS will determine which areas of the project require further attention and investigation, and will define where and how much more resources exploration is needed within and around the selected deposit.

TPFS considers mainly *indicated resources*, and *inferred resources*. TPFS studies consider only technologies with TRL equal and above 5.

TPFS considers in medium detail and gross dimensions all relevant disciplines involved in the successful completion of a space resources project, including geology, engineering (mainly mining, mechanical, aerospace, space, chemical, and processing), business (finance and economic), and legal. A space resources operations design is completed at TPFS level.

TPFS requires a small or medium multi-disciplinary group, and the level of accuracy of a TPFS is within an error range of $\pm 25\%$.

Technical Scoping Study

A *Technical Scoping Study (TSS)* is a low-certainty technical evaluation of a proposed space resources project, to assess and determine if the evaluated resource can be technically extracted, processed, refined, and delivered. A TSS is more detailed than a technical order of magnitude study, but less detailed than a *technical pre-feasibility study*.

A TSS is an initial technical study for resources with mid-low geological certainty of inferred resources and below.

TSS considers mainly *inferred resources*, *prospective areas*, and *speculative regions*. TSS studies consider technologies with TRL equal and above 2.

A TSS is used to determine whether to proceed with a space resources exploration program, whether to proceed with detailed engineering work, and whether to go ahead or not with a *technical pre-feasibility study*.

TSS considers low-detail and gross estimations of all relevant disciplines involved in the successful completion of a space resources project at TSS level, including geology, engineering (mainly mining, mechanical, aerospace, space, chemical and processing), business (finance and economic), and legal. Plans, figures, costings, and factors from existing projects are used to complete a TSS. A preliminary *Space Resources Mine Plan* is completed at TSS level.

TSS requires a small multi-disciplinary team, or even one individual. The level of accuracy of a TSS is within an error range of $\pm 50\%$.

Usable Resources

According to the LORS, *Usable Resources (UR)* are those resources that are technically, technologically, and geologically extractable, with very high certainty levels. *Indicated resources* and *measured resources* with a technical error equal and below $\pm 29\%$, and a TRL equal and higher than 5, are considered and classified as *usable resources*. These resources can be reported as *usable resources*.

Unclarified Technical Resources

According to the LORS, *Unclarified Technical Resources (UCTR)* are those resources that are considered technically, technologically, and geologically extractable, with the available geological data, available technology, and available technical knowledge (studies). UCTR have very low geological, technical and technological certainty levels.

Unclarified Technical Resources are made up of *inferred resources*, *prospective areas* and *speculative regions* with a technical error equal and above $\pm 30\%$, and a TRL below 5.

These resources are above the technical discovery test line (TDTL) as preliminary studies and data are showing potential extractability (usability). However, due to its very low certainty levels, further resource exploration, technical studies, technology development, and assessments are needed if the intention is to report *usable resources*.

Unrecoverable Technical Resources

According to the LORS, *Unrecoverable Technical Resources (URTR)* are those resources that are considered not technically, technologically, and geologically extractable, with the available geological data, available technology, and available technical knowledge (studies). URTR have very low geological, technical and technological certainty levels.

Unrecoverable technical resources are made up of *inferred resources*, *prospective areas* and *speculative regions* with a technical error equal and above $\pm 30\%$, and a TRL below 5.

These resources are below the technical discovery test line (TDTL) and further resource exploration, technical studies, technology development, and assessments are needed if the intention is to report *usable resources*.

30.2.3.3 Economic Extractable Reserves Classification System (GTE-CS3) Definitions

Commercial Resources

According to the LORS, it is only the amount of economic resources that have been signed and committed commercially through a contract (sales), that can be reported as *commercial resources*.

Commercial Test Line

According to the LORS, the *Commercial Test Line (CTL)* is a hard dividing line between *commercial resources* and *economic resources*. Only the amount of economic resources that have a signed and committed commercial contract (sales), can be reported as *commercial resources*.

Economic Resources

According to the LORS, the economically extractable part of *GTX probable resources*, and *GTX proved resources* are classified as *economic resources*.

GTE Probable Reserves

According to the LORS, *GTE Probable Reserves* are the economically extractable part of *GTX probable resources*. These are automatically classified as *economic resources*.

GTE Proved Reserves

According to the LORS, *GTE Proved Reserves* are the economically extractable part of *GTX proved resources*. These are automatically classified as *economic resources*.

Marginal Economic Resources

According to the LORS, these are space resources which have marginal economics (pay for all costs), but do not meet the specified financial objectives (NPV, IRR, IROR, etc.).

Non-viable Economic Resources

According to the LORS, these are space resources which are technically and technologically extractable but are not economically extractable. They do not break even, nor meet the specified financial objectives (NPV, IRR, IROR, etc.).

On-Hold Economic Resources

According to the LORS, these are *economic resources* that have not received nor granted all the required legal, social, environmental, or any other license to operate. However, these resources have been deemed economic.

Technical and Economic Detailed Feasibility Evaluation

A *Technical and Economic Detailed Feasibility Evaluation (TEDFE)* is a study of a proposed space resources project, to assess and determine if the evaluated resource can be economically and technically extracted, processed, refined, and delivered.

TEDFE is the most detailed type of economic assessment. Its primary goal is to determine the total value (NPV, IRR, ROI, etc.) of a space resources project, and whether to proceed to operation, re-assess, or to stop a space resources project.

TEDFE considers in the highest detail all expert fields involved in the successful completion of a space resources project, including economic (e.g., market prices), financial (e.g., OPEX and CAPEX cost models), geology, resources engineering, mechanical engineering, space engineering, processing engineering and legal. A detailed and final *Space Resources Mine Design* with financials is completed at *TEDFE* level.

TEDFE considers mainly *measured resources*, *indicated resources*, and *inferred resources*. *TEDFE* considers only technologies with TRL equal to and above 8.

TEDFE require a large team, a large amount of engineering, finance and economics work, very high geological certainty and a high level of confidence. The level of accuracy of a *TEDFE* is within an error range of $\pm 15\%$.

Technical and Economic Order of Magnitude Evaluation

A *Technical and Economic Order of Magnitude Evaluation (TEOME)* is a very low-certainty economic evaluation of a proposed space resources project, to assess and determine if the evaluated resource can be economically and technically extracted, processed, refined, and delivered. *TEOME* is the lowest level of economic and technical studies according to LORS, and less detailed than a technical economic scoping evaluation.

TEOME is the very first economic studies for resources with very low geological certainty, including *prospective areas* and *speculative regions*.

TEOME considers mainly *prospective areas*, and *speculative regions*. TEOMEs consider technologies with TRL equal and above 1.

TEOMEs are used to estimate the total value of a space resources project (NPV, IRR, ROI, etc.), and to determine whether to proceed with a space resources prospecting and/or exploration program, whether to start or continue initial engineering work, and whether or not to go ahead with a technical and economic scoping evaluation.

TEOMEs consider the lowest detail and very gross estimations of all expert fields involved in the successful completion of a space resources project at this level, including economic (e.g., market prices), financial (e.g., OPEX and CAPEX cost models), geology, resources engineering, space engineering, processing engineering, and legal. Plans, figures, costings (OPEX and CAPEX), and factors from existing projects are used to complete a TEOME. In the absence of relevant information to support the study, rules of thumb and rough estimates can be employed for the purposes of the study. A very high-level estimation of potential available space resources and their potential economic value is completed at this level.

TEOMEs require a very small multi-disciplinary team, or even a single individual. The level of accuracy of a TEOME is within an error range of $\pm 75\%$.

Technical and Economic Pre-feasibility Evaluation

A *Technical and Economic Pre-Feasibility Evaluation (TEPFE)* is a medium-certainty study of a proposed space resources project, to assess and determine if the evaluated resource can be economically and technically extracted, processed, refined, and delivered. A TEPFE is more detailed than a technical and economic scoping evaluation, but less detailed than a technical and economic detailed feasibility evaluation.

TEPFEs are used for due diligence purposes, to estimate the value of a space resources project (NPV, IRR, ROI, etc.), and to assess whether to go ahead or not with a technical and economic detailed feasibility evaluation. A TEPFE will determine which areas of the project require further attention and investigation, and will define where and how much more resources exploration is needed within and around the selected deposit.

TEPFE considers mainly *indicated resources* and *inferred resources*. TEPFE studies consider only technologies with TRL equal and above 5.

TEPFE considers in medium detail and gross dimensions all expert fields involved in the successful completion of a space resources project, including economic (e.g., market prices), financial (e.g., OPEX and CAPEX cost models), geology, resources engineering, mechanical engineering, space engineering, processing engineering and legal. A Space Resources Mine Design with financials is completed at TPF level.

This type of evaluation requires a small or medium-sized multi-disciplinary group, and the level of accuracy of a TEPFE is within an error range of $\pm 25\%$.

Technical and Economic Scoping Evaluation

A *Technical and Economic Scoping Evaluation (TESE)* is a low-certainty economic study of a proposed space resources project, to assess and determine if the evaluated resource can be economically and technically extracted, processed, refined, and delivered. A TESE is more detailed than a technical and economic order of magnitude evaluation, but less detailed than a technical and economic pre-feasibility evaluation.

TESEs are initial economic studies for resources with mid-low geological certainty of *inferred resources* and below.

TESEs consider mainly *inferred resources*, *prospective areas*, and *speculative regions*. TSS studies consider technologies with TRL equal and above 2.

TESEs are used to determine the total value of a space resources project (NPV, IRR, ROI, etc.), and whether to proceed with a space resources exploration program, whether to proceed with detailed engineering work, and whether to go ahead or not with a technical and economic pre-feasibility evaluation.

TESEs consider low-detail and gross estimations of all expert fields involved in the successful completion of a space resources project at TESE level, including economic (e.g., market prices), financial (e.g., OPEX and CAPEX cost models), geology, resources engineering, space engineering, processing engineering, and legal. Existing plans, figures, costings (OPEX and CAPEX), and factors from existing projects are used to complete a TESE. A preliminary Space Resources Mine Plan with financials is completed at TESE level.

TESEs require a very small multi-disciplinary team, or even a single individual. The level of accuracy of a TESE is within an error range of $\pm 50\%$.

30.2.3.4 Other Definitions

ISRU: In-situ Space Resource Utilization (NASA [2020](#)).

GP: Guidelines Project.

SR: Space Resources.

SRI: Space Resources Industry.

SRU: Space Resources Utilization.

Surface Material: Material exposed to the environment, on the surface of an object.

30.2.4 LORS Guidelines

The LORS Guidelines will be a major component of LORS, which is designed to give detailed direction on how to estimate and report SR exploration results, space resources, and space reserves, for a number different types of resources (e.g., water ice, H, O, Ti, Fe, Al, REE, etc.) located on different celestial bodies (moons, planets,

asteroids, etc.). This major component is intended to be slowly built by a number of research projects already completed and awaiting completion worldwide.

30.2.4.1 LORS Guidelines Structure

The LORS Guidelines will be made up of a number of projects (e.g., LORS-GP1), as seen in Table 30.4, where GP stands for Guidelines Project. Each circle (o) represents a potential project that will feed the content of the Guidelines. These may be managed by different research institutions, or any other interested entity that the LORS Steering Committee deems appropriate.

Table 30.4 Guidelines potential projects matrix

	Types of resources	Exploration		Resource estimation	Reserves estimation	Operations
		Prospecting	Ground truth exploration			
Moon	Water ice (e.g., permafrost)	LORS-GP1	LORS-GP2	LORS-GP3	LORS-GP4	LORS-GP5
	H ₂ O in minerals	LORS-GP6	LORS-GP7	LORS-GP8	LORS-GP9	LORS-GP10
	H ₂ O in glasses	LORS-GP11	LORS-GP12	LORS-GP13	LORS-GP14	LORS-GP15
	Volatiles (H, O, Helium-3)	LORS-GP16	LORS-GP17	LORS-GP18	LORS-GP19	LORS-GP20
	Metals (Fe, Ti, Al)	LORS-GP21	LORS-GP22	LORS-GP23	LORS-GP24	LORS-GP25
	RRE (Rare earth elements)	LORS-GP26	LORS-GP27	LORS-GP28	LORS-GP29	LORS-GP30
	Radioactive elements (Th, U, etc.)	LORS-GP31	LORS-GP32	LORS-GP33	LORS-GP34	LORS-GP35
	Others	LORS-GP36	LORS-GP37	LORS-GP38	LORS-GP39	LORS-GP40
Mars	Water ice (e.g., permafrost)	O	O	O	O	O
	H ₂ O in minerals	O	O	O	O	O

(continued)

Table 30.4 (continued)

	Types of resources	Exploration		Resource estimation	Reserves estimation	Operations
		Prospecting	Ground truth exploration			
	H ₂ O in glasses	O	O	O	O	O
	Volatiles (H, O, Helium-3)	O	O	O	O	O
	Metals (Fe, Ti, Al)	O	O	O	O	O
	RRE (Rare earth elements)	O	O	O	O	O
	Radioactive elements (Th, U, etc.)	O	O	O	O	O
	Others	O	O	O	O	O
Asteroids	Water ice (e.g., permafrost)	O	O	O	O	O
	H ₂ O in minerals	O	O	O	O	O
	H ₂ O in glasses	O	O	O	O	O
	Volatiles (H, O, Helium-3)	O	O	O	O	O
	Metals (Fe, Ti, Al)	O	O	O	O	O
	RRE (Rare earth elements)	O	O	O	O	O
	Radioactive elements (Th, U, etc.)	O	O	O	O	O
	Others	O	O	O	O	O
Other celestial bodies	Water ice (e.g., permafrost)	O	O	O	O	O

(continued)

Table 30.4 (continued)

	Types of resources	Exploration		Resource estimation	Reserves estimation	Operations
		Prospecting	Ground truth exploration			
	H ₂ O in minerals	O	O	O	O	O
	H ₂ O in glasses	O	O	O	O	O
	Volatiles (H, O, Helium-3)	O	O	O	O	O
	Metals (Fe, Ti, Al)	O	O	O	O	O
	RRE (Rare earth elements)	O	O	O	O	O
	Radioactive elements (Th, U, etc.)	O	O	O	O	O
	Others	O	O	O	O	O

References

- AusIMM. The JORC Code, Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves. 2012 ed. 2012. http://www.jorc.org/docs/JORC_code_2012.pdf. Accessed 19 February 2021.
- Casanova, S., and C. Espejel. 2019. Developing the Frameworks, Processes and Techniques to Evaluate the Commercial Viability of Off-Earth Resource Projects, in *Space Resources Round Table*. Colorado School of Mines, in Golden, CO. 2019.
- CRIRSCO. 2013. CRIRSCO International Reporting Template For Exploration Results, Mineral Resources And Mineral Reserves. http://www.crirSCO.com/templates/international_reporting_template_november_2013.pdf. Accessed 19 February 2021.
- Espejel, C. Lunar Ore Reserves Standards 101 (LORS-101), a First Code for the Reporting of Lunar Exploration Results, Lunar Resources, and Lunar Reserves. International Future Mining Conference. 2019. *AusIMM*, 2019. Sydney: Australia.
- NASA. 2020. In-Situ Space Resource Utilization (ISRU). <https://www.nasa.gov/isru>. Accessed 21 February 2021.
- NASA. Technology Readiness Level. 2012. https://www.nasa.gov/directorates/heo/scan/engineering/technology/txt_accordion1.html. Accessed 21 February 2021.
- SPE. 2018. Petroleum Resources Management System. <https://www.spe.org/en/industry/petroleum-resources-management-system-2018/>. Accessed 19 February 2021.
- UNECE. 2009. United Nations Framework Classification for Fossil Energy and Mineral Reserves and Resources 2009 (UNFC-2009). https://www.unece.org/energy/se/unfc_2009.html. Accessed 19 February 2018.

Chapter 31

The Economics of Space Resources: Future Markets and Value Chains



Mathias Link, Gary Martin, and Joseph Mousel

Abstract This chapter explores the potential benefits of extending the world's economic sphere from Earth deeper into space. The sustainability of exploration and the further development of the space economy are dependent on the creation of commercial markets based on the utilization of space resources. This chapter first describes the fundamental value chain for space resources and its potential for creating new added value in the future. It then explores how the rich history and experience of using resources on Earth can help the birth of a new space resources industry, with examples of innovative commercial space companies that are laying the foundation for a strong future industrial sector in space. The chapter concludes with a review of the important role of governments in nurturing the new space resources industry by describing the work being conducted in Luxembourg to build a sustainable space ecosystem and progress towards the long-term vision of a thriving space resources industry.

31.1 Future Markets and Value Chains

31.1.1 *Introduction to the Benefits of Space Resources Utilization*

The space sector has been a major enabler for science and technology advances, industrial competitiveness and the development of a knowledge-based society, contributing to a large number of other sectors in the economy in both direct and indirect ways. A number of terrestrial services and non-space industries and

M. Link (✉)

Luxembourg Space Agency, 19-21, Boulevard Royal, 2449 Luxembourg, Luxembourg
e-mail: mathias.link@space-agency.lu

G. Martin

915 Western Drive, Santa Cruz, CA 95060, USA

J. Mousel

Ministry of the Economy, 19-21, Boulevard Royal, 2449 Luxembourg, Luxembourg

markets depend on the existence and operations of space assets. The development of space systems alone generates invaluable technology spin-offs and spill-overs with extensive cascading effects on our daily lives.

Direct market-driven endeavors, once mainly confined to satellite communication systems, are rapidly evolving in growing portions of the upstream space infrastructures sector, embracing space launch capabilities, satellite imagery, and, prospectively, human space exploration and development of space. In this general context, progress made in robotic technologies used in uncrewed exploration missions, the implementation of sample return missions from celestial bodies, the advance of new and game-changing technologies (e.g., 3D printing, resource recycling), and the willingness to advance human ability to explore the universe, have generated increasing interest in the concept of space resources utilization (SRU).

In space agencies' current space exploration roadmaps, indeed, the use of in-situ resources is one of the central priorities. The ability to exploit local resources is broadly acknowledged to be mandatory for long-duration and extra-planetary missions.

The extraction and utilization of volatiles, mainly water, and other resources such as regolith or metals available on celestial bodies, requires the establishment of new supply chains that do not as yet exist. The establishment of these value chains would constitute a first-of-a-kind activity presenting a number of challenges both from the space perspective and from the resource-extraction perspective, but also opening the way to enormous opportunities in areas such as scientific and technological development, knowledge creation and transfer, and commercialization of space activities.

Although the time horizon for the first operational space resource applications is expected to be at least a decade away, preliminary studies and characterization of space resources technology development are currently underway. Further measures will, however, have to be taken to develop the SRU supply chain. As well as the few companies that have already decided to enter the SRU business, it is in the interest of space agencies and public players to help support the emergence of value chains by ensuring that SRU pioneers capture early opportunities and anticipate future needs to act as enablers for the emergence of these value chains.

In this context, the Government of Luxembourg commissioned a study¹ to assess future potential markets and value chains in space resource utilization. The following sections summarize this study, which includes:

- An in-depth evaluation of the potential future markets generated by space resource utilization and analysis of the associated value chains, from space component to utilization, both in space and on Earth,
- Identification of the main challenges and opportunities of establishing the value chains, and

¹ *This study was executed by a consortium of companies from Luxembourg, France, Belgium and the UK. It was funded by the Government of Luxembourg through the Luxembourg National Space programme (LuxIMPULSE) implemented by the European Space Agency.*

- Identification of the key players and enablers responsible for activation of the associated markets and value chains for space resources utilization.

31.1.2 Scoping of the Potential Value Chains

The scope of this study was defined as the combination of both the locations of the resources and the locations of their application or use, between the Earth and the asteroid belt, including natural bodies and artificial satellites (see Fig. 31.1).

Many resources are of potential interest for SRU. The resources considered in the study were selected considering their availability (between Earth and the asteroid belt) and/or their value in various space applications (Table 31.1).

Since the number of value chains emerging from the extensive lists of applications and resources was too large to be analyzed in detail, a confidence assessment was conducted, which aimed at evaluating those with the highest potential to materialize in the coming decades. These value chains were used for the subsequent analysis. This approach did not provide an absolute assessment of the value chains, but rather a comparative one. The approach also ensured that the subsequent opportunity analysis covered the most relevant value chains.

This confidence assessment relied on the evaluation of three criteria for each value chain:

- The volume of demand expected for the application,
- The expected difficulty of accessing the required resources, and
- The feasibility of the mission profile, from an orbital mechanics perspective.

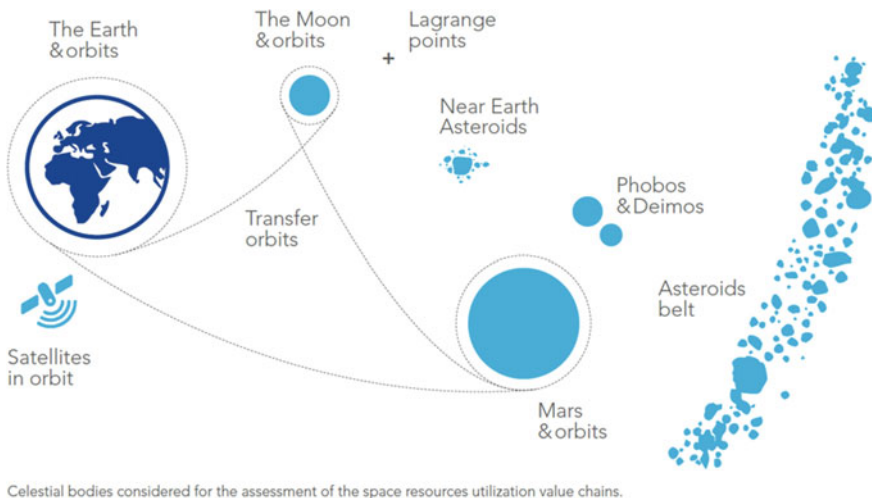


Fig. 31.1 Scope of locations of resources and potential applications (Opportunities for Space Resources Utilization: Future Markets & Value Chains (2018))

Table 31.1 Range of potential applications considered for space resources

Market	Application
<i>In-space applications</i>	
Propulsion	Fuel for rockets
	Fuel for satellites and space stations
Life support for astronauts	Liquid water for drinking and to grow food
	Provision of breathable air (oxygen produced from water via electrolysis)
Infrastructure and equipment	Construction of extraterrestrial ground infrastructures
	Manufacturing of equipment—in situ
	Manufacturing of equipment—for satellites and space station
	Satellites recycling
<i>On-Earth applications (examples)</i>	
Transport vehicles	Auto-catalyst, sensors for auto emissions/climate control
Luxury	Jewellery
Industry	Silicones, hard disk drives, electronic components, glass, crucibles
Energy	Petroleum, fuel cells
Investment	Metal bars
Medical	Anti-cancer drugs, biomedical components, dental

For each value chain, the demand, resources availability and mission profile were ranked and combined into a global score for the whole value chain. The value chains listed in Table 31.2 are those with the highest scores, which constituted the set of value chains framing the subsequent analysis for the study.

Based on the confidence assessment, these ‘high-confidence’ value chains could then be further assessed to identify technology gaps and opportunities.

31.1.3 Cost Savings Assessment for a Defined Value Chain

In parallel to the opportunities assessment, a cost assessment study was performed, with the objective of quantifying, on defined value chains and under given assumptions, the range of cost savings that could be achieved through SRU. The value chains and mission scenarios were selected both for their high confidence score and for the data required to feed the set of assumptions available in the literature (see Fig. 31.2).

In particular, this assessment was conducted for the provision of water from the Moon (for propellant and life support), supplied to spacecraft in low-Earth orbit (LEO), and used for missions going to the Moon and to Mars.

Table 31.2 Value chains with the highest potential according to the confidence assessment. EML: Earth–Moon Lagrange point, LEO: Low-Earth orbit

Utilisation	Application location	Resource location
Rocket refuelling for exploration missions	Earth orbits	Moon
Rocket refuelling for exploration missions	Earth orbits	C-type asteroid
Rocket refuelling for exploration missions	Mars & orbits	Phobos
Rocket refuelling for exploration missions	Mars & orbits	Deimos
Rocket refuelling for exploration missions	Moon orbits, EML	Moon
Rocket refuelling for exploration missions	Moon orbits, EML	C-type asteroid
Human facilities, roads, landing pads, power plants, extraction plants	Mars	Mars
Buildings radiation shielding	Moon	Moon
Buildings radiation shielding	Mars	Mars
Water for drinking/food and for O ₂ generation, N ₂ for breathable air	Moon and Gateway	Moon
Water for drinking/food and for O ₂ generation, N ₂ for breathable air	LEO	Moon
Water for drinking/food and for O ₂ generation, N ₂ for breathable air	LEO	C-type asteroid
Platinum group metals (PGM)	Earth	M-type asteroid
Rocket refuelling for exploration missions	Mars & orbits	Mars
Water for drinking/food and for O ₂ generation, N ₂ for breathable air	Mars	Mars
Water for drinking/food and for O ₂ generation, N ₂ for breathable air	Moon and Gateway	C-type asteroid

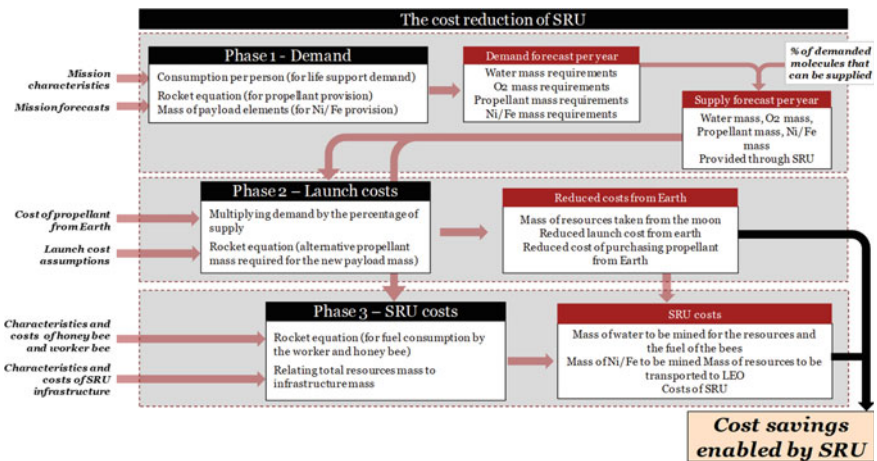


Fig. 31.2 Detailed logic of the cost assessment and inputs/outputs relationships (Opportunities for Space Resources Utilization: Future Markets & Value Chains (2018))

Table 31.3 Assumptions defined in order to compute the cost savings

Assumptions	Approach
Demand for life support	Yearly launch manifest for missions to the Moon and to Mars, with number of people per mission and duration of the mission
Demand for propellant	Each mission is assigned to a vehicles profile among different options
Demand for regolith and metals	Based on the demand for construction materials, conducted for the confidence assessment
Launch costs from Earth	Cost per kg for each vehicle per destination (LEO, Moon, Mars), and varying distribution between the vehicles depending on the scenario
Supply share from SRU	Yearly percentage of water / regolith and metals / platinum group metals procured through SRU
Cost of mining water	Cost of SRU extraction equipment and their expected lifetimes
Cost of mining regolith and metals	Due to the technical difficulty of separating and processing metals from the soil, it is assumed to be more expensive than water mining. A factor of 3 seems realistic and conservative
Cost of SRU permanent infrastructure	Cost of lunar outpost associated with its lifetime

The mission architecture to supply the space resources relies on a lunar outpost (permanent infrastructure), which serves as a water depot, supplied by water production from the lunar surface before supplying LEO. SRU spacecraft perform the mining operations and delivery to the outpost and the resources transportation from the outpost to LEO.

The cost savings assessment was then computed, using assumptions on the aspects listed in Table 31.3.

The cost savings were computed as the difference between what it would have cost to launch the material from Earth versus the cost to mine and deliver the material from space to the locations. Due to the uncertainty on a number of parameters, one conservative and one optimistic scenario were analyzed.

The conservative scenario was formulated with a precise analysis of planned missions from space agencies and private companies. The demand in the cis-lunar environment was driven by the number of satellites around Earth and the potential for servicing and refueling. On the Moon and Mars, the demand and population forecasts were characterized by the realistic ambitions of establishing permanent bases and settlements.

In the optimistic scenario, SRU was not enabled earlier, except for water for life support. It represents the case where efforts to reach other bodies are more effective and therefore happen earlier and in a greater number. This implies that demand grows faster alongside the number of humans in space.

Under these assumptions, and for the scope of demand defined above (launch manifest to the Moon and Mars), the total cost savings for the period up to 2045 were estimated at between €90 billion (conservative scenario) and €256 billion (optimistic scenario). All cost numbers are expressed in 2018 present value.

The provision of propellant was expected to be the first source of savings, reflecting the dominant source of demand for both conservative and optimistic scenarios. The manufacturing value chain was expected to come close to that of the propellant.

After a slow start, the cost savings increase in the different value chains, but substantially for the propellant value chain between 2030 and 2035, when the share of demand supplied by SRU starts to increase. The savings from the propellant value chain alone increase to over €7–8 billion per year after 2040 in the conservative scenario and to over €17 billion per year in the optimistic scenario.

31.1.4 Opportunities Linked to the Space Resources Value Chains

Along with the technical aspects described above, a stakeholder consultation was conducted, covering all types of stakeholders that are linked directly or indirectly to SRU value chains (institutional, commercial, scientific). As well as confirming and refining the technical analysis, this consultation led to a characterization of the SRU value chain in terms of risks, drivers and enablers.

This analysis of challenges, enablers and opportunities was performed for the whole SRU value chain, which was defined in a similar way to that for the Earth mining value chains. Among the several possibilities, the segmentation shown in Fig. 31.3 was chosen.

A high-level roadmap for the implementation of SRU was developed by adding stakeholders' views on the priorities for SRU and expected timeline to the analysis performed (see Fig. 31.4).

The discussion of each stage of the SRU value chain led to the identification of 13 major opportunities (Table 31.4), which are either initiatives that would lead to a commercial business case or an initiative that would provide a strategic role for the entity achieving it.

These opportunities represent potential areas of investment for commercial entities looking for viable business cases, or for public players looking for strategic contributions to support the emergence of SRU.



Fig. 31.3 Segmentation of the SRU value chain into stages (Opportunities for Space Resources Utilization: Future Markets & Value Chains (2018))

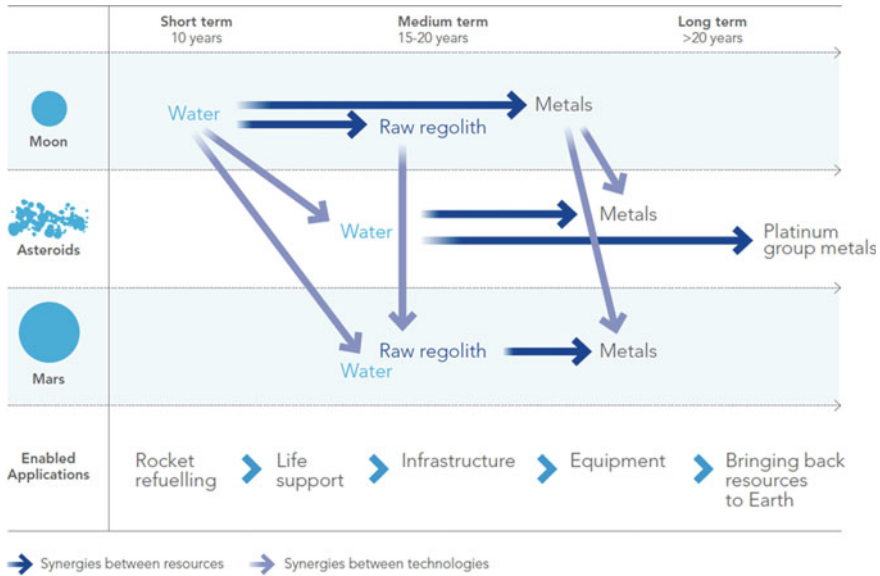


Fig. 31.4 Suggested roadmap for SRU implementation based on stakeholder consultation (Opportunities for Space Resources Utilization: Future Markets & Value Chains (2018))

Table 31.4 List of commercial opportunities and initiatives

#	Opportunity title
1	Cataloguing and standardisation of results for prospected sites
2	Reference methodology for the demonstration of proven reserves
3	Technology and process development for mining and refinement process
4	Creating opportunities between Earth mining companies and space companies
5	Development of manufacturing processes using regolith
6	Development of processes for in-space manufacturing
7	Development of processes to produce propellant and oxygen on Mars
8	Optimization of the SRU spacecraft architectures and propulsion
9	Development of long-term storage fuel depots
10	Development of technology for proximity operations
11	Attracting companies under regulatory framework for in-orbit operations insurance
12	Providing a framework ensuring interoperability of SRU processes and technologies
13	Commercial provision of deep-space communications and energy

31.1.5 *Socio-economic Impact*

The activation of SRU value chains is expected to create socio-economic benefits that are grouped into three categories:

- *Industrial effects*

From an economics standpoint, the core economic activity relates to the space resource utilization industry that will be created in order to fulfil the envisioned value chains. The operations of this terrestrial industry (and supporting supply chains) will provide employment (jobs supported) and generate economic value (measured in economics as gross value added (GVA)).

- *Spill-over effects*

A ‘spill-over’ refers to any economic value (may be positive or negative) that arises outside the core economic activity—both within and outside the space sector. Such additional value may arise to: end users of the commercialized SRU product/service (e.g., provision of fuel in space)—market spillovers; adopters of the technology/knowledge created—knowledge/technology spillovers; and/or the network, standards, reputation effects—network spillovers.

- *Wider impacts*

A range of broader catalytic impacts are expected in terms of positive external influences on the environment (e.g., reduction in launch fuel consumption due to the provision of in-situ resources), on strategic positioning (e.g., strengthening of space competitiveness), and on society and citizens overall (e.g., improved safety due to increased capability of cataloguing near-Earth objects).

A quantitative assessment was undertaken on the following five value chains:

- Fuel supply (Moon, Mars missions)—water (H₂O) and methane (CH₄);
- Life support (LEO, cis-lunar, Moon, Mars)—water (H₂O);
- Commercial sale (Earth return)—platinum group metals (PGM);
- Infrastructure construction (Moon, Mars)—regolith and local resources; and
- Equipment manufacturing (LEO, cis-lunar, Moon, Mars)—metals.

The two last value chains, infrastructure construction and equipment manufacturing, were merged during the analysis. It is difficult to estimate the concentration and needs of these materials (regolith and metals) distinctively at different places in the solar system. Nonetheless, it is reasonable to consider that regolith accounts for 75% of the total demand for regolith and metals. The developed model extends up to 2045 for industrial effects and market spillovers, and to 2070 for technology/knowledge spillovers. All benefit values were discounted to 2018 present value terms, such that they may be compared to the upfront investment costs.













In the economic analysis, the impact of an initiative depends on the degree of additionality generated. Additionality is defined as the difference between the realized outcome in the scenarios with the initiative and the hypothetical outcome in the counterfactual scenario without the initiative. The additionality is, therefore, measured

as the difference in socio-economic effects resulting from the fulfilment of all foreseen space missions until 2045 using space-based resources rather than Earth-based resources.

Based on the updated assumptions, the same socio-economic indicators were computed, using the same approach and methodology. Table 31.5 summarizes the results for both scenarios and compares the relative change between the conservative and optimistic scenarios.

In summary, the study of the benefits of SRU identified the strong potential for development of a future space industrial sector based on SRU and that now is the time to begin the research and development work needed to realize SRU space markets. To jump-start the establishment of the new industry, the Luxembourg Space Agency facilitated a dialogue between experienced mining industry experts and space resource utilization entrepreneurs. The results of these discussions are summarized in the next section.

Table 31.5 Comparison of socio-economic benefits of SRU between the scenarios (present value, 2018). *Note* Each blue square represents a 20% positive effect

Socio-economic benefit	Conservative scenario	Optimistic scenario	% change (optimistic–conservative)
<i>Industry</i>			
Market revenue	€73 billion	€170 billion	+133%
Gross value added	€49 billion	€108 billion	+120%
Employment	845,000 FTE-years	1.8 million FTE-years	+113%
<i>Spill-overs</i>			
Market	€54 billion	€135 billion	+150%
Technology	€2.5 billion	€2.5 billion	+0%
<i>Network</i>			
Coordination			Same
Standards			Same
Critical mass			Same
<i>Wider effects</i>			
Environmental			Same
Societal			Same
Strategic			Same

31.2 Commonalities Between the Terrestrial and Space Resources Industries: What Can Each Learn from the Other?

31.2.1 Commonalities Between the Terrestrial and Space Mining Value Chains

The SRU value chain shown in Fig. 31.3 is based on the terrestrial mining value chain. Although there are certainly differences, many commonalities between industries can be found in each stage of the value chain.

Prospecting

Before resources can be mined, they need to be located. This is where prospecting comes into play. In terrestrial mining, remote prospecting techniques are commonly complemented by on-ground campaigns to gather more information. Yet, in space, making ground truth assessments to validate remote sensing data is currently very limited. Until now, only a few missions to celestial bodies have actually returned information on the exact composition of the material, and only for very confined regions.

Establishment

In terrestrial mining, this stage includes administrative steps, such as obtaining the mining rights and gathering the required funding, as well as building the infrastructure and access to the mine. While some administrative steps might also be considered for space mining, such as obtaining authorization for the space mission from a government entity, a system for extraterrestrial mining rights does not yet exist. For space, this phase can be regarded as the mission planning phase, e.g., finding the most efficient mission timeframe, planning the infrastructure needed, launching and establishing the mine on a celestial body.

Mining

This stage concerns the actual mining of the reserves. On Earth, mining is generally done by people using large machinery, although new semi-autonomous technologies (e.g., robot swarms) are currently being developed. Automation will play a larger role in the future but is currently limited to a few applications, such as in trucks and trains. In addition, new techniques will need to be developed to account for the challenges of the space environment, such as vacuum, radiation, temperature fluctuations and low gravity.

In space, the size of the machinery and human workforce will be very limited or non-existent, therefore robotics and automation will play a crucial role. Here, the existing vast experience of mining technology and processes in the terrestrial mining industry will bring an important contribution to the space resources industry.

Transport

It is evident that the means of transport will be completely different in both industries, e.g., spaceship or rover vs. big trucks or trains. Nonetheless, a common point of interest can be the storage solutions used, especially for gases.

Refining

Due to the limitations in space including the vacuum, radiation and low-gravity environment, the large-scale refining that occurs on Earth cannot be reproduced using the same approaches. Nonetheless, the terrestrial industry's experience in refining processes can help guide the development of the new refining processes needed for space resources, which may in return bring improvements to terrestrial operations.

Manufacture

There exists a large number of manufacturing techniques on Earth, some of which can also be adapted to space, e.g., 3D printing.

Supply

This is probably the phase where the two industries differ the most. While supply chains on Earth are well established, the actual commercial market in space has not yet developed and no real customers exist. The first customers will probably be space agencies.

31.2.2 Engaging Space and Mining Industries

The Mining Space Summit was a one-day workshop organized twice, in 2018 and 2019, by the Luxembourg Space Agency (LSA), focusing on the two main challenges that will be key to enabling the success of the space resources utilization sector: (1) the viability of SRU business models and (2) the development of critical technologies and operations (Major Takeaways of the Mining Space Summit 2018, 2019).

To address these challenges, engagement between the space resources community and terrestrial industries, including mining, and oil and gas, is essential. To facilitate a productive dialogue on these points, stakeholders from across terrestrial and space resources industries, financial, and government communities took part in the two summits.

Over the course of the events, participants identified and discussed the opportunities and challenges faced by the space resources industry. The participants actively engaged in six parallel breakout sessions, three on business-oriented topics and three with a technical focus:

- **Market and Dynamics:** understanding space resources supply and demand dynamics by considering their use cases, prices, associated costs, and other factors.

- **Investment and Financial Planning:** financing space resources projects and ventures, and understanding potential financing models for projects in an innovative and high-risk frontier field.
- **Role of Government and Regulators:** enabling the growth of a nascent industry through public policy and regulatory actions.
- **Prospecting—Proving Value:** finding, identifying and analyzing resources to prove their value and justify mining operations.
- **Extraction—Creating Value:** establishing and operating mines in extreme and remote conditions and generating value from a mine in space in a sustainable manner.
- **Enablers—Optimizing Value:** increasing mining efficiency by leveraging critical support services, technologies and processes, such as logistics, communication services and power distribution.

In the following sections, reflecting the results of these discussions, the opinions reported are those of the participants.

31.2.3 *Markets and Dynamics*

According to the participants, the space resources market is currently in a development stage, with governments and space agencies as the main potential customers. The market is expected to evolve over time, with the near term being dominated by products derived from water, used as fuel or for life support. Due to its early stage, a lack of coordination between the different elements of the value chain and a communication gap between the space community, the customers, and investors can be observed. For the medium term, the market could evolve into in-space manufacturing along with in-space construction of facilities.

To transform the space resources industry into a strong, sustainable economic sector, more companies and a larger customer base are needed. Interaction between space resources companies and potential customers, as well as with terrestrial industry, needs to be increased to develop stronger business cases. In general, the space resources community needs to evolve from ‘technology push’ to (terrestrial) ‘market pull’.

With current risk levels too high, commitments from governments as anchor customers are needed to jump-start the space resources sector, and a profitable market is only expected to develop in a future too distant for most investors.

Consequently, it is difficult to convince potential investors to invest in the SRU sector. Although terrestrial mining projects share similar long time horizons, this established industry is well understood but is also known as a ‘slow-moving dinosaur’, often avoiding risky decisions.

The workshops generated six principles to guide the space resources community:

- Develop stakeholder communication guidelines for customers, investors, and regulators.

- Develop and integrate end-to-end services for space resources.
- Focus on risk reduction, starting at an early stage by better integration, and instituting standardization.
- Define the terms needed for the space resources industry and seek agreement with stakeholders.
- Consolidate international frameworks and regulations.
- Create a positive vision for the industry, including robust assumptions with a customer pull as well as an investor push.

These principles resulted in four recommendations:

- Develop a stable customer base by creating a need for products made in space, e.g., private and/or public outposts on the Moon. Involve potential customers in market research and product development. This will in turn create an important alliance of customers needed for long-term sustainability.
- Develop a stable commitment from governments by encouraging them to be anchor customers, guaranteeing to buy resources at a fixed price at a given time.
- Develop a stable legal and regulatory framework to bring down risk for investors and promote the entry of more companies into the space resources sector.
- Develop a critical mass of investments needed to sustain the sector. Big companies may initially play an important role in backing small space players.

31.2.4 Investment and Financial Planning

The space industry is undergoing a profound evolution, as affordable access to space coupled with small spacecraft architectures has invited new breeds of investors into the sector. Despite the increase in financing, investment in SRU has mostly been limited to government grants.

Several factors constrain potential investors: the long-term time horizons of space resources ventures, the lack of commercial customers, uncertainties regarding customer needs, profitable business models and market size, as well as the absence of an international legal framework.

The lack of current demand means that space resources companies should try to focus on 'dual-use' projects, using their technology for Earth applications in the near term to generate revenue and then to transition into space at a later stage.

At present, there are no concrete price tags for space resources, nor is their exact availability known. Participants mentioned concepts like a space resources commodity exchange to counter uncertainties in the SRU marketplace. In order to develop demand, focus on a specific resource for prospecting, like water or oxygen, where the potential use cases and thus business models are better known. This approach can be a first step to attract financing and accelerate the market development.

Overall, the risky regulatory environment of space resources requires new financial mechanisms that need to be investigated and developed. For the time being, governments and space agencies are needed to fund the necessary technology advances required to establish this new industrial sector in space.

31.2.5 Role of Government and Regulators

Currently, there is neither a production capacity nor a market for space resources. The main barriers to sustainable and commercially viable space resources activities are:

- Resources are surrounded by uncertainty, as exploration and estimation of reserves are still at an early development stage.
- Space mining technologies still have to be tested and demonstrated.
- The initial users have to be identified and a market has to be developed.
- Future space resources activities have to set a high sustainability of operations standard from a transportation, logistics, maintenance and infrastructure standpoint.
- There is uncertainty related to regulatory aspects, such as the required legal framework, property rights, standards and taxation.

Governments are able to, and should, play a key role in reducing these barriers, while not overwhelming the efforts of the private sector.

Should the government be a facilitator or a regulator? A balanced approach that would allow governments to ensure proper administrative function and manage the related risks will be necessary. The terrestrial mining regulatory system could easily be transferred to the space sector. Such an approach could rely on either a national authority or an international body similar to the International Telecommunication Union (ITU), both having advantages and disadvantages. Governments from all development levels need to be involved.

At the same time, discussions in international fora such as the United Nations Committee for the Peaceful Uses of Outer Space (COPUOS) need to continue. On a national level, governments should evaluate existing administrative requirements and consider how current barriers can be reduced.

There is a need for governmental funds to be invested in a sustainable and responsible manner. Governments should continue to invest in the space industry. These investments should also significantly contribute to solving pressing problems on Earth.

Finally, industry associations, non-political in nature, are needed to support governments in creating future policies and frameworks.

31.2.6 Prospecting—Proving Value

On Earth, prospecting is well established. In space, the situation is different: the prospecting results that do exist originate from a limited number of scientific missions and thus, while the information may be sufficient for some elements (e.g., iron, titanium), the data for the probable first product, water, is, according to the participating experts, not well known and is related to the unexplored cold traps on the Moon. The spatial resolution of current prospecting missions in space is low; it only maps the surface material, but does not provide information on depth of the deposits.

Additionally, current space prospecting does not address extractability, although the efficiency of extraction is critical. Current technology readiness of the necessary extraction technologies in space is still only in the experimental and bread-boarding stage, thus maturation is necessary.

Current prospecting missions are publicly funded; the work of the space agencies right now is comparable to pre-competitive geological survey work.

For prospecting to be done correctly, it needs to be treated as a campaign, not as a single mission. Getting the best results will require ground truth, which means accessibility for robotics that can resolve knowledge gaps such as: how deep do you need to drill? What spacing do you need between drill holes? Depth and continuity information will also be necessary.

These data will support development of stronger business plans with enough credibility to attract investors. But the first question will be: who will pay for the prospecting campaigns? On Earth, you can obtain a lease of the land after paying for the prospecting, but this is currently not possible in space, further exacerbating the need for a clear regulatory framework.

In the end, for investors to be able to make an informed investment decision, different prospecting campaigns need to be comparable. On Earth, this is done through reporting standards like JORC² or CRIRSCO³ in mining and PRMS⁴ in oil and gas. A similar standard should also be developed for space operations. Additionally, a clear definition of space resources terminology is essential, as, although the same terms are used, the current meaning in space is often different than the words used for terrestrial mining (e.g., resource, reserve).

31.2.7 Extraction—Creating Value

Participants stressed that terrestrial mining is a very old industry that must significantly evolve and adapt over the coming decades as resources become more and

² Joint Ore Reserves Committee.

³ Committee for Mineral Reserves International Reporting Standards.

⁴ Petroleum Resources Management System.

more difficult to access. Consequently, requirements on Earth and in space will become all the more similar, resulting in dual-use opportunities. The current trend for increasing automation in terrestrial mining is a prerequisite for successful space resources missions.

On the other hand, the space resources industry is in its infancy, such that while many mining concepts have been developed, no architecture has been proven in actual operation. Swarm architectures, where smaller robots collaborate in one operation, could be one of the key technologies for both industries to operate in extreme conditions and to leverage mass automation advantages as well as keeping people out of harm's way.

In near future, terrestrial and space resources industries can support each other by developing baseline technologies on Earth that can later be reutilized in space. Potential technologies therefore are robust sensors in combination with machine learning and artificial intelligence, waterless techniques, fuel cells, carbon-neutral steel production and, as already mentioned, robots and autonomous systems in general. Another possible contribution of the terrestrial mining industry may be their general expertise in handling and processing resources.

Targeting smaller mining companies and especially advanced equipment providers seems the most promising option for space companies.

Ground-based testing is crucial for success. But it is impossible to recreate the challenging environments of space on Earth. Therefore, processes and technologies need to be validated on the Moon, and Earth-based testing limited only to what can be done realistically to minimize space risks.

31.2.8 Enablers—Optimizing Value

Terrestrial mining companies are adopting a modular design architecture that can be modified more easily and tailored to fit the changing realities they face as mines mature. One current trend is to have small mining compared to mass mining approaches. This flexible architecture would also be beneficial for space resources mining, where there is a need to be as low in mass as possible due to launch cost considerations.

Supporting technologies that could create a market for technology transfer from terrestrial to space mining include:

- Nanobots/small robots, micro-tunneling and swarm robotics
- Remote sensing technologies
- Communications—dealing with latency and availability
- Reliability in communications, estimation, and automation.

The main challenges facing the space resources industry, where collaboration with the terrestrial industries could benefit, are:

- Energy: energy requirements drive innovative solutions, such as power beaming, which are in development.
- Exploration technologies: new techniques developed and used on Earth may be adapted and used in space.
- Market: there are currently no markets for mining industries in space. What needs to be done to create the markets in the next 20 years?
- Market stability: if launch costs are reduced, will there still be an LEO market for SRU water for propulsion? Like any market, this may evolve over time, especially once there are human settlements in space.

To bring these topics forward, a collaboration needs to be established through the identification of common topics and dual-use technologies for space and terrestrial mining. A global competition, similar to the XPRIZE, could be imagined as a catalyst for such ventures.

The dialogue between terrestrial and space mining industries helped to characterize the primary challenges facing the future SRU industrial sector. Many of the issues require the assistance of space agencies to mitigate the risks and provide financial support for these emerging commercial space companies.

31.2.9 Examples of Companies Generating Near-Term Value with Long-Term Goals in Luxembourg

Currently, there are several companies working on the future of space resource utilization while at the same time generating near-term value to achieve this goal. This section contains some examples of such companies, based on information they provided.

31.2.9.1 Blue Horizon

New approaches to the protection of the Earth's environment are required to fight global warming and the destruction of Earth's resources. At the same time, the exploration of other planets is becoming more and more important to reduce the impact on Earth and gain additional resources.

Blue Horizon is focusing on the identification of life (if any) on other planets, as this knowledge is a prerequisite for any further exploration and exploitation. Based on the technologies for Earth environmental protection and exploitation, the building of (possibly closed) ecosystems on other planets for the purpose of growing food for human colonies, the exploration of natural resources to accumulate building material, oxygen and other required products is at the center of Blue Horizon's long-term strategy.

The protection of the environment and the resistance to climate change are no longer considered solely a scientific activity or work exclusively conducted by NGOs,

but is moving into the center of the economy. Massive investments by public and private industry are needed to protect the life of mankind on Earth by reducing the impact on the environment. Blue Horizon aims to provide products and services that have a positive impact on the environment and its changes.

Terrestrial and space research and innovation are synergetic in the sense that they profit from each other or are even co-dependent. For example, the development of technologies that allow the cultivation of food on other planets (or in near term on the Moon) enables the identification of new approaches to fight desertification. Blue Horizon will establish as a node to use and apply the best of both worlds for a better protection of Earth's environment and a sustainable use of space resources on other planets.

31.2.9.2 ispace Europe

ispace is a commercial lunar services company specializing in the design of small lunar landers and rovers for payload delivery to the Moon (see Fig. 31.5). In addition to the company's payload service, in August 2020 ispace announced the launch of a data-centric platform concept which aims to support industry players with lunar market entry. The platform aims to leverage efficient, user-friendly mission planning tools and lunar data applications being developed in house at ispace. ispace's aim is to enable the establishment of a cis-lunar economy through space resources utilization (SRU), including legal and policy aspects, fundamental scientific research, and technology development across the SRU value chain.

ispace is approaching SRU in three phases. First, ispace plans to demonstrate its transportation, landing and exploration technologies and capabilities from 2022 to



Fig. 31.5 ispace's lunar lander (used with permission of ispace)

2023. Second, ispace is aiming to establish a high-frequency cadence of payload delivery to the Moon, as well as the capability to collect and process an array of lunar data from 2024 onward. And, finally, ispace aspires to act as an orchestrator of a cis-lunar economic ecosystem, of which SRU is widely anticipated to be a critical component.

ispace has plans to explore and extract essential lunar resources starting in 2030. This includes the production of fuel (H, O) in space, the support of lunar infrastructure development, and working towards the realization of the cis-lunar economy. As such, ispace will be focused on the exploration and extraction of elements including hydrogen (H), oxygen (O), iron (Fe), titanium (Ti), and rare-earth elements (REE), among others, as well as the use of lunar regolith for construction purposes.

ispace is developing plans to ensure that revenue streams are realized while returning value to its shareholders and investors. In the short term, ispace has already started to generate revenue through its payload delivery service, as well as through corporate partnerships of its HAKUTO-R lunar robotics program. Additionally, ispace is planning to build value in the near term through its lunar data business.

ispace is convinced that the technologies, knowledge, and experience of the space sector will be of great use to terrestrial companies. Similarly, the knowledge and experience of terrestrial companies will be of great benefit to the development of the lunar economy, and ispace is convinced that collaboration with terrestrial companies will create opportunities to support these companies with the expansion of their business from the Earth to the Moon. As such, ispace is actively engaging with terrestrial industries to co-develop technologies and business models that could bring value in both the short and long term to all parties involved.

31.2.9.3 Maana Electric

Maana Electric is a space and renewable energy startup that emphasizes the dual applications of space technologies for terrestrial and space purposes. The company is focused on the utilization of ISRU technologies for the localized production of solar panels using desert sand (on Earth) or regolith (in space). Maana is headquartered in Luxembourg and concentrates on the development of its ISRU and panel producing technologies (see Fig. 31.6).

Maana uses its proprietary ISRU technologies developed for the space industry to revolutionize the way in which solar panels are produced on Earth and in space.

- Solar cells (for terrestrial and space applications)
- Solar panels (for terrestrial and space applications)
- Glass panes and components
- ISRU equipment.

The products developed by the company could be used in remote regions of the world and on the Moon for the purpose of generating solar energy.

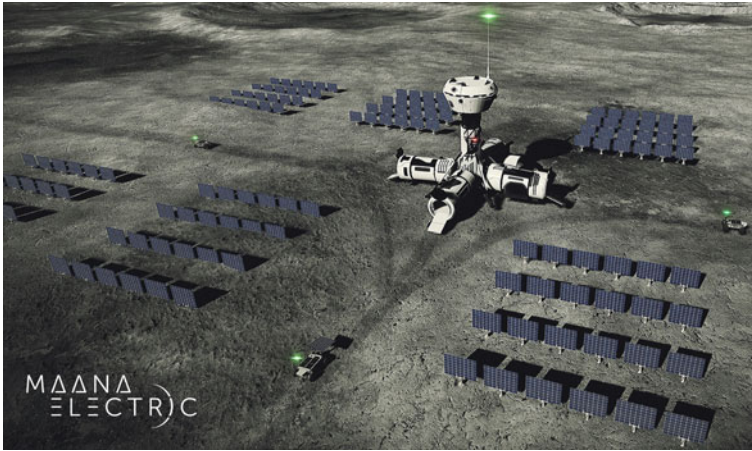


Fig. 31.6 Illustration of Maana’s Luna Box (used with permission of Maana Electric)

Maana’s terrestrial product, TerraBox, will be capable of producing 10 MW of solar panels a year utilizing desert sand as input material. This innovative way of producing solar panels:

1. Allows for the local production of panels on the installation site;
2. Enables a reduction of greenhouse gas emissions in the solar panel production value chain; and
3. Allows for a cost-effective and competitive alternative to the industry standard solar panels of today.

As a company integrated in both the terrestrial and space industry, Maana considers this cooperation essential for its business to succeed. Due to the high-tech nature of the technology, space industry contracts allow Maana to push the boundaries of ISRU technologies, while the terrestrial industry offers the company an established market to enter once the technologies are mature.

For the space industry to fully commercialize around space resources, it will require the capital available in terrestrial markets such as the energy industry. Once the prospects of space resources have become undeniable and the terrestrial industry starts considering investments in space, the space resource industry will be able to fully flourish.

31.2.9.4 OffWorld

OffWorld is working on the deployment of a machine-learning industrial robotic workforce to undertake mining, manufacturing and construction on planetary surfaces and in space. The deployment of their modular platform with machine-intelligent modules operating as a collaborative swarm of hundreds or thousands of

robots enables low-cost and robust robot deployments to operate anywhere in the solar system with expanding species capability (see Fig. 31.7).

OffWorld's robotic platform on Earth can perform extraction, separation and processing functions for mining operations. For construction they include demolition, assembly, additive construction, welding and fastening. By including transport, servicing and additional tasks, the OffWorld robotic platform builds into an integrated mining, construction and manufacturing capability serving terrestrial markets. As their machine intelligence and systems mature, these can be deployed into space for end-to-end industrialization for mining, processing, manufacturing and construction of habitats, vehicles, space stations, arrays and other components required to build civilization from cis-lunar space deep into the solar system.

OffWorld deploys its robots to mining and construction customers for Earth-based markets. They are developing a new industrial niche of rugged, swarming, machine-intelligent robots that do not require human presence in situ. For the mining and construction sectors, this may open up new paradigms of operations at an order of magnitude in cost reduction. The development of this architecture underwrites their space expansion in the future (see Fig. 31.8).

The principles of business practice that follow economic drivers are essential to expanding humanity's presence in the solar system. The implications are that business needs to be immediately profitable where possible. As challenging as this is, it is OffWorld's experience that the stronger the value proposition, the earlier revenues and profits are generated out of direct market need. Since customers are currently Earth based, the value propositions must directly benefit those customers within timescales that make strong business or investor requirements, typically within 1–3 years.

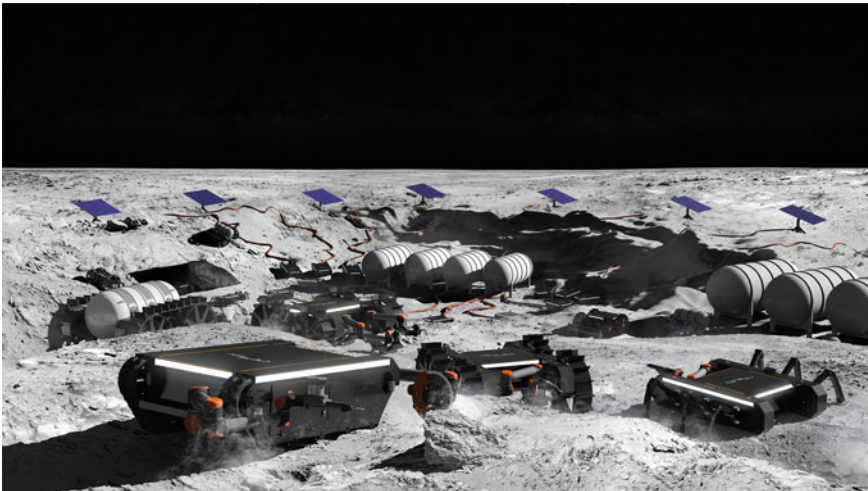


Fig. 31.7 Lunar operations (used with permission of OffWorld)



Fig. 31.8 OffWorld's Diggerbot in the field (used with permission of OffWorld)

31.3 The Role of Governments as Key Enablers for the Space Resources Industry

31.3.1 Governments and Agencies as Early Risk Takers and Anchor Customers

As mentioned earlier, there is currently no commercial space resources market. Often, potential providers of space resources cite a lack of customers, while potential customers mention the lack of providers.

To solve this problem, governments can play a decisive role. Apart from the satellite and commercial launcher industries, governments and their space agencies are the only current space 'users' that can benefit from space resources. As this industry requires a larger initial investment without guarantees, commercial entities will shy away from such a venture, while the risk may be acceptable for a space agency.

By providing the initial investment, significant cost reductions at a later stage are possible, and the role of the agencies as anchor customers allows the development of a space resources market.

Even without accessing space, supporting the development of space resources technologies can bring an additional benefit to a government, e.g., through terrestrial spin-offs. An example of such support is Luxembourg's SpaceResources.lu initiative.

31.3.2 Luxembourg's SpaceResources.lu Initiative: An Example of a Government-Supported Strategy Promoting Commercial Space Resources Utilization

The SpaceResources.lu initiative launched by the Government of the Grand Duchy of Luxembourg in February 2016 provides an example of what governments and agencies can do to promote forward-looking initiatives and to encourage the development of new markets, by establishing attractive frameworks and acting as early risk takers and anchor customers.

The SpaceResources.lu initiative aims to promote the peaceful exploration and sustainable utilization of space resources for the benefit of humankind. The initiative is putting in place a comprehensive ecosystem for the development of activities related to the use of space resources. Its different activities aim at gradually addressing the different technical, legal, economic and business aspects related to the use of space resources.

More specifically, the SpaceResources.lu strategy builds on the following five pillars:

1. To ensure national political support and promote international cooperation.
2. To build an attractive legal and regulatory framework.
3. To promote long-term development by supporting public research and education.
4. To offer dedicated support for private-sector research and development activities.
5. To develop long-term financial instruments tailored to the needs of private companies.

In just a few short years, by implementing activities along these five pillars, the Grand Duchy of Luxembourg became a key player on a global level, recognized for its ambition in the area of space resources utilization.

31.3.2.1 Ensuring National Political Support and Promoting International Cooperation

Luxembourg has been active on various levels in order to ensure national political support and promote international cooperation. On an international level, it has promoted space resources utilization in its relations with the European Space Agency (ESA), the European Union and the United Nations. Since 2016, numerous bilateral agreements have been signed to this end with countries such as the United States, China, Japan and the United Arab Emirates, as well as several European countries. These agreements generally express an intention to collaborate, which includes the exchange of information and expertise, scientific and technological cooperation, and exchanges on legal matters related to the use of space resources. As part of the bilateral agreements, the Luxembourg Space Agency (LSA) has also organized workshops and meetings with representatives of partner countries, in order to share their respective policies and present the development of their commercial space sectors.

LSA also invited representatives of partner countries to specialized events held in Luxembourg, such as the NewSpace Europe conference, the Mining Space Summit, the Space Resources Week and the Space Forum.

Over the past few years, LSA has been actively involved in various events and conferences abroad, such as the International Space Exploration Forum (ISEF-2) in Tokyo, Japan, the International Astronautical Congress (IAC) in Adelaide, Australia in 2017, Bremen, Germany in 2018 and Washington, USA in 2019, and specialized conferences such as the Planetary and Terrestrial Mining Sciences Symposium and Space Resources Roundtable in Golden, US and in Canada, or the International Future Mining Conference and Off Earth Mining Forum in Sydney, Australia. These events provided an opportunity to strengthen international ties and forge new relationships, while further promoting the use of space resources.

As a further key measure to promote and strengthen space resources utilization, the Luxembourg government established a high-level advisory board on space resources, which advises its ministers on matters related to space resources. Its members have provided strategic advice and further promoted the topic on a global level.

31.3.2.2 Building an Attractive Legal and Regulatory Framework

To gradually establish a clear and attractive legal and regulatory framework, Luxembourg chose to proceed on a step-by-step basis. Indeed, the situation is complicated and requires governance that is able to evolve by taking into account new technological and economic achievements.

Luxembourg is a signatory to the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (the Outer Space Treaty) and the Convention on the International Liability for Damage Caused by Space Objects. In accordance with these treaties, Luxembourg chose to start approaching the issue on a national level. In 2017, the parliament adopted a law on the use of space resources, which introduced an authorization and supervision system in Luxembourg. The law is aimed at regulating the use of space resources and is the first legal framework of its kind in Europe.

In parallel, Luxembourg has engaged on a multilateral and international level. At the United Nations, the question of space resources is handled by the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) in Vienna, Austria, as a specific point on the agenda of its legal subcommittee. Work is currently underway, and an increasing number of countries are expressing interest in establishing an international framework in the future. At the meeting of the legal subcommittee of COPUOS held in April 2019, Luxembourg presented the SpaceResources.lu initiative and the vast opportunities related to the exploration and use of space resources.

Representatives from Luxembourg-based entities were actively engaged in the work of the Hague International Space Resources Governance Working Group between its start in 2016 and the end of its work in 2019. The Group's aim was

to develop policy building blocks which could provide a basis for a future international framework for the peaceful exploration and use of space resources. The Government of Luxembourg provided financial support to the second phase of the Working Group in collaboration with the University of Luxembourg. The Working Group finished its study with the publication of a list of 20 building blocks.

The most recent action was to negotiate and sign the Artemis Accords, together with the United States, Australia, Canada, Italy, Japan, the United Arab Emirates and the United Kingdom. The Accords describe a shared vision for principles, grounded in the Outer Space Treaty of 1967, to create a safe and transparent environment which facilitates exploration, science, and commercial activities for all of humanity to enjoy. They will enable coordination and management of NASA's Artemis program operations and will also serve as useful input to development of a corresponding framework at United Nations level.

31.3.2.3 Promoting Long-Term Development by Supporting Public Research and Education

Within the SpaceResources.lu initiative, another important area of activities was the support of public research and education. The contribution of research is key to the development of the required technologies. Education is needed to train the workforce to staff this new field.

Various projects and initiatives were started with the main public research institutions in Luxembourg: the Luxembourg Institute of Science and Technology (LIST) and the University of Luxembourg. These projects were funded by LSA, ESA and the Luxembourg National Research Fund.

Workshops and conferences were also organized to support the scientific community and connect scientists with companies active in the SRU field. A good example is the Asteroid Science Intersections with In-Space Mine Engineering (ASIME) conference that was organized in 2016 and 2018, focusing on the interaction of asteroid sciences with space resources utilization.

On education, Luxembourg became active at both secondary and university levels, with various initiatives promoting space and space resources utilization. In September 2018, the LSA and University of Luxembourg inaugurated a specialized Master's program, the Interdisciplinary Space Master (ISM). The aim of the ISM is to train talented individuals in Luxembourg in the space skills needed by the Luxembourg space community and thereby support the development and sustainability of the commercial space-sector ecosystem. It includes courses on space resources utilization. In the framework of the Space Resources Weeks of 2019 and 2021, LSA organized an SRU professional development course together with the Colorado School of Mines and the International Space University.

31.3.2.4 Offering Dedicated Support for Private-Sector Research and Development Activities

Luxembourg has become very active in supporting private and public collaboration in space technology developments and innovation. Local space companies are able to conduct joint research with Luxembourg research laboratories to develop new processes, hardware, and test new approaches important to the space resources industrial sector.

Several research collaboration agreements have been signed with organizations such as ESA, NASA, DLR, and in addition, Luxembourg joined the International Space Exploration Coordination Group (ISECG). These agreements provide further opportunities to set up projects and programs related to the use of space resources. Numerous projects were started with private companies, supporting R&D activities related to innovative business plans and enabling further development of related economic activities in Luxembourg.

31.3.2.5 Developing Long-Term Financial Instruments Tailored to the Needs of Private Companies

On the funding and financing level, Luxembourg used different existing instruments to support R&D developments both at the national and European level. In addition, Luxembourg's national investment bank was able to provide debt financing and equity investments to new commercial space companies. A key milestone was also the signature of an agreement with the European Investment Bank, one of the world's largest multilateral investment organizations. Another important activity was the establishment of a new venture capital fund in which the Government of Luxembourg invested along with private investors.

31.3.2.6 Establishing the European Space Resources Innovation Centre

To support much needed space resources research, LSA launched a major initiative together with ESA and the Luxembourg Institute of Science and Technology (LIST) to create the European Space Resources Innovation Centre (ESRIC) in Luxembourg. Since the launch of the SpaceResources.lu initiative in 2016, ESA and LSA have identified several opportunities for cooperation in the field of space resources. As a means of jointly advancing these common objectives, a Memorandum of Cooperation was signed between the two organizations in 2019. ESRIC, launched in November 2020, aims to become the internationally recognized center of expertise for scientific, technical, business and economic aspects related to the use of space resources for human exploration and for a future in-space economy. It will foster open innovation in this field by providing access to unique research facilities and experts, support technology transfer between space and non-space applications, and nurture the emergence of a lunar economy by supporting related business-driven

initiatives and incubating start-ups. ESRIC will apply a comprehensive knowledge management approach to the benefit of the international space resources community engaged in related research and applications. In addition, the Centre will facilitate dialogue and information exchange between community members worldwide, and promote public and private investments in space resources-related activities.

As a result of the SpaceResources.lu initiative, a large number of new commercial companies have been established in Luxembourg, expanding the country's existing space sector. Some of them target space resources utilization in the long term, while pursuing short-term goals that allow them to create sustainable and viable businesses. Overall, it can be concluded that the success of SpaceResources.lu is based on a highly forward-looking political decision, coupled with an efficient implementation addressing various challenges through partnering with public and private actors on a national and international level. It shows that governments and space agencies are able to act as first movers encouraging the development of innovative fields such as space resources exploration and utilization.

31.4 Summary/Conclusions

Realizing the potential benefits of space resources remains challenging and requires global partnerships. There are technical and commercial challenges, but also regulatory and financial ones. The technical capabilities necessary for enabling the entire value chain from prospecting to utilizing space resources must be developed, tested, and matured. Identifying capabilities needed for space mining that also have near-term business potential, especially in terrestrial use cases, is required to ensure incremental growth as the market today is nascent and will need time to mature.

International and domestic legal frameworks need to be established to protect entrepreneurs, reassure investors, and ensure responsible business activities. Significant investment with long time horizons must be available to enable private firms to develop and deploy critical systems.

Although space resources utilization still has some uncertainties requiring planners to make several assumptions, analysis of the different value chains reveals a number of promising aspects for the future of this important industrial sector.

There are opportunities across all stages of the value chains that can be leveraged to create commercial benefits. In particular, early prioritization of specific space resources and uses has been identified along with areas for further technology developments. The collaboration between space players and the terrestrial mining industry is expected to play a central role in the successful development of space resources by leveraging the experience and know-how of mining companies.

Importantly, space resources utilization can generate sizeable socio-economic benefits, including economic impacts, employment, market spill-overs, and technology spill-overs. Cost savings assessments have demonstrated substantial cost reductions for space missions, revealing viable markets for space resources products.

Acknowledgements We would like to thank the European Space Agency, as well as the consortium behind the value chain study for their important work.

We are grateful to the participants at the Mining Space Summits for making them such worthwhile events and for their valuable insights and feedback.

We extend our thanks to Jochen Harms from Blue Horizon, Julien-Alexandre Lamamy and Carlos Espejel from ispace Europe, Joost van Oorschot from Maana Electric and Jim Keravala from OffWorld for providing us with input on their current and future plans.

And last but not least, we thank our colleagues at LSA for their input to various sections of this chapter and their daily devotion to making the SpaceResources.lu vision come true.

References

- Major Takeaways of the Mining Space Summit 2018. 2018. <https://space-agency.public.lu/dam-assets/publications/2018/MINING-SPACE-SUMMIT-2018-Whitepaper.pdf>.
- Major Takeaways of the Mining Space Summit 2019. 2019. <https://space-agency.public.lu/dam-assets/publications/2020/MINING-SPACE-SUMMIT-2019-Paper.pdf>.
- Opportunities for Space Resources Utilization: Future Markets & Value Chains. 2018. <https://space-agency.public.lu/dam-assets/publications/2018/Study-Summary-of-the-Space-Resources-Value-Chain-Study.pdf>.

Chapter 32

Lifetime Embodied Energy: A Theory of Value for the New Space Economy



George C. Lordos, Jeffrey A. Hoffman, and Olivier L. de Weck

Abstract Since the dawn of spaceflight, payload mass, or launched mass, has been used as the primary cost metric for mission-level trade studies. However, advances in the reusability and mass production of spacecraft have started decoupling mission cost from launched mass, leading us to consider novel metrics for trade studies and design of space systems. A new metric that is independent of launched mass will be especially useful to designers of complex space systems on the Moon or Mars where some of the system mass may be procured locally by way of in-situ resource utilization (ISRU). Further, since these worlds are hostile to complex animal life, technology and the sources of energy to build, maintain and operate it will be essential for all aspects of human life there. For these reasons, we propose an energy-based value and cost metric, Lifetime Embodied Energy (LEE), to replace launched mass as a useful proxy for the valuation and costing of infrastructure deployed, developed or operated on the Moon or Mars over long-term campaigns. Embodied energy principles have been used to calculate carbon footprints of Earth-based systems, however embodied energy does not and cannot underpin valuations of Earth systems or Earth infrastructure, largely because of the computational overhead imposed by the complexity, depth and breadth of terrestrial supply chains. However, space supply chains will be orders of magnitude simpler relative to whole-Earth networks, motivating the investigation into embodied energy as a universal metric. In this work, the lifetime embodied energy of a space system is the sum of the allocated portions of past and future, direct and indirect energy transformations that were or will be required to deliver

This chapter is based on George C. Lordos' 2018 MIT SM thesis, adapted for an audience specifically interested in space economics frameworks for the strategic planning and development of space re-sources.

G. C. Lordos (✉) · J. A. Hoffman · O. L. de Weck
Massachusetts Institute of Technology, 33-409, 77 Massachusetts Ave, Cambridge, MA 02139,
USA
e-mail: glordos@mit.edu

J. A. Hoffman
e-mail: jhoffma1@mit.edu

O. L. de Weck
e-mail: deweck@mit.edu

the intended value from a system throughout its useful life. The starting points are a system boundary and a common energy source outside this boundary; this energy source can be the Sun, nuclear energy or, as used in the case study presented here, the embodied energy of space logistics. Then, the LEE framework was applied to a simple toy model of a 20-year campaign to Mars, featuring different combinations of energy, resource extraction and processing, manufacturing, assembly and habitat sectors. This model supported the comparison of Apollo-like campaigns vs. in-situ resource utilization (ISRU) alternatives. Using reasonable assumptions for model parameters, the results showed that over the same 20-year time horizon, Apollo-like architectures came in at a specific lifetime embodied energy cost of ~ 210 MJ/kg, but the best ISRU architectures cost under 70 MJ/kg. We further found that these reductions in specific lifetime embodied energy were correlated with increased up-front investment, increased absorption of in-situ resources and longer design horizons. Near-term applications include optimal design of multi-decade Moon or Mars campaigns, such as the development of large-scale industry, habitats and other persistent infrastructure on these worlds. Over the longer term, the LEE paradigm could serve as an objective costing and valuation system. Since LEE can, in principle, accurately value all space systems relative to each other, it can also serve as an anchored unit of measure, i.e. a currency, with which to settle space-to-space transactions in the emerging commercial space economy.

32.1 Introduction and Motivation

The extremely high cost of spaceflight has been the main obstacle preventing the settlement of the frontier lands and worlds of our solar system. The laws of chemistry and physics make it hard, but not impossible, for our species to escape the bonds of gravity that tie us to the world of our birth. Faced with this challenge, over the last half-century we have mastered the efficient combustion of energy-carrying propellants and the efficient exchange of momentum to propel our ships through the ocean of space towards other worlds. But our grades so far in the subject of sustainable space economics have only just started improving. Launch costs to date have been high in practice because launch vehicles and upper stages were typically discarded after their first and only voyage, and because the reusable space shuttles ended up requiring extensive inspection, refurbishment and re-qualification between their infrequent flights. The entrenched reality of high launch costs led to space mission architects investing substantial fractions of total mission budgets in design, development, test and evaluation (DDT&E), so as to obtain maximum performance and reliability for the minimum amount of mass (Jones 2003).

In retrospect, it is self-evident that high launch costs per unit mass have been steering architects towards one-off designs optimized for minimum launch mass. But, by definition, one-off designs amortize their DDT&E costs into one production unit, thereby perpetuating the “vicious circle of mass” shown in Fig. 32.1. This entrenched dynamic, which had locked in long mission development cycles and a

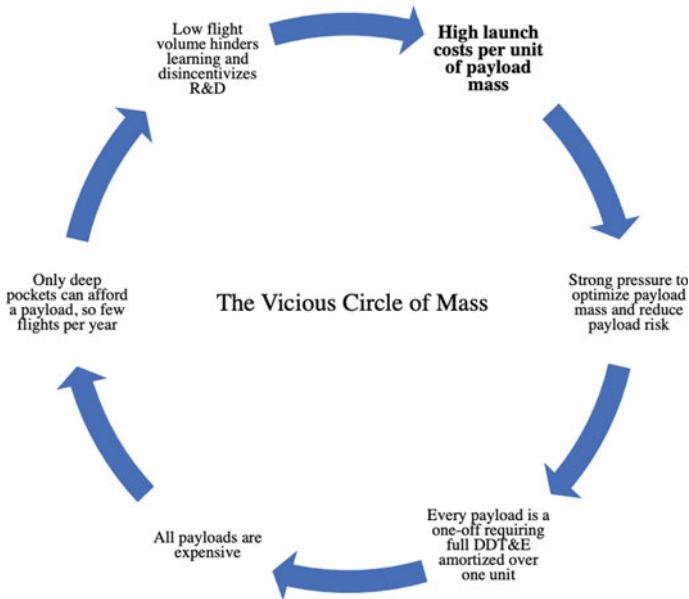


Fig. 32.1 The vicious circle of mass. For several decades, high launch costs led to high payload costs, which led to low flight volumes and dampened competitive incentives to reduce launch costs. Recent innovations in launch-vehicle/upper-stage reusability and modular spacecraft design are breaking through the vicious circle of mass, leading to a new era of spaceflight

high floor for space mission costs across the industry for decades, is currently in the process of being broken by the recent innovation of reusable first and second stages. As a result, the decoupling of payload mass from space mission cost is underway, opening up new approaches for estimating the true economic cost of space systems and missions.

32.1.1 *The Initial Mass in Low Earth Orbit (IMLEO) Cost Proxy*

This focus on treating payload mass as the main proxy for space mission cost had been universal and uncontroversial: space mission architects, from John Houbolt¹ to the present day, have been using Initial Mass in Low Earth Orbit (IMLEO) as a

¹ In 1961–62, NASA engineer John Houbolt was prioritizing minimum IMLEO when he famously persuaded his colleagues at NASA to adopt his Lunar Orbit Rendezvous (LOR) architecture for the Moon landings, leading to the behemoth 15 m diameter Nova rocket being abandoned in favor of the more realistic 10 m-diameter Saturn V design. The wisdom of this decision was later proven by the four pad explosions of the ill-fated 17 m-diameter Soviet N-1 Moon rocket: the N-1 never made it to space.

key figure of merit and as a proxy for dollar cost (Ho et al. 2016). Moreover, space mission architectures generally have been and still are being optimized to minimize IMLEO subject to meeting objectives and constraints. IMLEO as a proxy for cost has especially facilitated tradespace exploration at the early stages of concept generation, selection and development where dollar cost figures would be difficult, cumbersome or otherwise inappropriate to estimate and assign.

IMLEO has held up well for decades as a skilled proxy metric for total space mission cost owing largely to the high launch costs and the vicious circle of mass dynamic outlined above: high launch costs drive total mission cost both directly and indirectly, via the one-off mass-minimization and risk-retirement emphasis which drives very high DDT&E costs. Therefore, in general, reducing IMLEO also leads to reducing dollar cost. However, as shown in Tables 32.1 and 32.2, changes are under way that tend to weaken IMLEO’s suitability as a cost proxy for space mission architecture. These changes include increasing technology development efforts in the field of in-situ resource utilization (ISRU).

The progressive decoupling of IMLEO from true economic cost outlined in Table 32.2 continues to gain traction as of this writing. In the early 2020s, space exploration is being transformed by several interdependent trends: reusability of boosters, upper stages, and landers; plans for persistent and maintainable orbital and planetary

Table 32.1 Linkage between space mission architectural decisions, IMLEO and space mission costs: the historical experience of the past 60 years

Space mission architectural decision	1957–2020	Link between space mission architectural decision and IMLEO	Link between IMLEO and cost of space mission architecture elements
Source of mass and energy for mission	Earth	Mission mass required at destination directly drives IMLEO	Mission payload mass cost element was proportional to IMLEO
Repeat visits to same planetary surface site	No ^a	No pre-emplaced infrastructure, so mission mass directly determines IMLEO	
Number of uses per launch vehicle	One ^b	IMLEO requirement is proportional to total launch mass consumption rate	Launch cost element was proportional to IMLEO
In-orbit propellant refilling	No	ΔV requirement to reach destination from LEO directly drives IMLEO requirement	Transit cost element was proportional to IMLEO

Finding: Total space mission cost used to be proportional to IMLEO

^a Apollo 12 did not make use of the mass of the Surveyor III spacecraft

^b Cost and time of refurbishing the space shuttle was orders of magnitude higher than the cost and cadence of SpaceX building and/or reflying the Falcon 9/Dragon launch system. Hence, the space-shuttle architecture was arguably not truly reusable in the economic sense of the word

Table 32.2 Linkage between space mission architectural decisions, IMLEO and space mission cost: the next 30 years

Space mission architectural decision	2021–2051	Link between space mission architectural decision and IMLEO	Link between IMLEO and Cost of space mission architecture elements
Source of mass and energy for mission	Earth + ISRU	Mission mass employed at destination will increasingly be decoupled from IMLEO requirement	Mission payload mass will in future vary independently of IMLEO
Repeat visits to same planetary surface site	Yes	Pre-emplaced infrastructure may reduce IMLEO requirement for follow-on missions or increase it for early missions	
Number of uses per launch vehicle	Many	IMLEO requirement will no longer consume the entire mass and economic value of a launch vehicle (LV)	Launch cost element will be proportional to IMLEO divided by number of reuses of the LV
In-orbit propellant refilling	Yes	ΔV requirement to reach destination will be met using the IMLEO of the primary as well as of the refueling missions	Transit cost element will in future vary independently of the IMLEO of the primary (payload-launching) mission
Finding: total space mission cost is being progressively decoupled from IMLEO and will increasingly be driven by the mission architecture and its context			

surface infrastructure; modularity and commonality in spacecraft subsystems; and the ongoing mainstreaming of in-space manufacturing and in-situ resource utilization. Thus, as launch costs fall due to reusability of booster and/or upper stage (e.g., SpaceX’s Falcon 9, Falcon Heavy, Starship/SuperHeavy), other cost items which are not necessarily proportional to payload mass increasingly become more meaningful with respect to total mission and campaign cost. Falling launch costs make it economical to launch very large satellite constellations, motivating investment in assembly lines for the mass production of thousands of flat-packed Starlink satellites where the amortization cost of the investment depends more on the total number of satellites than on their mass. Moreover, campaigns of many sequential, similar missions to the same site on planetary surfaces are being planned, for example by NASA and other international actors to the lunar south pole or to Mars Exploration Zones, and by private companies to their own privately conceived outposts or bases on the Moon and Mars. These target sites on planetary surfaces will accumulate persistent infrastructure and are likely to feature in-situ propellant production from local resources. The capability to refill propellant at the destination resets the rocket equation and disrupts both the reach and economics of spaceflight (Lordos et al. 2020).

32.1.2 *An Energy-Based Metric to Replace IMLEO*

In view of the above, a new physics-based metric may be desirable to replace IMLEO as the standard non-monetary, technical measure of value and cost for early-stage concept development and selection. Since energy is the physical measure of work, and all value is created by work, we propose an energy-based metric. And since this metric is envisaged to be used by architects creating long-lived architectures of persistent infrastructure on other worlds and in space, this metric should be calculated and summed over the life cycle of the system being evaluated. Energy-based value and cost metrics such as embodied energy had been developed during the oil crises of the 1970s and are in use today generally only in the buildings energy performance sector and also in comparative macroeconomic studies across countries and industries. Embodied energy is calculated by summing and allocating all the past energy flows required to develop the system of interest. As these energy flows can be estimated analytically from first principles for any well understood physical process, embodied energy as proposed by Odum (1983) is a good candidate for a long-term, physics-based cost metric. Odum measured embodied energy in embodied joules or emjoules, a unit that is defined below in Sect. 32.4.3.1. When we add emjoules already expended to emjoules likely to be expended in the future to sustain a system in operation so that it can deliver its output, we obtain lifetime embodied energy – a generalized, physics-based life-cycle cost for any system.

And so we ask: what if we could develop a new way of assessing the cost and value of all in-space activities *in units of energy*? How might we do that? Might an energy-based value metric be more objective and universal in its coverage, more solution neutral and more directly to the point than any other measure, given that all value creation ultimately requires available energy? Might an energy-based metric be easier to implement than one might at first fear, given the precision with which we should be able to calculate its employment and embodiment at all steps in the (relatively) small value chain of Earth-made and in-situ manufactured space systems? Might an energy-based cost metric find near-term potential in assisting with the optimal architecting of complex campaigns consisting of very long sequences of diverse in-space activities, with myriad alternative future paths? And last but not least, might a more objective, formulaic, energy-based metric one day become highly relevant and useful to investors as a space economy emerges and starts to grow, in the way that options pricing theory serendipitously catalyzed the birth of many financial derivatives markets, which are today measured in trillions of dollars?

Hence the novel energy-based cost metric primarily for human space exploration missions, which has been termed *Lifetime Embodied Energy (LEE)*. The LEE approach is applicable to estimating the costs of in-space activities for any type of space mission, but, as we shall see,² it is especially well suited to architecting extended campaigns of missions to the same site on the surface of a celestial body, where the architect intends to develop an optimally designed, long-lived space settlement ecosystem.

² See, for example, Sect. 32.5.1 in this chapter.

32.2 Overview and Limitations of Current Value Systems

Over recent decades, the gradually weakening correlation between IMLEO and true economic cost has led system architects and space economists to develop alternative space mission cost analysis methodologies and metrics. However, as we show below, all these metrics implicitly or explicitly end up being correlated with launch mass, payload mass or upmass as the main driver of total economic cost and as the key figure of merit to optimize. As a result, none of these metrics will be ideal for a future space systems architecture world that is not characterized by the vicious circle of mass. These alternatives to IMLEO, which are currently in use by NASA and the industry, include the following.

32.2.1 Equivalent System Mass (ESM)

Previously Equivalent Mass (EM) (Levri, Vaccari and Drysdale 2000) Equivalent System Mass (ESM) is a metric used only for advanced life-support systems (ALS). ESM³ is the direct mass cost of the life-support system under study, plus the allocated indirect mass requirements for all or some of power, cooling, consumables, spares, crew time and structural volume, enabling the proper trade-off of life-support systems against a mass-based cost metric. ESM is effectively an implementation of an activity-based costing system with standard costs and cost drivers, where standard costs are defined for every mass, volume, power, cooling and crew time requirement at different locations. This database of standard costs is found in the Life Support Baseline Values and Assumptions Document (BVAD) (Anderson, Ewert and Keener 2018), which is updated regularly.

The key point about ESM which is of interest to the lifetime embodied energy argument is that in all cases, the quantity being allocated, summed and traced is *mass*, and specifically the system dry mass of a payload that had originally been launched from Earth. Therefore, ESM is in essence an elegant method to optimize space system designs for minimum IMLEO and as such it does not address the issue at hand, which is that IMLEO is losing its relevance due to the new trends in spaceflight.

32.2.2 Life Cycle Cost (LCC)

Life Cycle Cost (LCC) (Jones 2003, 2015) includes all costs incurred in the life cycle of a project, from analysis and definition to DDT&E, launch and emplacement and operations costs for the lifetime of the project, typically tabulating and reporting the

³ During the Apollo era, equivalent mass (EM) included mass, volume, power, cooling and materials and spares logistics. In 1992, crew time spent maintaining the systems was incorporated in EM, and in 1998 EM was renamed Equivalent System Mass (ESM).

cost as \$/kg of payload. Line-item costs in LCC calculations are estimated using parametric cost models, including the Advanced Missions Costs Model (AMCM), the NASA-Air Force Cost Model (NAFCOM⁴) and the commercial PRICE-H space hardware cost model. Life cycle costing models sit one level above non-recurring-cost DDT&E models and recurring-cost operations cost models, incorporating their outputs as inputs and producing multi-year cost phasing as their outputs. Launch and emplacement cost is driven by system (i.e., payload) mass, destination, timeframes and orbit selection, and the operations cost component of life cycle cost analysis relies partially on DDT&E as a cost driver (Jones 2015). All LCC models are based on analogs drawing from historic data of past projects, and all estimate DDT&E using system dry mass as the main input (Lordos 2018, Table 3). Difficulty, complexity, novelty, learning and number of units also factor in, but as we saw with the vicious circle of mass, all of these are correlated with the minimum-mass imperative, and therefore with mass itself.

Summing up, in life cycle cost models, system dry mass directly drives estimates of launch cost *and* also has a substantial indirect effect on estimates of non-recurring engineering and operations costs via the established models for estimating DDT&E. Therefore, optimizing LCC implies optimizing IMLEO and as a result, LCC as currently calculated cannot fully replace IMLEO as a new value and cost metric that would be appropriate to the new space economy.

32.2.3 Life Cycle Mass (LCM)

Life Cycle Cost (LCC) was an effort to merge and improve on both ESM, which was only used in ALS, and LCM, which was used throughout NASA. Dollar life cycle costs, including development, launch and emplacement and operating costs, are converted from millions of dollars to mass using a derived parametric equation. LCM was derived from the established life cycle cost (LCC) and first proposed (Jones 2004) as a “more inclusive and potentially better metric” than EM/ESM. The life cycle costs for (i) development, (ii) launch and (iii) operations are estimated in millions of dollars and then converted into mass.

Development costs for LCM purposes can be estimated in a variety of ways, with Jones’ preferred approach being the top-down Advanced Missions Cost Model (AMCM) (Guerra and Shishko 2000) which consists of one parametric equation with five⁵ input variables, derived by fitting to historic cost databases and given as:

$$\text{Cost} = 5.65 * 10^{-4} Q^{0.59} M^{0.66} 80.6^T G^{-0.36} 1.57^D$$

⁴ As of 2015, NAFCOM was being replaced by the Project Cost Estimating Capability (PCEC) (*ibid.*).

⁵ A sixth variable, time until entry into service, is also typically part of AMCM, but it was not included in the Jones (2004) formulation.

where:

Cost is expressed in units of millions of 1999 dollars,

Q = total number of units, including development and production units

M = system dry mass, in units of pounds

T = type of mission unitless parameter (2.46 for crewed planetary missions)

G = hardware generation unitless parameter (new design = 1, second generation = 2)

D = difficulty scale unitless parameter (e.g., -2 for very easy, 0 for average, $+2$ for very difficult).

From this equation, it is evident that once again, there is an implied and direct relationship between the development cost component of LCM and system dry mass. Within any given family of novel, difficult missions (such as for crewed missions to the surface of the Moon or Mars), all the terms of the equation except those based on Q and M reduce to constants. Thus, for any given mission profile and total number of required units Q , more mass translates monotonically into more development cost.⁶ This mass-driven development cost is then an input into LCM.

The launch cost component for LCM is typically estimated using mass ratios. For example, for a one-way Mars landing, or for a Moon landing and return to LEO, Jones gives a typical initial-to-payload mass ratio of 20. This initial mass is multiplied by a typical launch cost per kg, which Jones had taken as \$25 k/kg for space shuttle (Jones 2004). So in the case of launch cost, the dependence of LCM on system dry mass is direct and strongly amplified.

Finally, the operations cost component for LCM is estimated as a percentage of the system development cost per year, in accordance with the JSC Mission Operations Cost Model (MOCM). However, for a given mission profile development cost is already mostly driven by mass.

Therefore, in the case of LCM, the development, launch and operations costs components for a crewed mission to the Moon or Mars are once again, directly or indirectly, a function of system dry mass. Despite the inclusion of development costs and operations cost, we note that in the end LCM hews closely to the same launch-mass-focus paradigm, just like ESM and LCC.

32.2.4 Network Flow Models (Mass-Based)

All the above NASA-developed metrics implicitly assume a carry-along strategy, where all the mass required for the mission is launched from Earth. As we have described above, this implicit assumption will not apply for future human missions to the Moon and Mars, which will rely extensively on ISRU for life support and fuel.

⁶ This vicious circle of mass also tends to flatten design, lengthen system iteration cycle times by orders of magnitude, and eliminate degrees of freedom for space architects. Compare the innovative design and rapid evolution of the Starlink system, which is not subjected to the vicious circle of mass, vs. the Iridium/Iridium NEXT constellations which could not break out of the vicious circle.

The promise and potential of ISRU, however, has led researchers to develop more sophisticated analytical methods using network flow theory:

Generalized Multi-Commodity Network Flow (GMCNF) (Ishimatsu et al. 2015) is a method for designing and optimizing space missions in Earth–Moon–Mars space where lunar ISRU water production is assumed to be available, for example, enabling refueling of a mission to Mars from lunar resources. The optimization goal is minimum total IMLEO across all missions in the campaign. Once it is set up and running, GMCNF sums up the cost flows along the arcs to calculate and minimize MLEO. The sources of commodities which yield reductions in total IMLEO are ISRU propellants and consumables. The end result is still an exercise in minimizing IMLEO, and as a result the implicit launch mass focus is present in this approach, as well.

Time-expanded Networks (TEN) (Ho et al. 2016) is an improvement on GMCNF which introduces a temporal dimension in the analysis to solve chicken-and-egg problems (e.g., ISRU relying on itself), thereby improving realism and fidelity. Again, however, the optimization target is minimum total IMLEO.

Integrated Space Logistics Mission Planning and Spacecraft Design (Chen and Ho 2018) is an optimization framework which builds on GMCNF and TEN to allow for the simultaneous consideration of mission planning and spacecraft design, using mixed-integer nonlinear programming. Results show significant reductions in, once again, IMLEO relative to traditional mission-level design.

32.2.5 *Utility Theory Approaches*

One more general method to measure value or opportunity cost, which can sidestep the problems with mass, is utility theory. Specifically, for tradespace exploration of space system architectures, one method to tabulate benefit accruing to multiple stakeholders and the accomplishment of different objectives is Multi-Attribute Tradespace Exploration (MATE), a comprehensive approach which includes need identification, elicitation of preferences and exploration and evaluation at the architecture and the design levels (Ross 2003). The key step is the generation of a value function which aggregates and reflects the decision maker's various preferences. This implies the need to select relative weights among these preferences, and also to map how the performance of the system translates to satisfaction, or utility, in each preference dimension. Ultimately, this enables the scoring and ranking of all the alternative system designs in accordance with the value function. The weighted utility function of the decision maker is then the benefit function, and there is also some cost function, with the two coming together into a tradespace of alternative architectures or designs. However, utility is by definition in the eye of the beholder, so the original question—which physical figure of merit to actually measure as a proxy for cost or value—remains unaddressed.

32.3 Lifetime Embodied Energy (LEE) Modeling Foundations and Approaches

32.3.1 Energy Theories of Value

Life, as Ludwig Boltzmann pointed out in 1886, is primarily a struggle for available energy. All natural and artificial systems which survive the selection struggle perform a value-delivering function and exhibit some order. They rely on flows of available free energy⁷ to sustain their capability to perform their value-delivering functions and to maintain their state of order. This follows directly from the second law of thermodynamics: in the complete absence of available free energy inflows and assuming realistic systems where not all processes are reversible, it would be impossible to repair the inevitable degradations imposed on isolated, highly ordered systems by the second law. It would also be impossible to create any highly ordered systems out of an initial state of disorder without access to a high-quality source of available free energy. In this sense, *all* physical value creation takes place on foundations laid down by the laws of thermodynamics, and *only* on these foundations.

Intuitively, we can easily grasp and appreciate the possibility that physical value creation might turn out to be inseparable from its thermodynamic foundations, as Boltzmann argued. Yet, economists from Adam Smith in 1776 to Solow and Swan in 1956 eschewed the role of energy in value creation, instead attributing economic output to capital and labor.⁸ Their models treated capital and labor as independent and substitutable for each other via the intermediation of markets and the price mechanism. The direct and indirect essential high-quality energy flows, which enable capital and labor to create physical value, do not form part of these canonical economic models (Cleveland et al. 1984). There are indications that this choice of system boundary may be leaving something out: Solow, for instance, attributed the substantial observed divergences in the growth rates of income and the factors of production to an exogenous “technological progress” factor. He did not explore whether technological progress itself was enabled by the availability of high-quality energy surpluses. Earlier in the twentieth century, however, some physical science scholars had taken on this challenge of investigating the relationship between economic growth and energy. The Nobel laureate chemist Frederick Soddy, who had crossed disciplinary boundaries to consider issues in economics and finance, stated that “If we have available energy, we may maintain life and produce every material requisite necessary. That is why the flow of energy should be the primary concern of economists.” (Soddy 1933).

⁷ These flows of free energy, in the formulation of this embodied energy modeling approach, can be in the past, present and/or in the future. It is the notion of *reliance* on these flows which is of the essence here.

⁸ Partly yielding to criticism from Nicolas Georgescu-Roegen, Solow modified his production function to include resources, but maintained the claim that capital can fully substitute for resources, mediated only by marginal prices. This was described as a “conjuring trick” by Georgescu-Roegen (Daly 1997).

Building on Boltzmann's and Soddy's insights, a new generation of ecological economists (Cleveland et al. 1984; Costanza and Herendeen 1984; van Zon and Yetkiner 2003; Ayres and Warr 2005) modified Solow's classical production function to incorporate the contributions of high-quality energy flows in the modeling of the GDP of the United States for the twentieth century. Their modeling and empirical work demonstrated the improved fit of growth and production models which explicitly incorporate economic interactions based on energy theories of value. In Fig. 32.2, we show a cross-sectional analysis of embodied energy inputs and dollar output for the US economy in 1963 and 1967. The log-linear relationship indicates that embodied energy inputs are correlated to dollar economic output.

However, these modifications to dominant economic models never became mainstream. When compared to the fields of physics, chemistry and biology, all of which underwent fundamental transformations in the twentieth century and are continuing to do so in the 21st, the field of economics may be due for re-baselining along the lines of embodied energy.

32.3.1.1 Energy Flows Within and Across System Ecological Boundaries

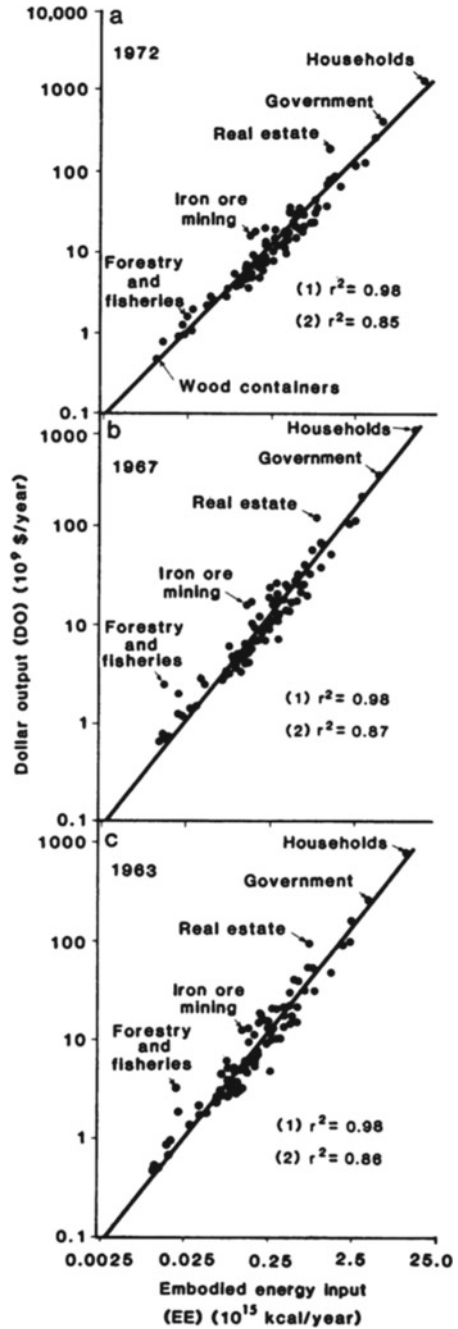
Howard Odum's energy language (Fig. 32.3), which was developed in a systems ecology context alongside his definition of embodied energy, permits the modeling of all types of natural and artificial systems at all scales, from a simple farm to the entire world economy (Odum 1983). In this work, we have adopted Odum's energy language as a modeling framework for calculating the lifetime embodied energy of space-based ecosystems.

32.3.1.2 Lotka's Maximum Power Principle

The maximum power principle (MPP) postulates that, in any survival-of-the-fittest scenario, such as biological evolution, market competition, or choice under economic scarcity, the systems that tend to be selected are those which generate the maximum useful power. This is based on the observation that such systems will be able to absorb the maximum available energy from the available sources for their maintenance, improvement, competitiveness and/or growth. The MPP was first enunciated by Lotka in 1922 and studied further in Odum (1983, p. 141). Odum operationalized the MPP through his energy-language depictions of "autocatalytic modules", high-order systems that tend to evolve via feedback reward mechanisms to maximize the flow of useful energy.

A simple example of the MPP in action is that of the family-owned farm, depicted using Odum's (1983, p. 9) energy language (Fig. 32.3). In the energy-language diagram of the family-owned farm (Fig. 32.4), arrows from left to right generally indicate flows of energy or matter, and arrows from right to left generally indicate flows of information and control actions. The food output enables the family to have

Figure 32.2 Embodied energy inputs on the horizontal axis strongly correlate with sectoral dollar-value outputs, supporting the claim of ecological economists that energy theories of value should be a basis for economic production functions. (Image: Cleveland et al. 1984)



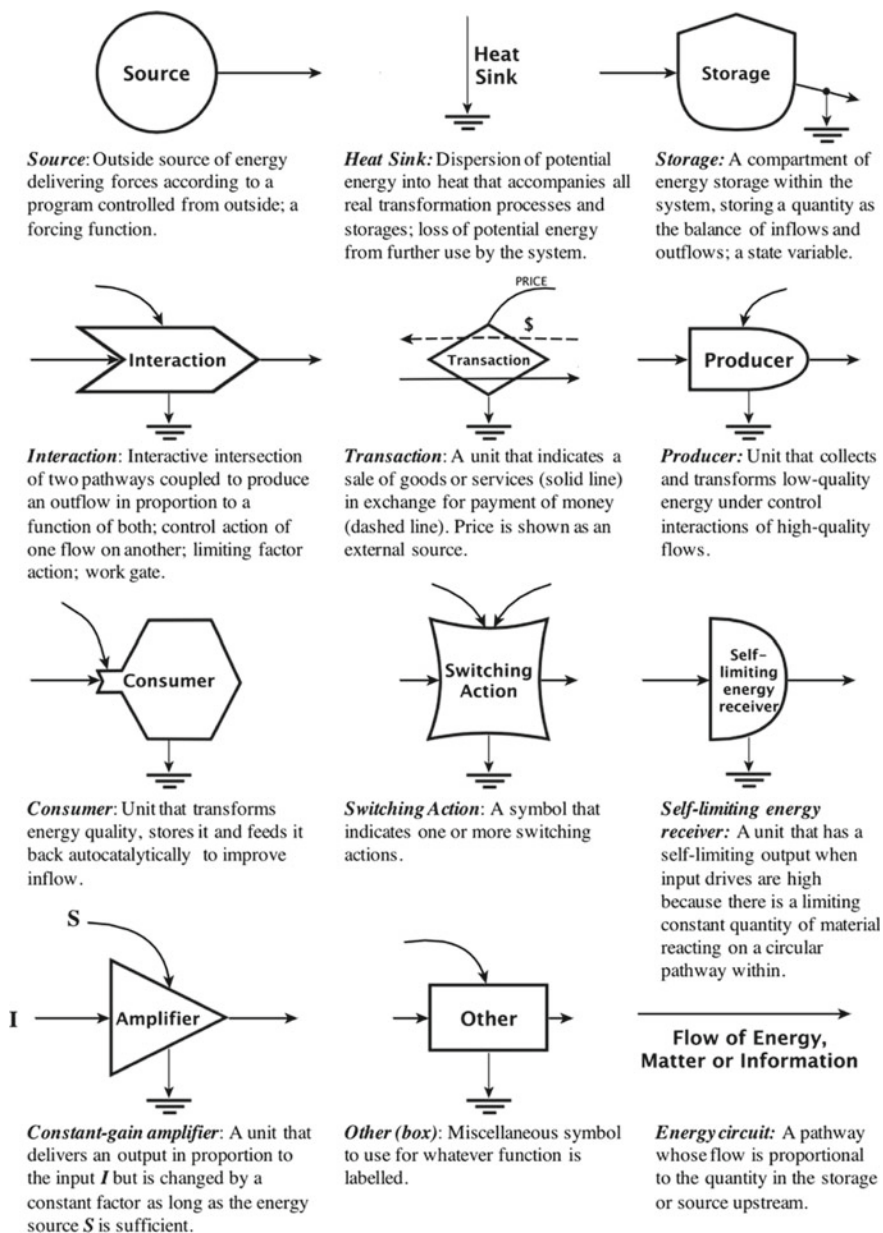


Fig. 32.3 The symbols used in Howard Odum's Energy Language. Any natural or artificial system, including all exchanges of energy, matter and information, can be modeled using energy-language diagrams. *Source* (Odum 1983, Fig. 1.4, 1996, Fig. 1.2)

32.3.1.3 Embodied Energy as Applied to Energy Performance of Terrestrial Buildings

Currently, the concept of embodied energy is in widespread use in the measurement of the energy performance of buildings. This measurement of performance has been mandated by law in the European Union and elsewhere. Life-cycle embodied energy modeling accounts for the energy expended in construction and demolition, both directly and indirectly, as well as the energy expended during operations of the building for its purpose, as shown in Fig. 32.5 (Dixit et al. 2010).

As can be intuited from the model diagram in Fig. 32.5, the computational overhead of embodied energy modeling can be enormous. This is due to the need to follow up on long chains of indirect embodiment of energy until one reaches a set of

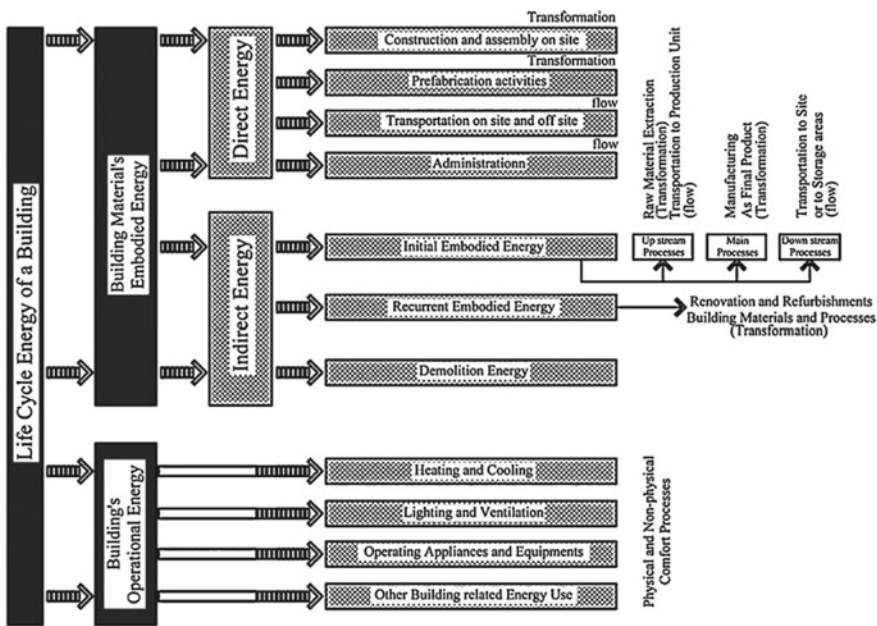


Fig. 32.5 Embodied energy modeling (Image credit: Dixit et al. 2010)

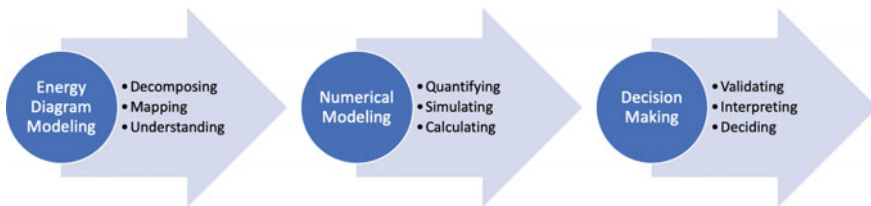


Fig. 32.6 LEE modeling framework

processes wherein the majority of the directly traceable input is energy. This is not practical to perform from first principles for every case, so one approach to streamline calculations is the tabulation and periodic updating of standard, widely accepted embodied energy values for common building materials. An extensive database of embodied energy for common building materials is maintained by Hammond and Jones in the United Kingdom (Hammond and Jones 2008).

32.3.2 Embodied Energy Modeling

Lifetime Embodied Energy (LEE) is defined as the thermodynamic sum of past, present and future work required to create, operate, maintain and decommission a system, including appropriate shares of indirect contributions from upstream systems as well as from other systems in a system-of-systems. The units of LEE are embodied joules; embodied joules are not conserved—they can be created and destroyed depending on losses to entropy and the method of calculation.

The important methodological choice of primary energy source and system boundary in embodied energy modeling is linked to the nature of the underlying figure of merit targeted for optimization in each specific case. For example, by specifying the embodied energy of space logistics as the primary energy source and common cost denominator of all technological systems and processed resources delivered to the Moon or Mars, it is possible to closely replicate the conclusions of standard IMLEO-based or equivalent system mass (ESM)-based analyses, but also to go beyond them.

There are three main methodological approaches to embodied energy analysis: sectoral input–output models, process-based models and hybrid models. The depth and breadth of Earth-based supply chains make the application of process-based models on Earth challenging, leading to the truncation of process trees and to large error bars. These difficulties would not apply to a Moon or Mars industrial development context, thereby making application of the LEE methodology easier for space systems than it is on Earth.

32.3.2.1 Process Tree Models

The main benefits of this widely used analysis method are the precision and reliability of the calculation result. In the energy performance of buildings use case, the method typically decomposes the building into its component materials and then follows each material upstream to account for all the energy input used in its manufacture. The main disadvantage is that the tree branches grow exponentially before good accuracy is achieved, leading to computational complexity. In practice, the analysis is truncated before reaching the end points at the base of the tree, leading to system incompleteness of up to 50% and errors up to 10% (Dixit et al. 2010).

32.3.2.2 Input/Output-Based Sectoral Models

This type of analysis is top down and it starts by decomposing an entire economy and, with more granularity, an industry of interest, into sectors which exchange products, services and money. The data is typically provided by a national government, so the analyst's task is to assign energy intensities to the measured money flows, thereby converting all inputs and outputs into units of energy. This method has the advantage of completeness, since it includes everything within the economy's system boundary, but it suffers from several classic modeling issues that arise when the scope of the model is an entire economy, such as uncertain assumptions about groupings, homogeneity, mapping, and even data entry accuracy. These insurmountable issues make the method error prone; Dixit et al. (2010) found errors ranging up to 50%.

32.3.2.3 Hybrid Analysis

A hybrid analysis combines the best of both worlds from the two methods, using process analysis up to one stage upstream of the process being analyzed and then completing the calculation using input–output analysis. This makes it possible to attain both system boundary completeness using I/O for the distantly indirect inputs, and precision in the summation of a substantial fraction of the direct and one-level-removed indirect energy inputs, while mitigating the penalty of computational complexity (Dixit et al. 2010).

32.4 Constructing a Simple LEE Hybrid Model for a Space Settlement Use Case

32.4.1 Model Structure and Overview

A hybrid input–output and process model was created using both energy-language diagrams and a computational spreadsheet model,⁹ taking advantage of the fact that the scope for a Mars outpost is orders of magnitude simpler than the US economy. The process begins with energy diagram modeling, which is an essential precursor to both the development and comprehension of a numerical model simulating the physical system under consideration. Simulation results in LEE values and other figures of merit, such as 'useful mass emplaced on Mars', for the subsystems and also for the system as a whole. The output of the simulation under alternative scenarios can be

⁹ There are differences in naming of sectors, level of detail, and the fact that the computational model uses a single period (no time dimension); the more detailed model replicating all the features of the energy diagrams is work in progress for George C. Lordos's doctoral dissertation (as at September 2021).

used to inform selection from alternative architectures on the basis of LEE cost, or on the basis of ratios such as benefit delivered per unit of LEE cost expended.

32.4.2 Model Assumptions and Inputs

As we have seen, lifetime embodied energy can be calculated for a variety of system boundaries and associated primary energy sources. For this case study, we have set the system boundary such that the primary source of energy will be the energy of space logistics, and we disregarded the energy cost of manufacturing a system on Earth. These choices enable direct comparisons with IMLEO-based costing systems as well as tradeoffs across the entire range of Mars surface system architecture elements, with immediate feedback to the embodied energy cost metric.

We also set a fixed campaign length (a time horizon) of 20 years, and keep the mass transported to Mars constant for all scenarios. This allows calculation of LEE for alternative campaign architectures, with or without in-situ resource utilization (ISRU) or in-space manufacturing (ISM), while holding other factors constant. These assumptions were made in order to support clearer comparisons between the LEE of alternative architectures that rely or do not rely on space resources and ‘living off the land’.

32.4.3 Model Elements

The major elements in our simple model are the Power, ISRU, ISM and Habitat sectors, depicted as Producer modules using Odum’s energy language. As we can see from a comparison of Fig. 32.4 with 32.7, the models of the farm and of the Mars settlement are analogous: the hexagonal Habitat/Household sector provides control and high-quality labor to the Producer sector(s), which in turn support the proper performance and long-term growth of the Habitat/Household sector: the relationship between Producer and Consumer modules is symbiotic and autocatalytic. The energy-language diagram¹⁰ in Fig. 32.7 shows the fundamental relationships between these four sectors, together with a level 2 view of the inner workings within each sector.

It is interesting to note the rich, emergent feedback relationships between the sectors and systems. Looking closely, we can identify and map out the significant

¹⁰ Please see Fig. 32.3 for an explanation of the meanings of the symbols, arrows and boundaries used in the energy-language diagrams shown here. For more details on Howard Odum’s Energy Language, see Odum (1983, 1996).

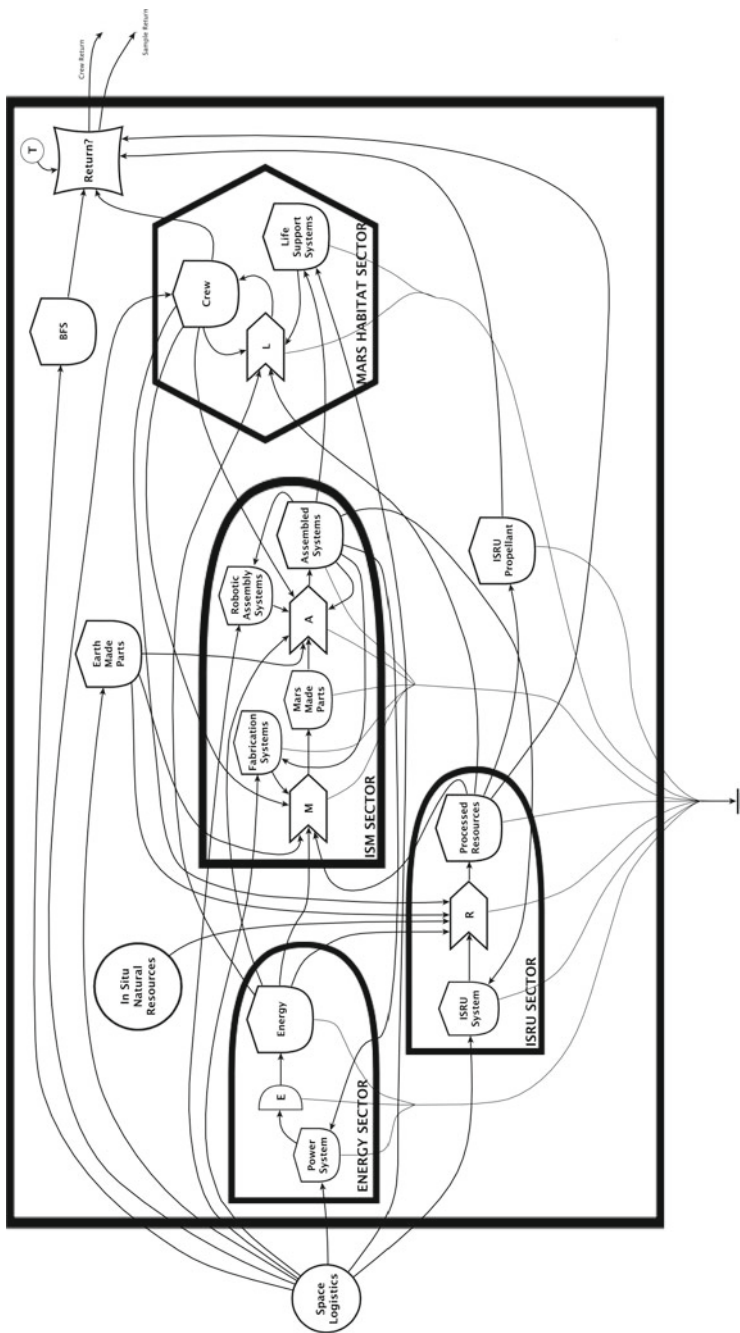


Fig. 32.7 Energy language diagram of the system overview of a generic architecture for an aspiring self-sustaining city on Mars. The logistical supply chain depends on space logistics as well as in-situ natural resources—the circular embodied energy source shown inside the system boundary

autocatalytic relationships between the Habitat sub-model and the other three sub-models, informing their proper simulation and the codification of governing equations and assumptions. In addition, we can intuitively see that there are autocatalytic relationships based on matter and energy flows between ISM and Power, and Energy/ISRU and ISM: the more we can manufacture, the more we can expand the energy and natural resource sectors, and the more energy and natural resources available, the more we can manufacture.

Based on the analysis thus far, a model city on Mars¹¹ which aspires to self-sustaining status is represented by at least four economic sectors: energy, natural resource processing, manufacturing/fabrication and habitation (the latter including food). At the same time, all sectors in a realistic model will, initially at least, require regular supplies of reliable Earth-made parts and systems, delivered to Mars via a space logistics system. Thus, the city and its four sectors can only come into existence once bootstrapped via space logistics. We therefore state that the embodied energy expended in a space logistics effort to create a permanent settlement on Mars will be embodied into the Power, ISRU, ISM and Habitat systems. This will allow us to trace the downstream flows of that energy into systems made on Mars.

32.4.3.1 Energy Sector Model

Since energy is the only output of the Energy sector, the entire embodied energy of the Energy sector is allocated to its lifetime energy output. We can therefore calculate the specific embodied energy of energy, which in our case study will be measured in space logistics emjoules (SLEJ) per joule of energy produced on Mars. Upstream of the Energy sector, we find its two direct sources of embodied energy: Space Logistics and Assembled Systems. The former refers to ready-made energy systems delivered from Earth, and the latter refers to systems assembled at industrial facilities on Mars. Generally, each source has its own upstream source(s) and its downstream sink(s). We can graphically show these relationships by producing a map of the predecessor processes and sources of the lifetime quantity of energy output, as shown in Fig. 32.8.

In the base case where there is no ISM, the embodied energy of the Energy sector equipment is equal to its fair share of the embodied energy expended by its sole remaining upstream input, which is the space logistics effort. This is equivalent to an IMLEO optimization scenario. In the absence of ISM, LEE can easily be converted to IMLEO. The lifetime payload-to-Mars mass of the Energy sector equipment is known to the designer, and the exchange rate between mass and lifetime embodied energy, such as the estimate of ~200 MJ/kg (Lordos 2018) as well as the gear ratio (~10) to convert Mars surface payloads to IMLEO would also be known.

In the case where the architect chooses to include ISM in the design, then again, based on Fig. 32.8, the embodied energy of the Energy sector will be equal to the sum of its fair shares of Space Logistics embodied energy, and Assembled Systems embodied energy. In turn, the embodied energy of the fraction of Assembled Systems

¹¹ Or anywhere, for that matter.

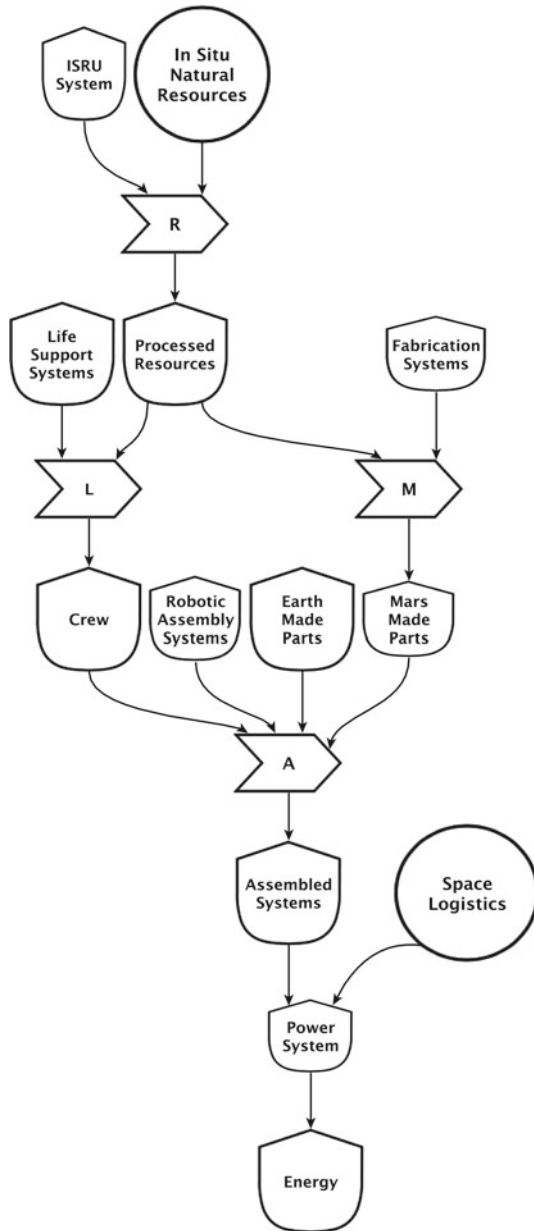


Fig. 32.8 Direct and indirect sources of embodied energy for the energy production sector on Mars. The embodied energy of the Energy sector is equal to its 'fair' share of the embodied energy of space logistics and the embodied energy of Assembled Systems. In turn, the embodied energy of Assembled Systems is derived from the embodied energy contributions from Crew, Assembly Systems, Mars-made Parts and Earth-made Parts. This attribution and allocation process is recursive up to the point of primary inputs to the Mars settlement system boundary

that are systems employed in the Energy sector will be the sum of the fair shares of the embodied energies of the inputs which went into Assembled Systems. These are: the labor of the crew, the work of the robotic assembly systems, the work done to deliver Earth-made parts and the work done to produce the Mars-made parts.

32.4.3.2 ISRU Sector

The storage representing the projected lifetime quantity of processed resources refined from local raw natural resources is shown at the bottom of Fig. 32.9, together with its upstream direct and indirect embodied energy inputs. Processed resources are the output of the resource processing interaction labeled **R**. This interaction requires as inputs the labor of the crew, the work of the ISRU system, the work of Energy and also consumes the mass of extracted raw, in-situ natural resources, as well as an appropriate fraction of the lifetime mass of Earth-made parts.

We calculate the total lifetime embodied energy for each of these upstream inputs, except the raw natural resources,¹² and in each case we allocate a fraction of this lifetime embodied energy to the ISRU sector, according to the relative intensity of each upstream input into a downstream process. In all cases, for each input–output edge, this fraction is calculated by dividing lifetime input at the downstream node (sector) by the lifetime output at the upstream node (sector). This realization provides a way forward to scale up the computation and prevents the double counting of embodied energy.

32.4.3.3 In-Space Manufacturing (ISM) Sector

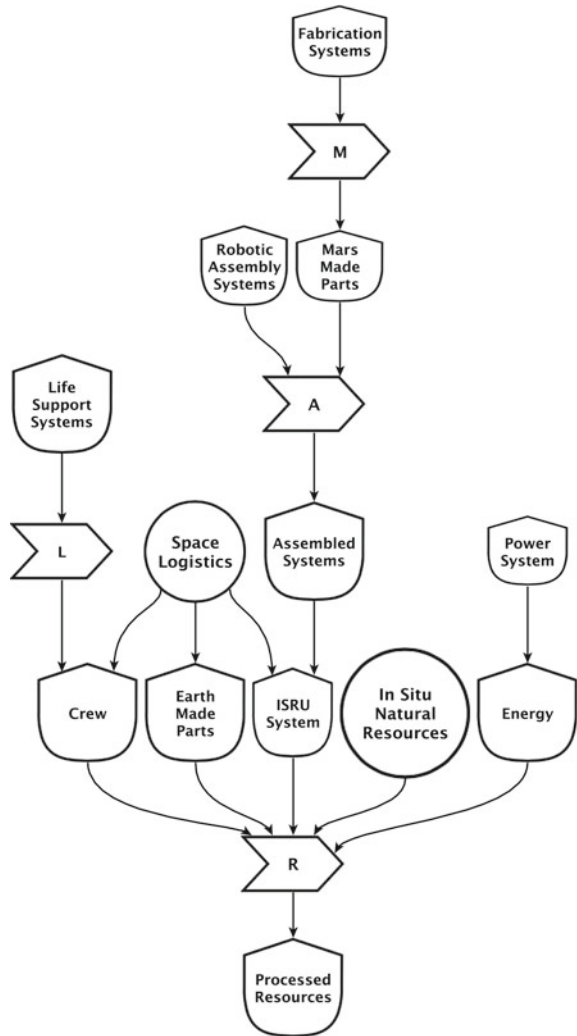
As we can see from the system overview energy-language diagram (Fig. 32.7), the ISM Producer sector consists of two sequential stages: (1) Fabrication using processed¹³ resources, represented by an interaction labeled with the letter **M**, and (2) Assembly, represented by an interaction labeled with the letter **A**. The output of fabrication is labeled Mars-made Parts. The direct and indirect sources of embodied energy of Mars-made parts are decomposed in Fig. 32.10.

In turn, the Mars-made Parts are inputs for the Assembly interaction **A**, along with Earth-made Parts. As shown in Fig. 32.11, Assembly requires the direct work of Robotic Assembly Systems, Crew and Energy, and consumes both Mars-made and Earth-made Parts. These in turn depend on other inputs. At all stages embodied energy flows from upstream and accumulates downstream. The output of the Assembly interaction is labeled Assembled Systems.

¹² We do not count the embodied (solar) energy of natural resources on Mars because the purpose is to evaluate technological tradeoffs for space mission architecture, not to put a price on the natural environment of Mars.

¹³ And recycled, but these are not shown in the model.

Fig. 32.9 Direct and indirect sources of embodied energy for processed resources on Mars. Processed resources are the primary output of the ISRU sector shown in the overview system diagram of the modeled Mars settlement



32.4.3.4 Habitation and Life Support Sector

The Habitation and Life Support sector includes habitats, food and water, environmental control and life support and generally all systems and consumables that are directly required to sustain human life. It is special in that it is both a part of the overall objective function (i.e., science or settlement), and also the main source of indirect embodied energy for crew labor hours. Figure 32.12 shows the direct and indirect upstream embodied energy sources for the Habitation and Life Support systems. This embodied energy associated with Habitation and Life Support must in turn be allocated downstream to the various products of the crew’s labor, according to the same

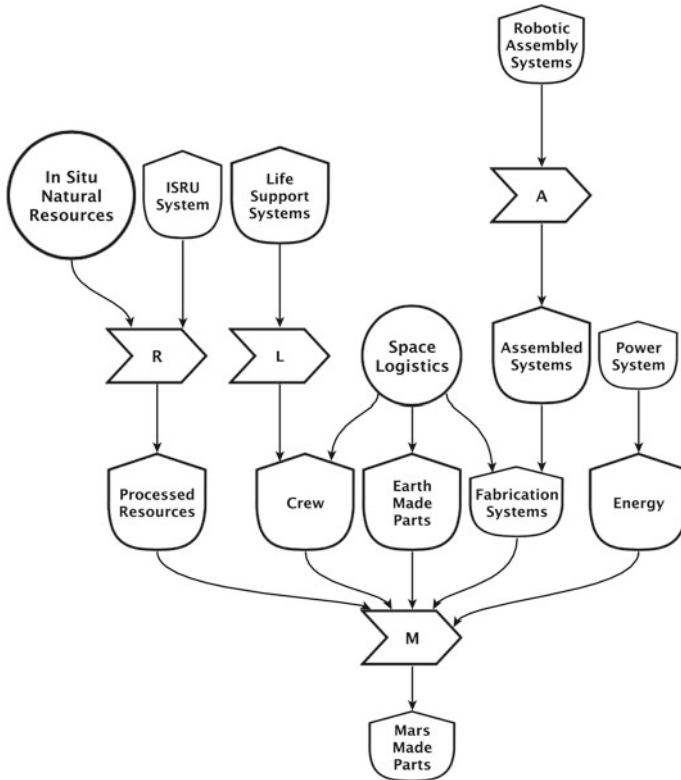
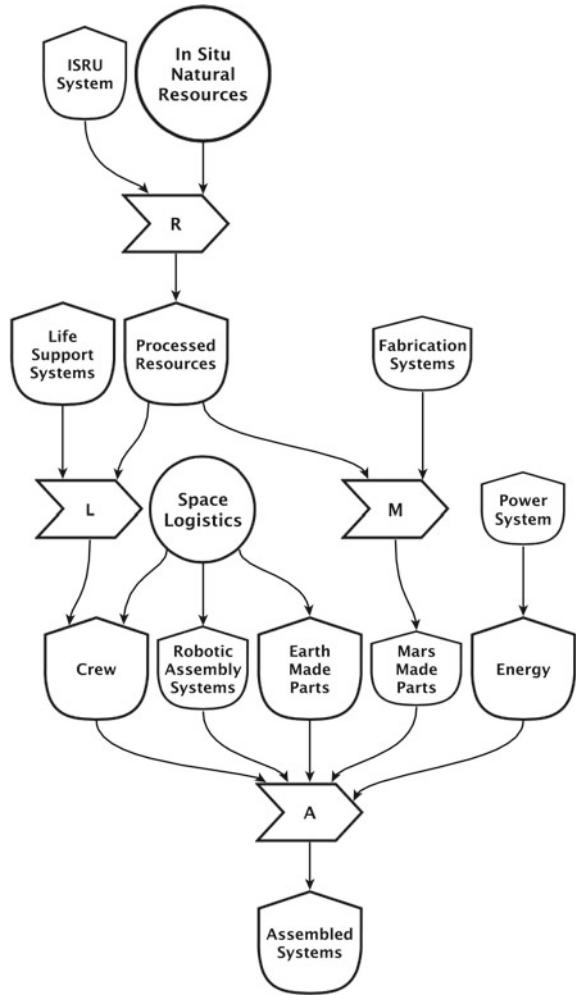


Fig. 32.10 Direct and indirect sources of embodied energy for parts fabricated on Mars. The primary output of the Fabrication first stage of the ISM sector is Mars-made Parts. These parts subsequently become inputs to a second stage, Assembly

principle of pairwise fraction calculation described in Sect. 32.4.3.2. This approach provides a rigorous calculation of the lifetime embodied energy (LEE) of human labor services and also supports the comparison and trade-off between automated and crewed modes using the common-denominator LEE metric.

Figures 32.12 and 32.13 unpack the substantial upstream sources of embodied energy for the Habitation sector and for crew. All systems/sectors on Mars contribute to supporting the lives of the crew. This accords with intuition and is also consistent with Odum’s finding that human labor almost always has by far the highest embodied energy (Odum 1983, p. 490). Of interest for modelers is the fact that this approach rigorously quantifies the embodied energy cost of human labor, in a way that extends NASA’s ESM approach to the more general LEE baseline. This enables the modeler to compare alternative architectures with different levels of intensity of robotic vs crew labor on the same common denominator of lifetime embodied energy. This has been implemented in the Excel model which is presented in Sect. 32.4.4.

Fig. 32.11 Direct and indirect sources of embodied energy for assembled systems on Mars. This process tree represents the second, Assembly (A) interaction in the ISM sector and includes both Earth-made Parts and Mars-made Parts as inputs

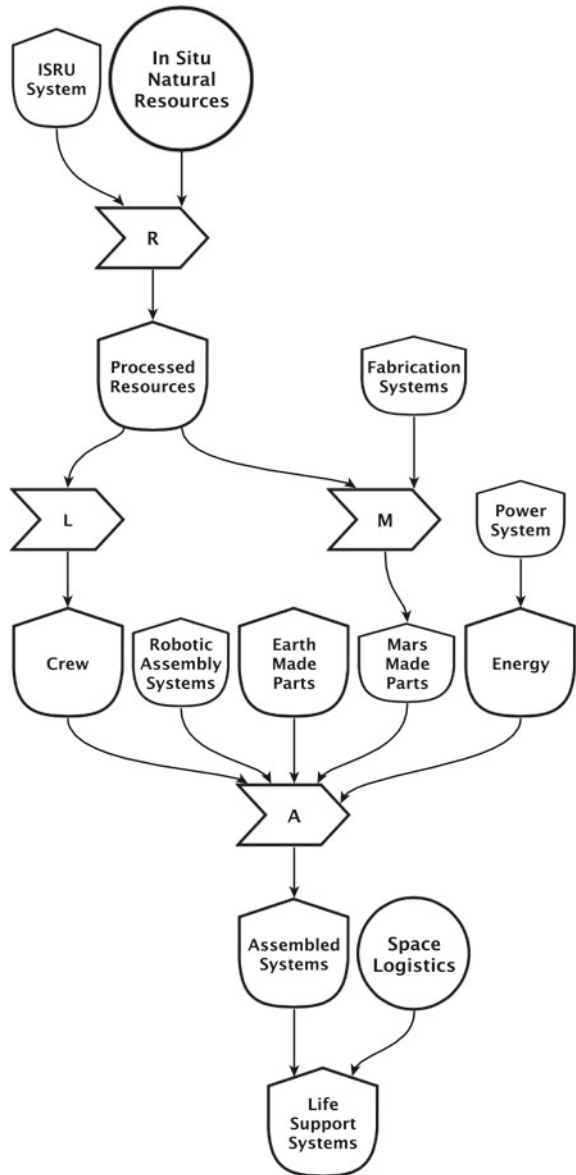


32.4.4 Model Structure and Equations

A computational model has been created which permits the generation of simple Mars surface mission architectures featuring a choice of nuclear or solar energy, whether to adopt recycling, ISRU, ISM, at different levels of automation and versatility. The computational model is organized in sectors and subsectors in the form of a value chain.

- The Imports sector delivers parts, systems and consumables from Mars to Earth, to all sectors that require them.

Fig. 32.12 Direct and indirect sources of embodied energy for life-support systems. Since crew labor is an input to all four sectors—Energy, ISRU, ISM and Habitation—and since life-support systems are required for crew (but not for robots) the embodied energy of life-support systems should be allocated fully to the embodied energy of crew services



- The Energy sector consists of the nuclear, solar and batteries subsectors. Its output, in the form of electrical and thermal energy, is consumed by all other subsectors that follow in the value chain.
- The Resource Processing sector consists of plastics, metal and recycling subsectors. Its output, plastic and metal, is consumed by most subsectors in the Fabrication and Habitation sectors, which are further ahead in the value chain.

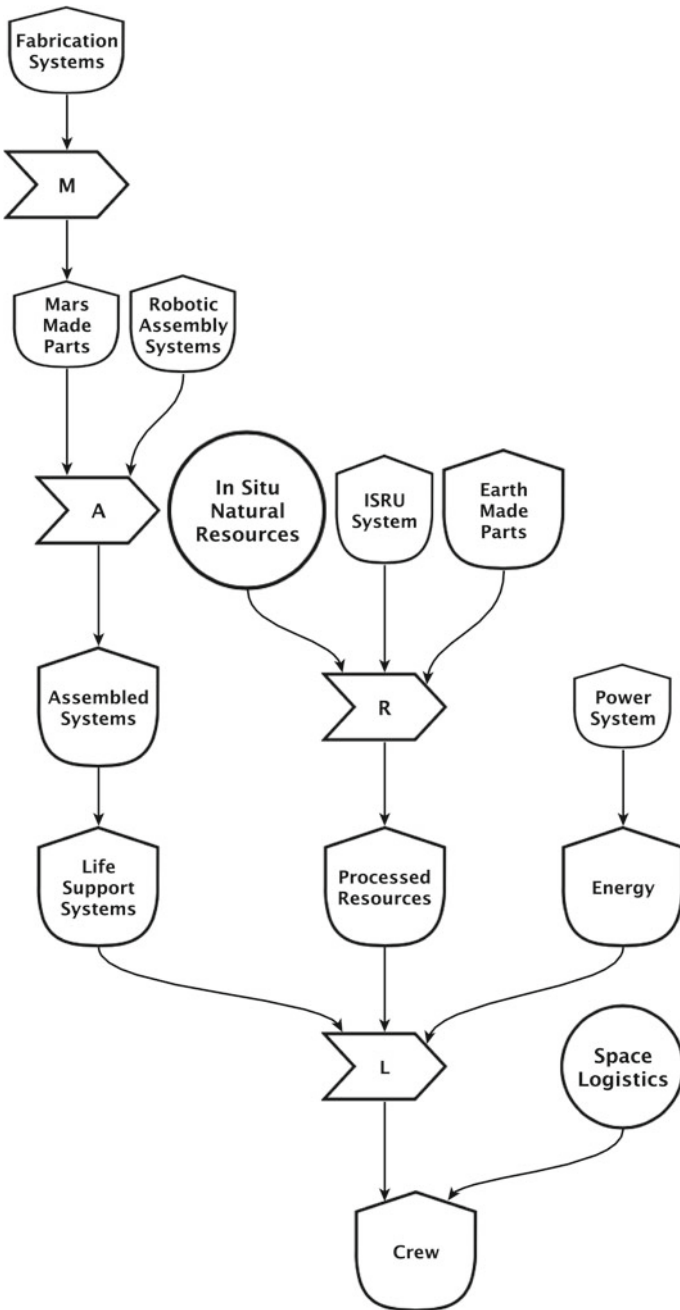


Fig. 32.13 Direct and indirect sources of embodied energy for crew. Notice from the tops of all the upstream branches that the crew literally depends on all the technology brought to Mars, as well as on the in-situ natural resources

- The Fabrication sector consists of 3D printing of plastics, CNC milling of metals and laser cutting of recycled metals. In a more elaborate model there would be more fabrication processes. The output of this sector, together with Earth-made parts, yields assembled end-use systems which are demanded by all sectors.
- The Habitation and Light Industry sector consists of crew including associated food production, habitat including associated ECLS systems, and semi-autonomous assembly systems.

The useful mass transported to Mars includes parts for Energy and Life-Support system subsectors, as well as equipment and end-use systems for the Resource Processing sector, the Fabrication sector and the Habitation and Light Industry sectors.

Transportation of useful mass to Mars takes place using a space logistics service. In this model we have assumed the planned SpaceX Starship service.

Different configurations of architectures are possible, from Apollo-like, where all the required lifetime mass is delivered from Earth, to ISRU-enabled architectures where significant fractions of the mass of systems and consumables are produced in situ from local resources on Mars. For those configurations which tap in-situ resources to manufacture consumables or systems on Mars, the total gross mass emplaced on Mars necessarily increases, but so does the overhead of resource processing and manufacturing systems. Hence, to fairly compare ‘make vs. take’ architectural configurations, the total lifetime *useful* mass on Mars is the sum of all logistical mass delivered from Earth plus all the mass produced on Mars from in-situ resources, *minus* the mass of the resource processing and manufacturing systems which were indispensable in the transformation of the in-situ resources into finished components ready for assembly. Here, useful is defined in terms of the end goals of the humans going to Mars.

The common denominator for evaluating all the alternative architectural configurations is lifetime embodied energy. The principle is simple—the direct and indirect energy consumed by an economic subsector to produce its designated output is considered to have been *embodied* in the subsector’s total output. But the output of one subsector is generally an input to one or more other subsectors. As downstream subsectors consume that output in order to add their own economic value, the embodied energy is absorbed by that subsector, along with the embodied energy of all other inputs, such as direct energy, direct labor and sophisticated parts brought from Earth. All these inputs have different units of measure, but at the base of the value chain of each input, we ultimately find a source of energy.¹⁴ This fact is not only computationally convenient but also physically realistic: hence, lifetime embodied energy is a useful and objective measure of cost that we can use to trade alternative architectural configurations.

In this computational model, to facilitate comparison between ‘make vs. take’ architectural configurations, the energy expended in the transportation of mass from

¹⁴ We also find a source of matter at the base of (almost) every value chain, however as energy is required to separate and beneficiate matter from its natural state, we can assume that all value chains originate in energy sources.

Earth to Mars was selected as the source of all embodied energy on Mars. It is measured in the embodied energy of space logistics using an exchange rate¹⁵ of 206 MJ of embodied energy per kg of useful mass carried from Earth to the surface of Mars. Thus, the embodied energy of all the useful mass transported to Mars is obtained by multiplying mass transported with this exchange rate.¹⁶

To facilitate the comparison of architectural configurations, we set certain variables to be equal. Campaign length is fixed to 20 or 40 years for all model runs, and the total mass delivered to Mars by Starship is fixed at two 80-ton payloads every two years.

For architectural configurations that rely on ISRU and manufacturing on Mars, some of this fixed quantity of mass delivered will be resource processing and manufacturing equipment and spares, together with complex sub-assemblies intended for complex systems assembled by semi-autonomous robots on Mars supervised by crew labor. The more Earth-made complex sub-assemblies are delivered to Mars, the higher the utilization of the resource processing and fabrication systems, the greater the mass of sub-assemblies and consumables that can be produced on Mars, and the more the Energy and Habitation sectors can grow, to support more crew who can then support more systems required for organic growth. In this way, alternative architectural configurations result in different total lifetime useful mass and a different lifetime embodied energy cost per unit of total lifetime useful mass.

One strength of the lifetime embodied energy approach for evaluating alternative architectures for a Mars settlement is that all possible flows of value between sectors can be accounted for, with precision limited mainly by modeling resources. So, for example, all the embodied energy that goes into the habitat sector, including the parts that come from direct inputs of energy, direct transport of ready-made systems from Earth, or life-support systems manufactured on Mars out of a mix of Mars resources and complex sub-assemblies from Earth, is ultimately accounted for as energy embodied in crew labor hours and can be allocated accordingly to processes that use labor hours. This puts a price on labor in units of embodied energy which can be traded against automation, also in units of embodied energy.

All of the above have been encoded in our simple computational model which calculates only lifetime values, with no time dimension. The starting point for the computational model was to specify the variables that the model should maintain for each system or process, as shown in Table 32.3.

¹⁵ This exchange rate is based on a total ΔV of 17.81 km/s plus 15% for inefficiencies, with conversion of the payload mass to energy using the law of conservation of energy. It has also been calculated using an estimated future cost of \$1,400/kg to Mars for the reusable Starship system, multiplied by the energy intensity of the US economy in 2016. In either case we obtain a similar result, 206 to 210 MJ / kg.

¹⁶ For simplicity, we disregard the embodied energy in manufacturing the materials sent to Mars because it would be approximately the same in all configurations.

Table 32.3 Model variables tracked for each system, sector or process

Sector <i>i</i> (system/process, or sector)	
$m_{lf,i}$ Lifetime mass of system, process or sector <i>i</i> including spare parts (kg)	$L_{,i}$ Lifetime labor hours required by sector <i>i</i>
$m_{is,i}$ Of which, lifetime mass manufactured on Mars from in situ resources	$m_{is,i} / m_{lf,i}$ Fraction of lifetime mass manufactured on Mars
$O_{lf,i}$ Lifetime output of system, process or sector <i>i</i>	Output units (e.g. MJ, kg)
$E_{m,i}$ Lifetime Embodied Energy (LEE) of mass transport for <i>i</i> (MJ) $E_{m,i} = e_m \cdot (m_{lf,i} - m_{is,i})$	$E_{s,i}$ Lifetime energy input OR output of sector <i>i</i> (MJ)
$\sum E_{m,ij}$ All direct allocated inputs of LEE into sector <i>i</i> from upstream sector(s) <i>j</i> (MJ)	$E_{s,i} / E_b$ Energy consumed by this sector as fraction of total energy budget
$E_{lf,i}$ Total LEE of system, process or sector <i>i</i> from all sources (MJ) $E_{lf,i} = E_{m,i} + E_i$	$E_{lfsp,i}$ Specific LEE (MJ per unit of output) $E_{lfsp,i} = E_{lf,i} / O_{lf,i}$

32.4.4.1 Decision Options

As the input options are changed, the model recalculates the lifetime mass of systems accumulated on Mars, as well as the embodied energy cost of the architecture. These input options are the following:

- Fraction of surface power from nuclear energy (P_n , from 0 to 100%)
- Is ISM enabled? (True/False)
- Is ISRU enabled? (True/False)
- Is material recycling enabled? (True/False)
- Level of automation in ISRU and ISM (from 0 to 80%)
- Level of versatility¹⁷ of the ISM systems (Low/Medium/High).

32.4.4.2 Lifetime Mass of Systems Accumulated on Mars

In the case of no ISRU, the lifetime mass (m_{lf}) of systems accumulated on Mars is the same as the mass of systems brought from Earth; it is equal to the lifetime mass transported through space over a 20-year campaign, which is also known as logistical mass. The actual amount of lifetime mass can only exceed the logistical mass if ISRU is enabled. In all scenarios, a fixed amount of logistical mass is transported from Earth to Mars, to enable comparison of outcomes. What varies is the manifest: do we transport only machines that support life, or machines that can help make and maintain machines that support life? The energy of space logistics embodied into this logistical mass is found by multiplying mass by the LEE Exchange Rate which for the purposes of this case study is 206 MJ/kg. This figure was calculated in two ways,

¹⁷ A more versatile ISM system can produce a broader selection of different systems and components, thereby allowing the outpost to aim for higher levels of manufacturability. Versatility effectively proxies the technological sophistication of the ISM system.

both simplistic and rough-order-of-magnitude approximate: the first calculation starts from the projected future cost in \$/kg for payload mass delivered to the surface of Mars by a fully reusable transportation system (e.g., SpaceX's Starship). This cost is converted to embodied energy using the embodied energy density of the US economy in 2016, yielding a figure of 210 MJ/kg. The second calculation adds up the total ΔV from the surface of the Earth to the surface of Mars plus 15% for losses and inefficiencies, converts it to energy using the conservation laws and yields a figure of 206 MJ/kg. The latter figure was adopted because it requires fewer assumptions, and only physical assumptions. The full calculation is provided in Lordos (2018, Appendix I).

32.4.4.3 Lifetime Mass of Systems Made from In-Situ Resources

For any given subsystem, the part of the lifetime mass of that subsystem produced from in-situ resources, m_{is} , is given by equations of the following form:

$$m_{is} = m_{lf} \cdot P_n \cdot w_n \quad (32.1)$$

Equation (32.1) is specific to the example of nuclear energy systems, where m_{lf} is the lifetime mass of energy systems including spares required over the fixed campaign horizon (20 or 40 years),

P_n is the fraction of nuclear (Kilowatt) systems in the energy mix, and

w_n is the average lifetime mass fraction of the nuclear power system that is manufacturable from in-situ natural raw materials.

This choice of model structure signals the designer's anticipation that heavy energy systems which require substantial structural or thermal mass can be partially constructed out of in-situ materials. This will be assumed to be feasible for low-technology structural and thermal mass. However, for resource processing, fabrication and robotic assembly operations, in this case study of an early outpost on Mars, m_{is} is assumed to be zero, signaling that for those productive sectors which will directly be generating other mass from in-situ resources, the equipment will be pre-optimized for performance and reliability from Earth, and that the outpost will initially rely on having sufficient spare parts instead of producing machines that can make machines.

32.4.4.4 Solar Panels

The projected lifetime output of the different systems, processes or sectors is estimated in a different way for each sector, according to the physics and economics of the situation. Here is the form of the equation for O_{lfp} , the lifetime energy output of solar panels, measured in watts:

$$O_{lfp} = O_{pk} \cdot t \cdot 365 \cdot lh \cdot 60 \cdot 60 \tag{32.2}$$

Equation (32.2) is specific to the example of solar energy production, where O_{pk} is the lifetime-average peak power of the solar panels in watts, t is the time horizon of the campaign in years, typically 20 or 40, and lh is the average hours of daylight, in hours per day, typically 12.

Equations such as (32.2) which calculate the useful economic output of a sector are subsequently used by downstream processes to allocate the embodied energy of upstream sectors. In the case of energy, the total lifetime embodied energy of the energy production sector is fully and fairly allocated among all energy-using sectors in proportion to the fraction of their energy consumption divided by the total lifetime energy output, Eq. (32.2).

32.4.4.5 In-Situ Resource Processing

For simplicity, it is assumed that demand from manufacturing directly drives the production output of resource processing. In turn, manufacturing need is determined by the availability of extra Earth-made parts that are deemed to be needed in fixed proportions to Mars-made parts; counter-intuitively, the mass of Earth-made parts increases as the outpost becomes more self-sufficient, because of the fixed logistical mass constraint. Once again, this lifetime output value is used in the computational model to allocate the embodied energy of resource processing systems among all downstream consumers of processed resources:

$$O_{lfrp} = O_{lfm} \cdot t \cdot eff_m \tag{32.3}$$

Equation (32.3) above yields the lifetime¹⁸ output of the resource processing system, O_{lfrp} , in units of kg, where:

O_{lf} is the annual rated output capacity of the resource processing system in units of kg of processed resources produced per year,

t is the time horizon of the campaign in years, typically 20 or 40, and

eff_m is the dimensionless efficiency of conversion of inputs to outputs.

32.4.4.6 Fabrication Equations

In the model, it is assumed that the lifetime output of the fabrication system, O_{lff} in units of kg, is sized to the required inputs of the Manufacturing sector. This output

¹⁸ Lifetime = campaign length, in this case study both 20 and 40 years were modeled.

is proportional to the fabrication system's mass,¹⁹ its efficiency converting inputs to outputs, and a special productivity multiplier which corresponds to the versatility attribute described above. Versatility of a fabrication system is the capability to produce a larger variety of components that will be required by outpost systems.

$$O_{lff} = m_{system} \cdot (m_{output} / m_{system}) \cdot t \cdot eff_{mf} \cdot prod \quad (32.4)$$

Equation (32.4) is specific to the example of fabrication of simple components for later assembly into subsystems and systems, where:

m_{system} is the mass of the fabrication system (including the average mass of required spare parts per annum) in units of kg,

m_{output} is the rated output of needed²⁰ components by the fabrication system, in units of kg per year,

t is the time horizon of the campaign in units of years, typically 20 or 40,

eff_{mf} is the dimensionless efficiency of conversion of inputs to outputs in manufacturing on Mars, and

$prod$ is a dimensionless average productivity/versatility multiplier which depends on technological choices made, as discussed above.

In this simple model, this production function for fabricated components implicitly assumes that all other required inputs (i.e., crew labor hours, energy, processed resources) are available as and when required. A higher-fidelity model by Lordos et al. (2020) simulates shortages or delays in these required inputs.

32.4.4.7 Embodied Energy of Labor

To calculate the embodied energy contributed to various economic subsectors by crew labor, we require the sum of all the direct and indirect embodied energy required for crew life support. This budget would include habitat, environmental control and life-support systems, food production, etc. Then, the lifetime embodied energy of the Habitat and Life Support sector can be allocated among downstream labor-consuming subsectors in proportion to the fraction of the total lifetime labor hours used up by each subsector. By way of example, the equation to calculate the embodied energy of food consumed by the crew is shown:

$$E_{m, food} = C \cdot f \cdot t \cdot 365 \cdot e_f \cdot L \quad (32.5)$$

¹⁹ Including spare parts.

²⁰ It is important to note that this output is net of equipment idle time, i.e., if a campaign scenario underutilizes the fabrication equipment, then m_{output} is correspondingly reduced to only the amount needed.

Equation (32.5) above yields the part of the lifetime embodied energy of crew labor hours attributable to food production, $E_{m,food}$, in units of embodied MJ, where:

C is the average number of crew over the lifetime of the campaign,

f is the food consumption per crew member per day, in units of kg per day,

t is the time horizon of the campaign in years, typically 20 or 40,

e_f is the energy requirement for food production per person per day, in units of MJ per kg of food, and

L is the dimensionless fraction of total lifetime crew labor hours allocated to production functions, with the rest of the available crew labor hours being allocated to all other purpose(s) for which the crew is on Mars.

A limitation of the above equation is that in this simple model, we did not provide for indirect embodied energy inputs to food production because food production was not modeled as a standalone subsector.

A further complication that is especially prominent in the Labor sector is that its downstream labor-consuming sectors are also upstream of the Habitat and Life Support sector, i.e., there is circularity here. In this simple, static, one-period computational model the circularity was sidestepped by not attempting to allocate the lifetime embodied energy of labor to other sectors. Instead, the lifetime embodied energy of labor is summed with the embodied energy of other sectors to yield the total LEE of the campaign. This limits trades using this model to between complete architectural configurations, however a more elaborate model can track and account for the allocation of labor among subsectors and adjust their embodied energies accordingly.

32.4.4.8 Lifetime Embodied Energy of Mass Transport

Depending on the selected mission configuration, the mix of end-use systems vs. manufacturing systems in payload manifests will vary. The embodied energy of end-use systems due to the mass transport from Earth is directly proportional to their mass and to a pre-determined constant being an exchange rate between payload mass and embodied energy. It must be calculated for each sector so that it can be directly allocated to its embodied energy budget, for further downstream allocation(s) as needed:

$$E_m = (m_{lf} - m_{is}) \cdot e_t \quad (32.6)$$

Equation (32.6) above is calculated for each sector in units of embodied MJ to directly allocate the embodied energy E_m of finished systems to the sectors that require them, according to their payload mass:

m_{lf} is the lifetime mass of sector systems including spares required over the fixed campaign horizon (20 or 40 years), in units of kg,

m_{is} is the part of the lifetime mass of a sector's systems that is produced on Mars from in-situ resources, in units of kg, and

e_t is the exchange rate between logistical mass and embodied energy, described in Sect. 32.4.4.2 above, in units of MJ/kg. In this model, this exchange rate is 206 MJ/kg.

32.4.4.9 Lifetime Embodied Energy of the Energy Sector

Nuclear, solar and battery energy systems are modeled as being partly manufacturable on Mars, with the import only of complex sub-assemblies and the fabrication of bulky metal or other parts from local resources. For computational simplicity and ease of exposition, in this demonstration model we do not assume other embodied energy inputs into the Energy sector such as labor for maintenance or any consumables. Thus the equation for $E_{m,energy}$, the embodied energy of energy systems measured in units of embodied MJ, is as follows:

$$E_{m,energy} = (m_{lf,energy} - m_{is,energy}) \cdot e_t \quad (32.7)$$

$m_{lf,energy}$ is the lifetime mass of energy sector systems including spares required over the fixed campaign horizon (20 or 40 years), in units of kg,

$m_{is,energy}$ is the part of the lifetime mass of the energy sector's systems that is produced on Mars from in-situ resources, in units of kg, and

e_t is the exchange rate between logistical mass and embodied energy, described in Sect. 32.4.4.2 above, in units of MJ/kg. In this model, this exchange rate is fixed at 206 MJ/kg for all model runs.

32.4.4.10 Lifetime Embodied Energy of Direct and Indirect Energy Inputs for Value-Adding Subsectors in the Value Chain

Here, we use the fraction of the total energy budget directly consumed by each subsector as the cost driver to allocate the embodied energy of the energy sector, $E_{m,energy}$, among all the subsectors which directly consume energy produced on Mars. In addition, we take into account the direct material inputs to each subsector, which carry their own previously embodied energy. In the Eqs. (32.8–32.10), $E_{lf,erp}$, $E_{lf,mfg}$ and $E_{lf,hab}$ are the sums of the direct plus indirect embodied energies, originating from the space logistics primary source and from the Energy sector, in the Resource Processing, Manufacturing and Habitat sectors respectively:

$$E_{lf,erp} = (O_{lf,erp} \cdot e_{srp} / O_{lfe}) \cdot E_{m,energy} \quad (32.8)$$

$$E_{lf,mfg} = E_{lf,erp} + (O_{lf,mfg} \cdot e_{smfg} / O_{lfe}) \cdot E_{m,energy} \quad (32.9)$$

$$E_{lf,hab} = (C \cdot t \cdot 365 \cdot 24 \cdot e_{ecls} \cdot L / O_{lfe}) \cdot E_{m,energy} + (m_{lf,hab} - m_{is,hab}) \cdot e_t \quad (32.10)$$

where:

$E_{m,energy}$ is the lifetime embodied energy of the energy of space logistics in the energy sector, in units of embodied MJ of energy expended in space logistics,

E_{lfrp} , E_{lfrmfg} , E_{lfhab} is the direct and indirect energy embodied in the Resource Processing,²¹ Manufacturing and Habitat sectors respectively, in units of embodied MJ of energy expended by the energy subsector on Mars,

O_{lfe} , O_{lfrp} , O_{lfrmfg} are the physical outputs (in units of MJ or kg, as appropriate) of the Energy, Resource Processing and Fabrication sectors,

e_{srp} , e_{smfg} , e_{shab} are constants in terms of energy required in MJ per kg of output in the Resource Processing, Fabrication and Habitat²² sectors,

C is the average number of crew over the lifetime of the campaign,

t is the time horizon of the campaign in years, typically 20 or 40,

e_{ecls} is the energy requirement for environmental control and life support, in units of MJ per person-hour supported,

L is the dimensionless fraction of total lifetime crew labor hours allocated to production functions, with the rest of the available crew labor hours being allocated to all other purpose(s) for which the crew is on Mars,

$m_{lf,hab}$ is the lifetime mass of habitat subsector systems including spares required over the fixed campaign horizon (20 or 40 years), in units of kg

$m_{is,hab}$ is the part of the lifetime mass of the habitat subsector's systems that is produced on Mars from in-situ resources, in units of kg, and

e_t is the exchange rate between logistical mass and embodied energy, described in Sect. 32.4.4.2 above, in units of MJ/kg. In this model, this exchange rate is fixed at 206 MJ/kg for all model runs.

32.4.4.11 The Lifetime Embodied Energy of a Sector, System or Process

For each sector, its LEE, E_{lf} in units of embodied MJ, is calculated by Eq. (32.11):

$$E_{lf} = E_m + E_{is} \quad (32.11)$$

²¹ For ease of exposition, we make the computationally simplifying assumption that all output of the resource processing subsectors is used up by the fabrication subsectors. A more elaborate model would track inputs and outputs across multiple types of inputs and multiple subsectors at different parts of the value chain.

²² Note that the useful output of the habitat is "number of persons supported".

where:

E_m is the direct lifetime embodied energy of systems directly provided by space logistics, in units of embodied MJ, and

E_{is} is the indirect lifetime embodied energy of space logistics, embodied into the sector by upstream in-situ resource utilization. It may also include the embodied energy of space logistics for any complex Earth-made parts.

Taking care to avoid double counting, the sum of all value-adding, directly embodied energies is the lifetime embodied energy of the settlement, and therefore a proxy for the physical economic cost of the architectural configuration.

32.4.4.12 Visual Representation of Model Structure, Inputs and Outputs

The computational model described above, as implemented in the spreadsheet, provides the controls for the selection of the architectural configurations and displays the results on the same sheet, as shown in Fig. 32.14. The model as implemented is capable of yielding summary cost and benefit values for hundreds of architectural configurations which can then be analyzed in a tradespace of useful mass on Mars vs. lifetime embodied energy per unit of logistical mass. It is also capable of allocating embodied energy by subsector, providing insight into the relative energy intensity of different productive activities. In this simple study, however, and considering the length of the chapter, only the results of seven representative configurations are shown and discussed for the purpose of analysis.

32.4.5 Modeled Mission Configurations

Seven mission configurations for our modeled fledgling Mars industrial outpost, each with different levels of reliance on ISRU and ISM, were generated using the same consistent set of assumptions and parameters.²³ These configurations, numbered from 1 to 7, were carefully chosen to be sufficiently distinct, representative alternatives, while also cast in increasing order of utilization of Mars in-situ resources to reduce reliance on Earth in order to visualize the return on more investment in ISRU.

The main assumptions common across all configurations were crew size (8), campaign period (either 20 or 40 years), a fixed amount of mass delivered to Mars (1,600 tons for the entire campaign, delivered evenly with every launch window throughout the campaign length) and a moderate, but comparatively energy-rich, level of power availability (720 kW). The model calculates the final useable mass emplaced on Mars at the end of the campaign period, the fraction of useable mass that was made from in-situ resources and the specific lifetime embodied energy cost

²³ Please see Appendix II of Lordos's (2018) thesis for a listing of the assumptions and parameters.

of the total logistical mass. Useable mass excludes the overhead mass of the industrial ecosystem required to produce mass from in-situ natural resources, because this industrial ecosystem is the alternative to the classic long supply train. This exclusion of the industrial ecosystem overhead from the ‘useable mass’ metric makes ISRU-supported configurations directly comparable to no-ISRU classical configurations, where all systems required for survival and exploration are delivered using a classically optimized space logistics architecture which only minimizes the cost of space transportation, including future spare or replacement parts.

The production capacity²⁴ of the modeled industrial ecosystem is exogenously set as a configuration assumption, and its specific productivity (in terms of ‘mass of output, per unit time, per mass of production system’) depends mainly on the versatility of the equipment as prescribed in the definition of each given configuration. ‘Versatility’ in this context is a term which combines the attributes of high flexibility, variety of producible outputs and technological sophistication for a given packaged industrial ecosystem, resulting in high utilization and higher output rates as measured in useable (i.e., needed) mass. As used in the model, versatility abstracts away many of the modeling details in favor of a simple exposition of the LEE method; however, the range of versatility multipliers modeled, from 1 to 12, is in line with findings from past ISRU studies, such as lunar molten regolith electrolysis (Schreiner et al. 2015) or Martian solid oxide electrolysis (Hecht and Hoffman 2016). The total mass of the resource processing, fabrication and assembly systems, which drives the exogenously set production capacity, was deliberately *not* made endogenous for this simple model, so as to isolate and illuminate the first-order productivity differences implied by the various alternative industrial development scenarios, without relying on the second-order endogenous growth of production capacity. Growth trajectories of autocatalytic (self-reinforcing) systems are likely to be highly sensitive to initial assumptions and to future uncertain events and decisions. Thus, endogenizing growth would unnecessarily complicate this initial exposition of LEE as a cost and value metric that is well suited to the study of long-lived space infrastructure projects.

Seven different configurations were run, representing alternative strategies for the first multi-decade campaign of crewed missions to Mars in increasing order of investment in industrial capabilities tapping in-situ resources. These were set across the industrial development spectrum, ranging from no efforts to industrialize, up to high-technology strategies which include ISRU, ISM, recycling and substantial automation with highly versatile, small-footprint smart manufacturing platforms. The seven configurations are described in the following subsections.

²⁴ In the model, the ISRU and ISM systems are treated as being maintained using parts brought from Earth, parts fabricated on Mars from in-situ resources and labor available on Mars, such that total productive capacity remains constant at the exogenously set scenario rate throughout the campaign horizon.

32.4.5.1 Flags and Footprints (# 1)

The Flags and Footprints configuration is Apollo-like, meaning there is no ISM and no ISRU. As a result, the sole source for all requirements is space logistics, and therefore the specific embodied energy cost of this configuration stays very close to the ‘exchange rate’ between payload mass/IMLEO and embodied energy.²⁵ The purpose of this configuration is to provide an embodied energy baseline, against which we can assess the other six configurations.

32.4.5.2 Low-Tech (no Robots) (# 2)

The Low-tech configuration introduces basic, low-versatility ISRU and ISM capabilities together with recycling, with almost no automation in the assembly phase. The energy mix is predominantly solar backed up with batteries, with some nuclear.

32.4.5.3 Simple Mining (# 3)

In this configuration, the ISRU/ISM equipment is upgraded to medium versatility, which in practice translates to the mining of more classes of resources and the fabrication of more types of systems. As a result, the outpost can import larger quantities of fewer types of ready-made complex sub-assemblies from Earth which it can combine with the new resources mined and processed to produce a higher mass fraction of their needs from local resources, at a slight cost to their productivity. The energy mix is solar plus batteries, and there is no recycling of metals and plastics.

32.4.5.4 Medium-Tech (# 4)

The Medium-tech configuration has medium-versatility equipment, more robotic automation at the assembly stage, plus recycling of metals and plastics and an all-nuclear energy mix.

32.4.5.5 Robotic Tech (# 5)

This configuration is similar to Medium-tech, but has even higher levels of robotic automation at the assembly stage. This configuration recognizes that the time of humans will be at a premium.

²⁵ This exchange rate is analogous to the gear ratio concept. Given an interplanetary transfer orbit, a propulsion technology and an EDL technology, it is possible to calculate gear ratios of mass required in LEO to mass delivered to a planetary surface. Typical gear ratios for LEO to Mars are of the order of 7.5 to 11. An exchange rate of ~206 MJ / kg was calculated in App. I of Lordos’s (2018) thesis.

32.4.5.6 High-Tech (no Robots) (# 6)

High-tech is a control configuration with high versatility, meaning that the flexible manufacturing systems could manufacture and/or assemble a substantial fraction of sub-assemblies, however in this case this is simulated with zero automation at the final assembly stage, but with manual labor.²⁶ This selection forces the model to allocate all habitat-embodied energy costs to the production function, and embody all the costs of keeping humans alive into the energy cost of the outpost's economic output. The energy mix is all nuclear.

32.4.5.7 High-Tech (# 7)

This configuration has high versatility, high automation at the assembly stage, and a balanced energy mix.

32.4.6 Model Results

The model was first run to calculate outcomes for these seven configurations with a campaign length of 20 years. The campaign length was then doubled to 40 years, and the same seven configurations were run again, to quantify the impacts of longer time horizons on long-term, forward-looking lifetime embodied energy costs. The results of the model runs together with the input parameters are shown in Table 32.4.

Among other things, these results show that doubling the analysis horizon from 20 to 40 years, which is the same as saying that the decision maker's investment horizon is 40 years instead of 20, leads to significant reductions in specific lifetime embodied energy cost. It is also clear from Table 32.4 and Fig. 32.15 that specific LEE reductions resulting from longer decision-making horizons are greater for the more capital-intensive, higher-industrialization strategies (i.e., scenarios 4–7). This result is in accordance with an intuitive understanding of amortization for capital-intensive projects.

The key insight from these results is that reductions in specific lifetime embodied energy are correlated, as expected, with increased absorption of in-situ resources (i.e., with higher scenario numbers) and with longer design horizons. Hence, the LEE method can lead to validated ecosystem models which can be used to quantify the returns of up-front investment in industrial ecosystems to support sustainable, Earth-independent planetary settlements.

²⁶ Which means that the crew of eight would have to assemble systems at an average rate of ~58 kg per person per day for 20 years, accumulating a useable mass of 4,000 tons of infrastructure. Assuming they could keep up with such a rate, this hypothetical crew is unlikely to have much leisure time for exploration or science.

Table 32.4 Design vectors and model outputs of lifetime embodied energy simulation for the seven configurations in order of increasing efficiency of utilization of in-situ resources, for two campaign horizon durations per configuration

Scenario number	1	2	3	4	5	6	7
Scenario name	Flags & Footprints Baseline	Low tech (no robots)	Simple Mining	Medium tech	Robotic Tech	High tech (no robots)	High tech
Energy mix: Nuclear %	50%	20%	0%	100%	50%	100%	50%
Energy mix: Solar %	50%	80%	100%	0%	50%	0%	50%
ISM?	False	True	True	True	True	True	True
ISRU?	False	True	True	True	True	True	True
Use of robotic automation in assembly operations	5%	10%	20%	50%	80%	0%	80%
Use of manual labor in assembly operations	95%	90%	80%	50%	20%	100%	20%
Versatility of ISRU/ISM equipment (range, types of outputs that can be made)	Low	Low	Med	Med	Med	High	High
Mass Multiplier Index for ISRU/ISM equipment (LOW versatility = 1)	1.0	1.0	5.0	5.0	5.0	12.0	12.0

(continued)

Table 32.4 (continued)

Scenario number	1	2	3	4	5	6	7
Productivity Index for ISRU/ISM equipment (LOW versatility = 1)	1.0	1.0	0.9	0.9	0.9	0.8	0.8
Recycle scrap metal and plastic?	False	True	False	True	True	True	True
Final Useable Mass (kg)	1.36 E+06	1.91E+06	2.63E+06	2.90E+06	2.79E+06	4.06E+06	4.00E+06
% Useable Mass Produced In-Situ	0%	25%	52%	49%	52%	66%	67%
Specific Lifetime Embodied Energy of Useable Mass (20 years) MJ/kg	212.8	150.4	108.8	99.5	89.7	80.8	68.3
Specific Lifetime Embodied Energy of Useable Mass (40 years) MJ/kg	208.0	113.8	88.9	59.0	49.2	49.2	34.0

32.4.7 Model Limitations

The model presented here, though fully quantified and parameterized, has no dynamics and no time dimension. The time horizon under study (either 20 or 40 years) is treated as a single instantaneous period, and all equations are solved deterministically and instantaneously for the entire time horizon. This limitation is a benefit for the purpose here, which is to present the method of lifetime embodied energy in the context of an easy-to-understand calculation. Realistic growth dynamics were not considered: only the first-order capacity of each alternative industrial configuration to produce useable mass was calculated, along with the metrics of interest here,

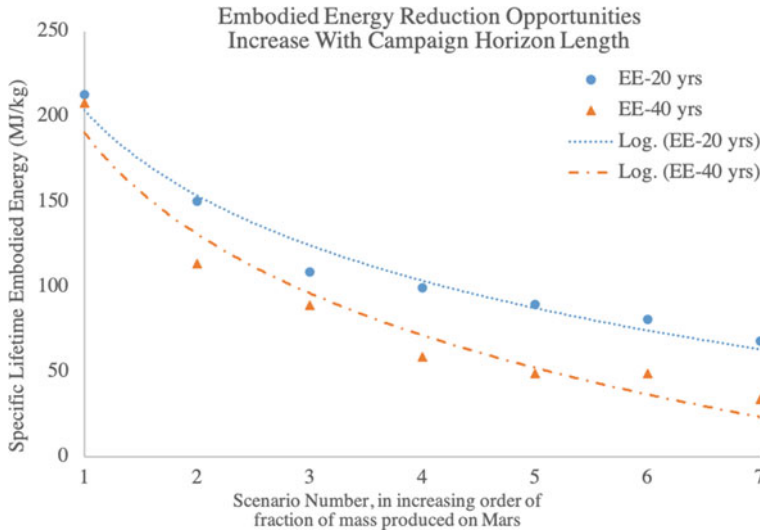


Fig. 32.15 Results of lifetime embodied energy modeling for seven architectural configurations in order of increasing efficiency of utilization of in-situ resources, for two campaign horizon durations per configuration

such as the lifetime embodied energy for each economic sector and the fraction of cumulative useful mass produced from in-situ resources.

Another set of modeling and methodological limitations revolves mainly around data-gathering difficulties. These affect data or projections on system lifetime, productivity, manufacturability, materials and labor requirements. While some of these assumptions have been validated with references to literature, many are estimates that remain to be validated and refined before the output of the model can be used in a realistic campaign design exercise. The impact of lack of data is multiplied because of the practice of truncation of hierarchical process trees: the high cost of obtaining more data motivates the pruning and truncation and the replacement of missing data with estimates and assumptions, which negatively impacts the accuracy of the resulting embodied energy calculation. However, if we assume that all assumptions are correct to rough order of magnitude, then the order-of-magnitude differences observed in specific lifetime embodied energy costs of some of the high industrialization scenarios vs. the no-ISRU scenario 1 are a significant finding.

Another perspective on the same limitation is that embodied energy calculations remain a time-consuming process and are less relevant to most people than mass or dollar costs. They are certainly not immediately and intuitively tangible. Mass-based approaches are very well established in the space industry. This gives rise to resistance to changing established methodologies.

Most of the above limitations are, in the final analysis, Earth specific. They have to do with the depth and breadth of our supply chains, and centuries of economic history and inertia. Some of the above limitations can be mitigated by integrated

project planning of LEE applications (see examples in Sect. 32.5). This is especially true in our specific ‘space settlement’ context of clean-sheet industrial ecosystem and planetary settlement designs, where supply chains are open to de novo design along with everything else. Hence, lifetime embodied energy models and the lifetime embodied energy approach can usefully inform mission campaign architecture and the design of sustainable industrial ecosystems and settlements on other worlds.

32.5 Lifetime Embodied Energy Applications

As we saw in Sect. 32.1, the mass-as-cost-proxy approach was quite successful in past decades when space systems were one-off, expendable and/or too expensive, when every mission was unique, and when no habitation infrastructure was being accumulated on Mars or the Moon. However, now, at the dawn of the reusability, ISRU and ISM eras of human spaceflight, it is proposed that an energy-based metric will in future turn out to be more appropriate to architects of planetary infrastructure, industrial ecosystems and permanent planetary settlements.

32.5.1 *Design of Planetary Industrial Ecosystems and Supply Chains*

Intuitively, when contemplating the results of the lifetime embodied energy model of a Mars outpost, the modeling of energy flows through industrial ecosystems on other worlds is likely to be a useful exercise. The architect of an extended human exploration spaceflight campaign may be interested in, among other things, promoting Earth independence by reducing the flows of logistical mass, and in increasing the cumulative mass of useful infrastructure emplaced at the destination world. The lifetime embodied energy paradigm proposed here eschews logistical mass and dollar costs in favor of an energy-based common-denominator metric that can be used to value all future investments and activities on the destination world, given the requirement of a permanent presence there. In this way, the designer has at their disposal a quantified common denominator of cost that they can use to ask and answer questions such as whether they should be transporting less *mass* and more *capabilities* through and across gravity wells, and how they could be putting such capabilities together with the mass available at the destination to good use to solve the problem of emplacing the infrastructure needed for human activities.

Thus, the hypothesis is as follows: that the lifetime embodied energy-based metric can reflect everything classical mass-based metrics can do, and it can also do more that mass-based costing systems could not do. As we saw from the Excel model analysis, direct comparisons with mass-based costing systems are possible by selecting the embodied energy of space logistics as the source of energy to the LEE model. In this

model, we converted payload mass cost to embodied energy of space logistics using an appropriate ‘energy per unit of landed payload mass’ exchange rate.²⁷

More interestingly, LEE goes beyond replicating what mass-based cost models can do. As we found from our analysis of predecessor nodes in the energy diagram model (Figs. 32.8, 32.9, 32.10, 32.11, 32.12 and 32.13), the same energy diagrams used to set up the embodied energy calculation also reveal linkages and dependencies between nodes, clearly illustrating the role every system or process plays in the effort to keep crews alive and productive. This side benefit was not accidental or fortuitous. It emerged precisely because energy is the natural metric to measure any and all efforts to sustain systems in the struggle against the relentless degradation of all natural and artificial systems which is imposed by the second law of thermodynamics. Maintaining high order requires available high-quality energy, and maximizing the flows of energy in self-reinforcing ways is what all dominant forms of life—and dominant technical architectures—have evolved to do under competitive pressure.

Therefore, when we model networks of energy flows in order to measure the embodied energy costs at various nodes, we are also automatically recreating the pathways which keep the natural or artificial system not only functional, but also moving forward and ever closer to an idealized optimality. Going forward, this alignment between nature’s ways and the way we represent our system designs in terms of embodied energy flows is likely to lead to more insights and discoveries in the field of optimizing the performance of planetary industrial ecosystems, which is of interest to system architects, systems engineers, government agencies and private industry eyeing up opportunities for the long-term industrial development of, especially, the Moon.

Moreover, in modeling lifetime embodied energy (with energy expended in space logistics as the primary source), we are actually working with a metric which simultaneously targets not just reduced lifetime logistical mass, but also increased lifetime energy efficiency of the systems under consideration. Given that the industrial revolution only really began in earnest when James Watt started raising the energy efficiency of the steam engine, the space systems architect who would like to design self-sustaining cities on other worlds should be interested in metrics which can be used to optimize both logistical cost and planet-side lifetime energy efficiency in one stroke. The importance of keeping an eye on in-space energy efficiency cannot be overstated: unlike here on our home planet, where we co-evolved with and within our habitat ecosystem, all human activity in space will have to be paid for out of our own expensive energy sources that we will bring or make there. Far from the cradle of our birth planet, we will also be far from its free life-support system which sustains the lives of billions of us. In the economic sense, therefore, the free-rider problem should be diminished on Mars, because we will enjoy far fewer free services from the local environment than we do on Earth.

²⁷ This exchange rate will turn out to be unique for different transportation systems, different reusability strategies and even different launch dates in a launch window. Effectively, the optimal exchange rate is the optimized outcome of mass-based space logistics costing systems, given a desired landed payload mass.

Going one step further, the optimization opportunities offered by LEE-based frameworks go beyond mass and energy in their narrow, physical sense. The mechanics of measuring the cost of past effort in the way we described for embodied energy are generalizable to any type of high-level effort which is itself the result of compounded, indirect consumption of energy in lower-level required processes. Thus, we saw that the embodied energy cost of labor can be incorporated into the modeled flows of embodied energy simply by consolidating the LEE of all systems that the crew depend upon, such as Habitat and ECLS, and labeling it 'LEE of labor'. In so incorporating it, labor becomes a consequential input variable to the in-space production function. This opens up the possibility for the designer to assign an objective cost to labor using LEE, which opens the way to designing industrial architectures that vary the amount of high-cost, high-value labor input along with inputs of resources, capital and energy of different kinds, and observing the effect on predicted performance outcomes and on the predicted lifetime embodied energy total cost. If the principle of diminishing marginal utility applies, as it probably will, non-dominated designs will be consistent with equal marginal returns of all these inputs. In other words, in the context of a detailed, realistic model, non-dominated, Pareto-optimal designs will reveal not just the minimum amount of embodied energy, but also how the mix of capital, labor and materials, of different kinds and in different proportions, results in viable tradeoffs of LEE investment for additional system performance or system margin.

The systemic viewpoint afforded by the lifetime perspective of LEE and the associated easy-to-use energy-language diagrams emphasize another dimension that the architect should pay attention to. Namely, that past 'make vs. take' decisions will alter the trajectory of future decision options, and can create both negative and positive lock-in. Thus, if the first outposts on Mars are constructed using life-support systems which are very difficult to manufacture on Mars, these outposts will end up being dependent on a constant flow of very specific spare parts, and future mission architects will face both overt and subtle switching costs which will discourage them from adopting radical new designs. Such constraints, or 'limits to growth' will likely be everywhere to be found in our system designs for the Moon and Mars. It will be the responsibility of the architect to target a self-consistent set of sustainable growth rates for any variables threatening to become constraints to the future growth of the outpost into a city. The ability to model all interactions and measure the impact of the up-front decisions on a single objective cost metric will assist in that task of balancing the future growth potential of important internal model variables. Again, the natural energy linkages between nodes on the energy-language diagrams in Figs. 32.8, 32.9, 32.10, 32.11, 32.12, and 32.13 help to tell the story of the interacting growth trajectories of the main state variables.

32.5.2 Design of Sustainable Space Settlements

As we have seen, the LEE metric and the associated energy-language diagrams can inform insightful architecture decisions leading to the establishment of efficient and productive planetary industrial ecosystems. In turn, these industrial ecosystems are the essential foundations that provide logistical support and resilience to human settlements and other long-lived planetary infrastructure. However, lifetime embodied energy methods can also inform the design and concept of operations for permanent space settlements.

Currently, life-cycle embodied energy is in widespread use only in the measurement of the energy performance of terrestrial buildings. This measurement of performance has been mandated by law in the European Union and elsewhere. However, embodied energy calculations are cumbersome because of the depth and breadth of Earth's supply chains, and because energy use is effectively subsidized by nature, so very few people build houses designed to minimize lifetime embodied energy. Instead, with few exceptions, the embodied energy is generally calculated as a secondary information after a design has been optimized along other dimensions, and in many cases after a building has been constructed. Certainly, to the authors' knowledge, no one has designed or built an entire city which is intended to minimize lifetime or life-cycle embodied energy.

However, on the Moon or Mars, we'll be building new settlements that rely on new supply chains, with all of that starting from a blank sheet. Moreover, all energy consumption in space will come at an opportunity cost to be traded off with other activities, with fewer unpriced positive or negative externalities when compared to our historical experience on Earth (see: fossil fuels, water quality, climate change). Therefore, taking both of these differences into account, there is both the opportunity and the motivation to optimize future space settlements according to their lifetime embodied energy.

32.5.2.1 Embodied Energy Metadata Can Become a Standard Part of Digital Twins Developed as Part of the Design of Space Settlements

Digital twins of entire cities would be a logistically daunting proposition on Earth, but for de novo developments such as smart cities (Austin et al. 2020), or space settlements, embodied energy could be tracked for all elements by incorporating embodied energy metadata in the digital twin model. As the in-space supply chain has yet to be created, the opportunity arises to structure this supply chain in such a way as to make it easy, if not automatic, to calculate embodied energies for all manner of in-space systems, products and processes, at all levels of aggregation or decomposition. It is a matter of tagging every material system or subsystem intended for use in space with the required metadata, and creating a universal framework where every in-space system has a digital twin, from where all the required metadata

would then be accessible. This recommendation is well aligned with current trends towards model-based systems engineering (MBSE).

This metadata would also be useful not only to the space architect, but also to the logistics team of a mining or in-space fabrication operation on the Moon or Mars. Again, the Earth makes up many of our losses to entropy with free gifts of energy; this won't be the case in space. So in future there will be operational reasons for accurately tracking the energy cost of complex development activities, such as in the month-to-month planning to balance the energy budget of a fast-growing settlement on Mars. Such uses are supportive of the development and growth of human activities in space and go well beyond the architect's initial interest in mass, dollars or energy numbers for the purpose of designing a space system.

32.5.3 Valuation of Space Infrastructure and Space Economic Output

The objective function in space mission architecting almost always includes the minimization of total mass, whether for a single mission or across a campaign of missions. However, in a future characterized by reusability, ISRU and planetary industrial development, minimizing mass will not guarantee the minimization of cost. This gives rise to a need for improved proxies for estimated costs of long-lived space architectures. Since energy measures the capacity to do work, embodied energy—a concept which Howard Odum described as “the memory” of past work (Odum 1983)—is a measure of all the past work that was required for the creation or sustainment of a product or system. Given the physical basis of work in thermodynamics, the embodied energy of objects or systems may come closer than other metrics to being an ideal objective measure of their value or cost.

32.5.3.1 Embodied Energy is a Measure of Work, and Therefore of Cost

It is often repeated in the aerospace engineering profession that “mass attracts cost”. While this is generally true, we must consider whether the observed correlation between mass and cost is better understood as being due to a *third* causal factor that affects both mass and cost in the same way. Cost in fact attaches to *work*, not to *mass*, and in physics concentrated energy is the metric of the capacity to do useful work. All other things being equal, more mass requires more work, which requires more energy and therefore more cost. However, more or less mass does not always mean more or less cost, or more or less energy. Thus, energy is the natural metric for cost, because it is the natural metric for work.

32.5.3.2 LEE is a Native-Language Value and Cost Metric for Designing Multi-Decade Campaigns

The embodied energy of systems delivered to a destination in space can easily be converted to IMLEO using well-known gear ratios and a consistent, appropriate lifetime embodied energy (LEE) exchange rate, such as the exchange rate for Mars calculated using a method similar to that shown in Lordos (2018, Appendix I). Therefore, all IMLEO-based calculations and methodologies can be replicated with LEE. Beyond this, LEE does not come with the implicit assumption that payload mass will always be something that must be minimized: it is the lifetime embodied energy which must be minimized, *not necessarily* the mass. This makes LEE a native-language value and cost metric for designing multi-decade campaigns which will rely on ISRU and ISM, where the embodied energy of space logistics, delivered industrial ecosystems and in-situ raw mass combine to produce in-situ useful mass. This locally produced useful mass in turn displaces future payload mass from manifests. The embodied energy metric naturally keeps track of such displacements of future payloads as well as of the relative efficiency and productivity of alternative ISRU and manufacturing methods at the destination.

32.5.3.3 LEE Can Be Tracked in Dynamic Models to Create Models with a Time Dimension

Space economists interested in modeling the future growth of the space economy may use LEE-based models, such as systems dynamics models, to study the growth dynamics of the space economy and related potential constraints. In addition to the lifetime embodied energy cost per system, the figures of merit for different subsets of the overall settlement system can include: life-support system capacity per capita, pressurized volume per capita, stock of consumables per capita, mass of production output per unit of labor hour, energy consumed per unit mass of production output in the production sector, energy consumed per person per time period by habitat and life-support systems. Every stock variable in these system representations—habitat volume, industrial equipment, persons burdening the life-support system—can be tagged with its evolving LEE cost over time (Lordos et al. 2020). In this way, LEE provides a uniform common-denominator cost variable supporting trades and multi-objective optimization between all of the above functional intents and their figures of merit.

32.5.3.4 Standardizing Sector Breakdown of Space Economy to Aid in Mapping LEE Flows

In support of the objective to include detailed lifetime embodied energy metadata in digital twins of space infrastructure systems, it will be useful to agree early on a comprehensive, standard categorization and mapping of sectors of economic activity

for the space economy, so that all participants gather and record the data used in LEE calculations in the same way. Such an initiative will greatly facilitate later analysis of space economic activities using LEE.

32.5.3.5 Eventual Emergence of an Energy-Based Value System for the Space Economy

From the above-mentioned foundations throughout this Sect. 32.5.3, if LEE does find widespread application in tracking the cost, value and economic output at both micro and macro levels, it is possible to envisage that LEE can also form the foundation of a more widespread value system, including being a basis for valuing assets or capabilities.

More near term, LEE analysis can inform key NASA and commercial space company TRL development decisions. Organizations may use LEE to quantify the benefits of important technology investment decisions regarding TRL development for ISRU, ISM and life-support systems. Specifically, an LEE analysis could assist NASA to determine whether and in which cases it is worthwhile to pay the high switching cost up front to walk away from legacy flight-heritage systems which may not be easily manufacturable at the destination and to invest in modified designs that meet extensive 'design for manufacturability' requirements.

Longer term, LEE analysis can objectively inform the valuation of existing in-space infrastructure for the purposes of securitization, construction, sale or purchase, in the same way that options pricing theory crossed the boundary from academia to applied finance, informing the valuation and pricing of contracts that are traded daily in markets worth trillions of dollars.

32.6 Summary and Conclusions

From the Apollo era to the present day, various mass-based metrics such as payload mass, system dry mass and IMLEO have served as proxies or key inputs into calculations of total mission cost. However, IMLEO is being progressively decoupled from true economic cost due to changes including the advent of reusable rockets, the inclusion of ISRU in design reference missions and the new objectives for permanent Moon and Mars settlements or outposts where infrastructure will be accumulated. At the same time, cost reductions arising from new forms of contracting and new investments are starting to undermine the validity of old cost databases and CERs, leading to space system architects requiring a new metric for cost that is not denominated in units of mass or money.

Since cost is directly driven by thermodynamic work, a natural and objective metric of past work is energy, and specifically the embodied energy, which is the sum of past work that went into the creation or maintenance of a system and its predecessors. When adding the projected future energy expenditures for operation,

maintenance and decommissioning, we arrive at lifetime embodied energy. This concept was first created in the 1970s amidst the oil crises, and has found limited application, mainly in the study of the lifetime energy performance of buildings. Its use in a space application is novel.

Flows of embodied energy were represented using Odum's energy-language diagrams, and also with equations in a simple Excel model. The Excel model of the initial Mars outpost was run under seven scenarios, representing increasing investment in a variety of modeled industrial capabilities, for campaign durations of 20 and 40 years. The LEE method captures and quantifies long-term ISRU benefits that IMLEO will miss, because in the model, all seven scenarios start with exactly the same IMLEO but vary widely in lifetime embodied energy, up to by an order of magnitude. The resulting distribution of specific lifetime embodied energy is aligned with intuitive expectations about the high return on investment in ISRU and manufacturing capabilities at planetary destinations. The key enabling methodological advance is the ability to measure all types of short- and long-term economic costs, including labor costs, on the same objective denominator of embodied joules of energy from the selected primary energy source, which in the case of the model presented here is the energy of space logistics.

This common denominator of LEE makes it possible to compare all types of diverse human spaceflight architectures, with or without investment in ISRU and ISM, without unfairly disadvantaging any one family of concepts. This graphical and quantitative analysis demonstrated the usefulness of the proposed new metric in the architecting of human space exploration campaigns, the design and development of long-lived planetary industrial ecosystems and permanent planetary settlements such as the award-winning 'Star City', Mars (Lordos and Lordos 2019), as well as other types of long-lived space infrastructure campaigns, such as the newly announced CLD program for commercial LEO destinations with seed funding by NASA, as envisaged in Lordos et al. (2019).

Odum's energy diagrams, as applied to the design of industrial ecosystems, space settlements and long-lived space infrastructure, represent the natural energy pathways which create and sustain both life and technological systems under competitive pressure. In this context, LEE-based models can support the study of more detailed architectural trades, such as subsystem-level energy efficiency studies as well as alternative mixes of energy, capital, labor, Earth imports and in-situ natural resources for the in-space production function.

Finally, the embodied energy paradigm, with its close, natural and physical connection to the creation of value in highly ordered systems, can also serve as a new cost and value system for the emerging commercial space economy. Such an endeavor will require effort and multilateral coordination, not least to ensure the tracking of LEE-related data in the design and operational models of these future space infrastructures. However, as we have not yet built any permanent lunar or planetary infrastructure, the opportunity for space systems engineers to build in convenient ways to track and build up costs in terms of embodied energy is still available for the taking.

Ad Astra, per Industria.

References

- Anderson, M.S., M.K. Ewert, and J.F. Keener. 2018. *Life support baseline values and assumptions document*. Houston, TX. CTSD-ADV-484 A.
- Austin, M., et al. 2020. Architecting smart city digital twins: combined semantic model and machine learning approach. *Journal of Management in Engineering* 36 (4): 04020026. [https://doi.org/10.1061/\(asce\)jme.1943-5479.0000774](https://doi.org/10.1061/(asce)jme.1943-5479.0000774).
- Ayres, R.U., and B. Warr. 2005. Accounting for growth: the role of physical work. *Structural Change and Economic Dynamics* 16(2 SPEC. ISS.):181–209. <https://doi.org/10.1016/j.strueco.2003.10.003>.
- Chen, H., and K. Ho. 2018. Integrated space logistics mission planning and spacecraft design with mixed-integer nonlinear programming. *Journal of Spacecraft and Rockets* 55 (2): 365–381. <https://doi.org/10.2514/1.A33905>.
- Cleveland, C.J., et al. 1984. Energy and the U. S. economy: a biophysical perspective. *Science* 225 (4665): 890–897. <https://doi.org/10.1126/science.225.4665.890>.
- Costanza, R., and R.A. Herendeen. 1984. Embodied energy and economic value in the United States economy: 1963, 1967 and 1972. *Resources and Energy* 6 (2): 129–163. [https://doi.org/10.1016/0165-0572\(84\)90014-8](https://doi.org/10.1016/0165-0572(84)90014-8).
- Daly, H.E. 1997. Georgescu-Roegen versus Solow/Stiglitz. *Ecological Economics* 22 (3): 261–266. [https://doi.org/10.1016/S0921-8009\(97\)00081-5](https://doi.org/10.1016/S0921-8009(97)00081-5).
- Dixit, M.K., J.L. Fernández-Solís, S. Lavy, and C.H. Culp. 2010. Identification of parameters for embodied energy measurement: a literature review. *Energy and buildings* 42(8): 1238–1247.
- Guerra, L., and R. Shishko. 2000. Estimating the cost of crewed space systems. In *Human spaceflight: mission analysis and design*, eds. Larson, W.K., and L.K. Pranke. New York: McGraw Hill.
- Hammond, G.P., and C.I. Jones. 2008. Embodied energy and carbon in construction materials. *Proceedings of the Institution of Civil Engineers-Energy* 161 (2): 87–98. <https://doi.org/10.1680/ener.2008.161.2.87>.
- Hecht, M.H., and J.A. Hoffman. 2016. The Mars oxygen ISRU experiment (MOXIE) on the Mars 2020 rover. In *3rd international workshop for planetary missions*, 2.
- Ho, K., et al. 2016. ‘Campaign-level dynamic network modelling for spaceflight logistics for the flexible path concept. *Acta Astronautica* 123: 51–61. <https://doi.org/10.1016/j.actaastro.2016.03.006>.
- Ishimatsu, T., et al. 2015. Generalized multicommodity network flow model for the Earth–Moon–Mars logistics system. *Journal of Spacecraft and Rockets* 53 (1): 25–38. <https://doi.org/10.2514/1.A33235>.
- Jones, H.W. 2003. Equivalent mass versus life cycle cost for life support technology selection. In *33rd ICES*. Vancouver: SAE International, 18. <https://ntrs.nasa.gov/search.jsp?R=20040015099>.
- Jones, H.W. 2004. Equivalent mass (EM), life cycle mass (LCM), and mass (M) metrics compared in advanced life support (ALS) analysis. In *International conference on environmental systems*. Colorado Springs, CO: SAE International.
- Jones, H.W. 2015. Estimating the life cycle cost of space systems. In *45th international conference on environmental systems*. Bellevue, WA: ICES-2015-041, 1–13. <https://ntrs.nasa.gov/search.jsp?R=20160001190>.
- Levri, J.A., D.A. Vaccari, and A.E. Drysdale. 2000. Theory and application of the equivalent system mass metric. In *30th international conference on environmental systems*, 12. Warrendale, PA: SAE International. <https://doi.org/10.4271/2000-01-2395>.
- Lordos, G., et al. 2020. Tradespace exploration of space settlement architectures using long-term cost and benefit metrics. In *IEEE aerospace conference proceedings*, 1–15. <https://doi.org/10.1109/AERO47225.2020.9172320>.
- Lordos, G.C. 2018. *Towards the sustainable industrial development of Mars: comparing novel ISRU/ISM architectures using lifetime embodied energy*. Massachusetts Institute of Technology.

- Lordos, G.C. et al. 2019. Investment decision model for a commercially owned and operated space station in low earth orbit. In *Proceedings of the international astronautical congress, IAC*, 2019-October(October), 21–25.
- Lordos, G., and A. Lordos. 2019. Star city, Mars. In *Proceedings of the 22nd international Mars society convention*. Los Angeles, CA.
- Odum, H.T. 1983. *Systems ecology*. Wiley.
- Odum, H.T. 1996. *Environmental accounting: energy and environmental decision making*. New York: Wiley.
- Ross, A.M. 2003. *Multi-attribute tradespace exploration with concurrent design as a value-centric framework for space system architecture and design*. Massachusetts Institute of Technology.
- Schreiner, S.S. et al. 2015. Development of a molten regolith electrolysis reactor model for lunar in-situ resource utilization, 1–21. <https://doi.org/10.2514/6.2015-1180>.
- Soddy, F. 1933. *Wealth, virtual wealth and debt: the solution of the economic paradox*. United Kingdom: E.P. Dutton.
- van Zon, A., and I.H. Yetkiner. 2003. An endogenous growth model with embodied energy-saving technical change. *Resource and Energy Economics* 25(6571): 81–103. <http://mpira.ub.uni-muenchen.de/6571/>.

George C. Lordos is a PhD candidate at Massachusetts Institute of Technology in the Department of Aeronautics and Astronautics, the founder of MIT's Space Resources Workshop and a member of AIAA's Space Resources Technical Committee. He received a B.A. in Philosophy, Politics and Economics from the University of Oxford in 1991, a MBA from MIT's Sloan School of Management in 2000 and a Masters in Engineering and Management from MIT's System Design and Management Program in 2018. George is researching the industrial ecology of human settlements on the Moon and Mars and previously had a 25-year professional career as a technical project manager, strategy consultant, system architect, company director and entrepreneur.

Jeffrey A. Hoffman is a professor in MIT's Aeronautics and Astronautics Department. He received a B.A. in Astronomy (summa cum laude) from Amherst College (1966); a PhD in Astrophysics from Harvard University (1971); and an MSc in Materials Science from Rice University (1988). As a NASA astronaut (1978–1997) Dr. Hoffman made five space flights, becoming the first astronaut to log 1000 hours of flight time aboard the Space Shuttle. He was a member of the spacewalking team that repaired the optics of the Hubble Space Telescope. His primary research interests are in improving the technology of space suits and designing innovative space systems for human and robotic space exploration. He is Deputy PI on the Mars 2020 rover MOXIE experiment, which for the first time will produce oxygen on Mars using local resources. Hoffman is director of the Massachusetts Space Grant Consortium. In 2007, Dr. Hoffman was elected to the US Astronaut Hall of Fame.

Olivier L. de Weck is Professor of Aeronautics and Astronautics and Engineering Systems at MIT where he teaches Technology Roadmapping, Satellite Engineering and Systems Engineering as well as Multidisciplinary Design Optimization. He has authored over 400 publications (12 best paper awards since 2004) and is a Fellow of INCOSE, Fellow of AIAA, Senior Member of IEEE and the Editor in Chief of the Journal of Spacecraft & Rockets. He is a former Senior Vice President of Technology Planning and Roadmapping at Airbus where he was responsible for roadmapping a \$1 billion R&D portfolio for the world's largest aircraft manufacturer. His passion is to improve life on our home planet Earth through research and education while paving the way for humanity's future off-world settlements.

Chapter 33

Policy, Legal Processes and Precedents for Space Mining



Scot W. Anderson, Korey J. Christensen, Julia La Manna, Katherine Wood, Alex Gilbert, and Morgan Bazilian

Abstract When humankind began to reach into outer space, it thought of space like the ocean. The high seas are beyond the law of any nation, and so are governed by international law. Many nations (though not the United States) entered into an international treaty, the United Nations Convention on the Law of the Sea, to set out some rules for actions on the high seas. Outer space is similarly beyond the jurisdiction of any nation, and so the space-faring nations looked at international law to regulate space activities. The Law of the Sea is analogous to the law of outer space, but it is not identical. The foundational document for the law of outer space is The Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, known as the Outer Space Treaty. That treaty, subsequent treaties, enactments and precedents, and the fundamental principles of international law, will guide us as we consider the legal framework that will best allow the development of resources in outer space.

S. W. Anderson (✉) · J. La Manna
Hogan Lovells US LLP, 1601 Wewatta Street, Suite 900, Denver, CO 80202, USA
e-mail: scot.anderson@hoganlovells.com

J. La Manna
e-mail: julia.lamanna@hoganlovells.com

K. J. Christensen
4601 DTC Blvd., Suite 550, Denver, CO 80237, USA

K. Wood
Hogan Lovells International LLP, The Colmore Building, 20 Colmore Circus, Birmingham B4 6AT, UK
e-mail: katherine.wood@hoganlovells.com

A. Gilbert
Fellow, Payne Institute for Public Policy, Colorado School of Mines, 425 Elm St NW, Washington, DC 20001, USA
e-mail: alex.gilbert@powerandresources.com

M. Bazilian
Colorado School of Mines, 816 15Th St., Golden, CO 80401, USA
e-mail: mbazilian@mines.edu

33.1 Policy Introduction

The space age began with the launching of Sputnik by the Soviet Union in 1957, the first artificial satellite to orbit the planet. Occurring only a decade into the Cold War, Sputnik posed both a military and “soft power” threat to the United States. The United States responded, and the space race began as the Soviet Union and United States raced to reach new technological milestones and achieve prestige on the international stage. When initial negotiations for what became the international space treaties began, there were only two major space powers: the United States and the Soviet Union, and no commercial space industry. The foundational treaty, “The Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies,” was signed in 1967 and became commonly known as the Outer Space Treaty. The Outer Space Treaty is the constitution of international space law; by defining outer space as an area beyond national jurisdiction, the Outer Space Treaty means that space law IS international law (United Nations Treaty Series 1967, Art. III). Under the Outer Space Treaty, every country that launched a space object would retain jurisdiction over, and responsibility for, that space object as its launching state (von der Dunk 2001, pp. 3–28). Similar to flagging in maritime law, this means that all human space objects are ultimately tied to a nation on Earth. Section 33.2.1 describes these portions of the Treaty in more detail.

Although the space age began in a post-World War II environment that facilitated the creation of the modern international system, the driving motivation behind space technologies and treaties was the intense geopolitical competition of the Cold War (Quinn 2008). With the United States and Soviet Union engaging in a space race amidst an increase in nuclear armaments, the genesis of the Outer Space Treaty was as much about arms control and preventing a territorial rush in outer space. The Outer Space Treaty links the governance of outer space to existing international law and the nascent United Nations system. As a result, it is a foundational treaty for nuclear weapons arms control. Worried by the destabilizing nature of nuclear weapons based in outer space, the Soviet Union and the United States pushed for the Treaty to ban the stationing of nuclear weapons in orbit or on any celestial bodies. Further, the Treaty was intended to prevent a “land rush” between the Soviet Union and the United States as both nations eyed missions to the Moon and even other planets. The specter of a new age of exploration, dominated by two opposing nuclear powers claiming celestial bodies, was, for a time, a grave concern.

Accordingly, the Outer Space Treaty has a non-appropriation provision that prevents national claims of sovereignty over celestial bodies (United Nations Treaty Series 1967, Art. II). The limitations in managing space resources seen today result directly from the need in the 1960s to develop a treaty to prevent a nuclear arms race in outer space. While the Outer Space Treaty was ratified before humans reached the lunar surface, the subsequent crewed Apollo missions and uncrewed Lunokhod missions extracted and returned lunar rocks to the United States and Soviet Union respectively. Some argue that this government use of resources gathered in outer

space establishes state practice for the use of space resources for scientific purposes under the Outer Space Treaty. In modern times, there is no serious international debate that scientific samples are owned by the government that extracts them, as evidenced by relatively little controversy surrounding Japanese, US, and Chinese missions to extract samples from asteroids, Mars, and the Moon, respectively. Nevertheless, the applicability for commercial use remains ambiguous (see Sect. [Application of the Outer Space Treaty to Resource Development](#)).

Following the Outer Space Treaty, four major space treaties were negotiated with three broadly signed and recognized. Together with the Outer Space Treaty, these treaties provide the basis for national activities in outer space. The fourth treaty, the Moon Treaty (United Nations Treaty Series 1979), represented an attempt to provide greater international control over outer space. In particular, it included a scheme to govern commercial space resources with an international body similar to the one created by the United Nations Convention on the Law of the Sea (UNCLOS) for the deep seabed in international waters. However, concerns over this system and other issues with the Moon Treaty led to only a handful of countries signing it. As discussed in Sect. [33.2.1.2](#), the applicability of the Moon Treaty to space resources today is limited. Nevertheless, it contains provisions about due regard and international benefit sharing that many countries will look to as space resource activities begin (see Sect. [33.3.1](#)).

Collectively, the space treaties establish outer space as something like a global commons within the international system, with private activities to be overseen by nation states. The use of the term “global commons” is contentious within some United States circles, due to concerns that it implies acceptance of the Moon Treaty or an UNCLOS-type international regulation of resources. However, the term is neither defined in customary international law nor in any of the space treaties. Generally, global commons are recognized simply as areas beyond national jurisdiction: high seas, Antarctica, the atmosphere, and outer space. Each of these areas have international agreements that govern national activities in them, ranging from controls on emissions for the atmosphere to a ban on resource utilization in Antarctica to a mixed resource regime for the high seas. The precise nature and implications of outer space as a global commons remains contentious, and state practice in coming years will refine its nature under both statutory and customary international law.

Due to their scientific, commercial, and military value, space resources are an emerging area of geopolitical competition. The United States, Russia, China, India, Japan, Europe, and many other countries are now investigating technologies and policies to unlock space resources production. In the near term, lunar resources can enable greater lunar exploration and even support crewed missions to Mars. In the longer term, space resources production can generate large economic benefits for Earth by supplying energy (through lunar Helium-3 for future fusion reactors or for materials to build space-based solar power) (Schmitt 2006), supporting greater space commerce, producing rare earths for consumption on Earth, and by enabling in-space manufacturing of specialized goods. Despite the expansive nature of outer space making, space resources seem virtually unlimited, and the realities of distance and

resource concentrations mean that there are limited numbers of economically attractive areas in the near term. In particular, nations are eyeing a limited number of lunar craters on the South Pole expected to have large water deposits with readily available sunlight to power bases. As these geopolitical ambitions drive space resource activities, law will need to play a mediating role between countries and the private interests of each.

In light of this geopolitical competition, several countries are establishing domestic frameworks to facilitate commercial extraction of space resources. As described in Sect. 33.2.2.1, the United States is emerging as a clear leader in this regard. The United States passed the first space resources law, the Commercial Space Launch Competitiveness Act of 2015 (U.S. Space Launch Act 2015). This law codified the property rights of United States citizens who extract resources in outer space, while also granting United States courts the ability to oversee certain disputes in space. In April 2020, the Trump administration issued an Executive Order to further support commercial space resources development. In September 2020, NASA issued a solicitation to purchase lunar regolith or rocks in place on the Moon from a commercial company; if completed, the solicitation would represent the first commercial space resources extraction and support United States interpretation of international law as allowing commercial production (Gilbert and Bazilian 2020). Possibly the most promising United States activity is its recent proposal for the Artemis Accords to govern space resources production and other activities on the lunar surface. These represent a multilateral agreement between the United States and its space allies that can form the framework for future space resource activities for most space powers (Stimers and Dinegar 2020) (Sect. [Recent Trump Administration Actions: Executive Order 13,914 and the Artemis Accords](#)). If successful, the Accords can establish state practice and shape international customary law for space resources, perhaps forming the foundation for a global treaty. Beyond legal measures, the United States is taking policy actions to support space resources technology research and development. Luxembourg has established a conducive framework for space resources (Sect. 33.2.2.2), while other countries are examining how to do so (Sect. 33.2.2.3).

As nations develop domestic laws consistent with their interpretations of international space law, there are many issues that need to be addressed. Since the failure of the Moon Treaty to achieve widespread adoption, efforts to create international hard law governing space resources production have been unsuccessful. Recently, the non-governmental Hague International Space Resources Governance Working Group has developed soft-law guidance in the form of Building Blocks (Hague Working Group 2019). These Building Blocks, described in Sect. 33.2.3, are a result of collaboration with space lawyers and other stakeholders in multiple countries. They focus on the application of existing statutory and customary international law to commercial space resources production, while also recommending consideration of certain factors that are not currently established in space law. The Building Blocks can shape and guide national activities to create domestic legislation while also influencing the development of international agreements, such as the Artemis Accords.

There are many potential issues to address and manage as nations look to begin commercial space resources production. The Outer Space Treaty's invocation of

international law applying to outer space provides the basis for understanding and navigating concerns as a space mining industry emerges. The proximate question of access to locations in space and property rights over space resources gathered is foundational for commercial development. As described in Sect. 33.3.1, the existing international regime and domestic activities provide a basis to understand these issues, but their application still presents concerns.

Natural resource extraction has substantial environmental impacts on Earth and there is a potential for similar impacts in space as well. Yet international environmental space law is in its infancy. The Outer Space Treaty and Liability Treaty create some obligations but how they will apply to the risks involved in space activities are unclear. Section 33.3.3.1 provides an overview of these environmental considerations, with a focus on space debris, the impact of mining, and nuclear contamination. Perhaps most worrisome is the potential for forward and backward contamination. Forward contamination, also called planetary protection, involves the introduction of terrestrial life to an environment with extraterrestrial life and could complicate the scientific search for life. Backward contamination concerns the potential for materials or crew returning to Earth with extraterrestrial life, potentially posing public health risks. Some international fora can help guide standards to prevent both but these remain in the realm of soft law. Similarly ambiguous are efforts to conserve portions of outer space to protect natural resources or to preserve historic sites such as the Apollo landers (see Sect. 33.3.3.2). While both may be necessary for public acceptance of space mining, efforts to achieve them require domestic and international policymaking.

Other than property rights, the most contentious issues facing nations looking at space resources production are liability (Sect. 33.3.4) and dispute resolution (Sect. 33.3.5). With national space resource activities, the Outer Space Treaty and Liability Treaty provide a clear if untested legal regime that can enable nations to handle claims arising from interference. However, the involvement of private actors in commercial utilization schemes creates more complexity and challenges as the treaties govern relationships between nations. National legislation to channel national liability to private operators is likely necessary but requires careful design. Similarly, handling disputes between multiple private actors from different countries can prove challenging and make private grievances into matters of foreign policy.

33.2 Legal Framework

Developing natural resources in outer space will be technically challenging. Similarly, states and private parties engaging in extraterrestrial resource development will need to think creatively about how to finance these ventures, and how to secure an economic return. In addition to these technical and financial considerations, those exploring for and developing natural resources in outer space do so under considerable legal and regulatory uncertainty. Stakeholders continue to debate whether international law allows for private ownership of resources extracted in outer space.

Countries like the United States and Luxembourg have developed statutory frameworks designed to ensure that private entities can own outer space resources (Space Launch Act 2015) (Luxemburg Space Law 2017), but even these statutes do not specify how an enterprise is able to obtain those rights. Nonetheless, it is likely that space miners could move forward with the development of resources in outer space under existing laws and treaties. According to Hertzfeld and von der Dunk (2005, pp. 81–82), leading authorities on space law:

International agreements declare that no government can claim outer space or celestial bodies in outer space as its own. Private firms seeking to invest in potential space enterprises frequently point to these provisions as a major barrier to the future commercial development of space. Such businesses contend that the absence of property rights prevent them from obtaining external financing, hinder the protection of their investments in space, and deprive them of the assurance that they can appropriate income from their investment. In short, the lack of sovereignty in space jeopardizes the ability to make profits from private investment.... [But] most property rights exist in space and... the lack of sovereignty does not pose current or near-term problems for the types of business ventures likely to be developed in space. Furthermore, even in the case of future ventures, solutions based on terrestrial models would permit private companies to operate in space with reasonable reliance of the right to appropriate income from their investments.

In the subsections below, we consider the legal framework that applies to the extraction of resources in outer space. This includes a review of both relevant international law and the domestic laws of individual nations which seek to provide a framework for space resources activities conducted under their authority and oversight.

33.2.1 International Law

Terrestrial mining companies work all over the world. In mining, a company has to go where the minerals can be found and developed. Minerals are often located in areas where the legal framework for mining is unsettled. The factors considered by a mining company or investor when assessing whether to proceed with a project in a jurisdiction with an incomplete or uncertain legal regime are the same factors that will be considered when determining whether to develop resources in outer space.

A mining project has an economic structure that is different from most industries. A mine requires capital to secure mining rights, conduct environmental and engineering analysis, engage in exploration, navigate a permitting process, and then build a mine and process facility. Only after the expenditure of all this time and money will a mining project begin to generate revenue, and it is even longer before a mine will start to show a profit. A mining project in outer space, whether on the Moon or an asteroid, will follow a similar (but more dramatic) investment and return curve. The magnitude of these initial capital investments is almost certain to be substantially greater for space resources ventures than for their terrestrial mining counterparts, as space mining companies will have to first invest in the initial development of all of the hardware necessary to extract resources (i.e., akin to a terrestrial miner being

forced to invent a new form of ore transportation mechanism at the outset of a project, rather than simply using existing technologies such as haul trucks). Also, there are no existing markets for resources produced by space miners, and those markets will need to be developed.

When developing a mining project on Earth, an international mining company considers the following key issues when determining whether to proceed with the project.

Security of Tenure. Research on terrestrial mining investment indicates that security of tenure is a key element in investment decisions, especially given the uncertainties and long-term nature of mineral projects (Vilhena 2005). Security of tenure gives the mineral developer a sufficient term within which to carry out exploration and mining operations, a right to mine what has been discovered, and clear and objective obligations in order to maintain those rights in effect (Vilhena 2005).

Central to security of tenure is the right of exclusivity. For Earth-bound mining, countries with a well-developed minerals industry have a cadaster system or other method for ensuring that a miner either has the exclusive right to conduct mining operations in an area, or a pathway to acquiring exclusivity. In countries that are in the process of developing a mining industry, sometimes the allocation of an exclusive right to mine may be less certain. Even in those circumstances, mining companies may invest in a project in those countries, with the thought that the company's exclusive right to mine can be established at some future point.

As discussed in Sect. 33.2.1.1, the Outer Space Treaty does not provide an unequivocal process to assure that a space mining company can achieve the security of tenure necessary to move forward with an investment in space mining. The U.S. Space Launch Act 2015 and the Luxembourg Law on the Exploration and Use of Space Resources are both designed to address that uncertainty, and provide a legal framework for securing and recognizing the right to extract resources in space. The adoption of laws or treaties consistent with the Hague Working Group Building Blocks (Hague Working Group 2019), discussed in Sect. 33.2.3, would provide greater certainty. The Building Blocks recommend a legal framework for access to space resources, which would facilitate exploration, and an international framework to ensure the lawful acquisition and use of space resources.

Nor is there an agency or governing body that can grant the exclusive right to mine on the Moon, on Mars, or on an asteroid. There is at least the foundation for the recognition of mining rights on the Moon (see Sect. 33.3.1.3), but there is no formal, agreed process for registering Moon mining claims, or acknowledging the priority of rights to minerals on the Moon. Similarly, as discussed in Sect. 33.2.3.1, the Hague Building Blocks recommend creation of a registry to acknowledge the priority of rights to resource development. At some point, there may be an international agreement that sets up a mining claim registration system. It is more likely, however, that the development of minerals in space will take place under evolving legal systems and structures, and become commonly accepted over time. The creation of "safety zones" under the Artemis Accords (discussed in Sect. Recent Trump Administration Actions: Executive Order 13,914 and the Artemis Accords), for example, has been

accepted by several space-faring nations, and could provide a conceptual foundation for the recognition of exclusive areas for mineral development. Even absent an affirmative international agreement, these practices could become accepted as the standard approach and become customary international law.

Security of Title. The World Bank describes “security of title” as a system where “[l]icenses and mineral rights cannot be suspended or revoked except on specific grounds, which must be objective and not discretionary, and which must be clearly specified and detailed in the legal framework” (Girones et al. 2009). A mining company or investor, whether operating on Earth or in space, will want comfort that its right to develop minerals is stable, and can be challenged only in a principled manner (Girones et al. 2009, p. 8).

Fiscal Regime. It is common in terrestrial mining for the mining company to pay a royalty or fee to the nation with jurisdiction over the minerals being developed. The Outer Space Treaty does not include a process for charging a fee for the use of outer space, and there is similarly no mechanism for imposing a rental or royalty on minerals extracted from the Moon, Mars, or an asteroid. As the industry of resource development in outer space advances, however, governments may look for ways to tax the enterprise. If, for example, a company takes advantage of the legal frameworks established by the United States or Luxembourg, those countries could impose some severance tax or royalty payment in addition to the fees associated with forming companies under their laws. Of greater concern to a space mining company is the risk that other countries will invoke the concept in the Moon Treaty that space is the “common heritage of mankind,” or perhaps the language in the Outer Space Treaty that the exploration and use of outer space shall be carried out “for the benefit of all countries.” This language might be used to assert some economic interest in space resources, payable in a royalty or perhaps in kind. This concern could be addressed in part by adopting the “due regard” standard in Law of the Sea, as proposed in the Working Group Building Blocks (Hague Working Group 2019, Sect. 9). Under that standard, a space mining company would have the freedom to conduct activities in space so long as those activities do not adversely affect the use of outer space by nationals of other states. At present, the risk of the future imposition of a fiscal burden on resource development in outer space remains one of the potential risks to be factored into a decision to move forward with a celestial mining project.

Bankability. A viable mining project requires more than the presence of minerals in the ground. Those minerals must be capable of extraction, processing and sale in a manner that provides an economic return on the investment in the mine. A terrestrial mining project typically requires a detailed feasibility study to determine whether minerals can be developed economically. The Canadian Institute of Mining, Metallurgy and Petroleum (Canadian Reporting Standards 2014) provides a useful definition of a feasibility study:

A Feasibility Study is a comprehensive technical and economic study of the selected development option for a mineral project that includes appropriately detailed assessments of applicable Modifying Factors together with any other relevant operational factors and detailed

financial analysis that are necessary to demonstrate, at the time of reporting, that extraction is reasonably justified (economically mineable). The results of the study may reasonably serve as the basis for a final decision by a proponent or financial institution to proceed with, or finance, the development of the project.

The “Modifying Factors” include mineral reserves, mining, processing, metallurgical, infrastructure, economic, marketing, legal, environmental, social and governmental factors (Canadian Reporting Standards 2014). A feasibility study becomes “bankable” when it presents a project that is of sufficient quality to attract financing. A bankable feasibility study should include information sufficient to make a decision to proceed with the project and to enable an investor or financial institution to decide whether to commit finances to the project (Messer 2015, Sects. 6–2). While the term “bankable” is often “mistakenly applied as if it represented a settled form of product or output,” and represents an inherently fluid concept that depends on the specific market, deal and institutions involved, the term generally denotes that the study achieves a quality and standard acceptable for submission to bankers (Messer 2015, Sects. 6–3). The bankability of a space mining project will be less certain. In terrestrial mining projects, lenders are protected by their ability to take a secured interest in the assets of the projects, for example through a mortgage on the project’s assets on which the lender can foreclose if the miner defaults on its debts. By contrast, it may be hard to take out a mortgage on an asteroid, and lenders and miners will need to be creative in setting up secured interests that will provide enough comfort to the lender to facilitate financing at interest rates that are acceptable to the miners. Even more crucial to bankability, though, will be the certainty of the legal and commercial regime applicable to the mining venture. As noted above, the Outer Space Treaty—and specifically the interpretation of the Outer Space Treaty underlying the legal regimes created by the United States and Luxembourg—remain unsettled. One commentator, for example, argues that that the United States’ space mining law violates the Outer Space Treaty by purporting to create property rights in outer-space resources for private entities that extract them (Taylor 2019). Investors may require further clarity around those risks before making an investment. It is likely that investments will be made in stages, as space mining ventures develop and test the technical, commercial and legal structures necessary to move forward with a project.

Enforceability. A mining company wants to know that its agreements are enforceable. Mining often takes place in jurisdictions where the rule of law is uncertain or incomplete. Mining companies may use international arbitration clauses and international treaties to enhance the enforceability and stability of their mining rights and related agreements. For resource development in outer space, a private or national mining company will rely on the Outer Space Treaty and other international law, but as discussed in this chapter, those instruments do not provide a clear and unequivocal legal framework for resource development. As a result, a mining company will face some uncertainty if a private entity or nation claims a prior right to the resources being developed, or asserts a claim to some portion of the proceeds derived from resource extraction and use. A space mining venture is more likely to move forward

if the mining company has some assurance that its rights will be recognized and enforced, and if the company has access to a dispute resolution mechanism that will provide for the adjudication of those rights. The Hague Working Group includes in its Building Blocks the recommendation that such disputes be subject to arbitration under the Rules for Arbitration of Disputes Relating to Outer Space Activities (Hague Working Group 2019, Sect. 19). These rules, however, apply only when parties have agreed to such arbitration (Permanent Court of Arbitration 2011). The arbitration rules would not be available to adjudicate claims from competing companies or individuals, or claims made by non-spacefaring nations under the “common heritage of mankind” construct, (for example, Williams 2017; von der Dunk 2001). Also, adjudication of these claims in a single country may not be honored in other countries. As a result, it may be advisable to establish an international adjudicatory body to address these claims, similar to the International Tribunal for the Law of the Sea that exists under UNCLOS (United Nations Treaty Series 1982, Art. 186 and Annex VI). Dispute resolution for resource development in outer space is discussed in greater detail in Sect. 33.3.5.

33.2.1.1 Outer Space Treaty

Key Provisions

The Outer Space Treaty is the foundational text of international space law (United Nations Treaty Series 1967). It entered into force in 1967 and has been signed and ratified by over 100 nations, including the United States and most other space-faring nations. While the original impetus for the Treaty was to prevent any one nation from gaining a military advantage in space, the Outer Space Treaty sets out broad principles that provide a framework for commercial activities in outer space, including resource development.

The Outer Space Treaty provides (United Nations Treaty Series 1967, Art. I):

[t]he exploration and use of outer space, including the moon and other celestial bodies, shall be carried out for the benefit and in the interests of all countries ... Outer space, including the moon and other celestial bodies, shall be free for exploration and use by all States without discrimination of any kind, on a basis of equality and in accordance with international law, and there shall be free access to all areas of celestial bodies.

The Outer Space Treaty also addresses the ownership of the Moon, planets, and asteroids: “[o]uter space, including the moon and other celestial bodies, is not subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means” (United Nations Treaty Series 1967, Art. II).

An important principle in the Outer Space Treaty imposes supervisory obligations on nation states (United Nations Treaty Series 1967, Art. VI):

State Parties to the Outer Space Treaty shall bear international responsibility for national activities in outer space, including the Moon and other celestial bodies, whether such activities are carried on by governmental agencies or by non-governmental entities, and for

assuring that national activities are carried out in conformity with the provisions set forth in the present Outer Space Treaty. The activities of non-governmental entities in outer space ... shall require authorization and continuing supervision by the appropriate State Party to the Treaty.

As private parties become more active in outer space, the supervisory role of states will gain greater significance (Marboe and Traummüller 2013, p. 73):

In view of the diversification of actors it became necessary to find means to transform the international obligations imposed on States to obligations incumbent on private actors in order to ensure that private entities comply with international space law and its principles.

Application of the Outer Space Treaty to Resource Development

The Outer Space Treaty assures a right of free access to celestial bodies for all nations, and it prohibits appropriation or national ownership of the bodies themselves. The Outer Space Treaty does not directly address space resource extraction or provide an unequivocal statement that such activities would be authorized under the Treaty. There are several ambiguities that must be considered when assessing how natural resource development in space will proceed under the terms of the Outer Space Treaty.

The Outer Space Treaty does not include a uniform definition of “celestial body,” and does not specifically state that asteroids fall within the scope of “celestial bodies” (Tennen 2010, p. 796). If asteroids are not celestial bodies, then the Outer Space Treaty would not apply to them. But the standard view is that asteroids are celestial bodies (Tennen 2010, pp. 796–97). The International Astronomical Union, for example, includes asteroids in its definition of celestial bodies (Tennen 2010, pp. 796–97). There is little practical doubt that asteroid development would fall within the scope of the Outer Space Treaty.

Also, it is not clear how to interpret the Treaty’s statement that the exploration and use of space “shall be carried out for the benefit and in the interests of all countries.” One interpretation is that this clause mandates an international profit-sharing mechanism. The United States and others have taken the position, however, that it merely reiterates the right of free access articulated in Article I (Gold 2017). This language is different from the language of the Moon Treaty, which invokes the language of the Law of the Sea and describes space as “the common heritage of mankind” (United Nations Treaty Series 1979, Art. 11(1)). Those advocating for a narrow reading of this provision of the Outer Space Treaty often cite this difference as evidence of an intent in the Outer Space Treaty to use “benefit and interest” as a less expansive concept than the notion of “the common heritage of mankind.” This distinction is discussed in Sect. 33.3.2.

One of the more fundamental questions is whether the prohibition on national appropriation would limit the exercise of private rights over extracted resources. In other words, can private entities own resources that have been extracted from the celestial body without any nation owning the body itself? The Outer Space Treaty includes the phrase “exploration and use” twice in its terms. The word “use” seems to

indicate that development and deployment of space resources is expected, and thus, not prohibited (Gabrynowicz 2014). The Outer Space Treaty does not describe how mining rights would be allocated and recognized when no nation may appropriate the celestial body being mined. The diplomatic history of the Treaty indicates that this point may have deliberately been left ambiguous in order to gain broader support (Roth 2016, pp. 841–42). As a result, those considering the future of space mining are looking for ways to create a system of recognition or registration that will fill this gap.

This legal framework is similar to the framework for the extraction and utilization of resources, such as fish and minerals, from the high seas under UNCLOS. While the high seas are outside the jurisdiction of any single nation, resources extracted from the ocean can still be subject to domestic property laws (Roth 2016, p. 851). The same framework might apply to the Moon and asteroids: granting private property rights to asteroid resources does not conflict with the Outer Space Treaty prohibition on national appropriation of celestial bodies. At least one commentator has raised, but ultimately rejected, the possibility that the legal principle of accession—through which a person acquires ownership of property by improving it or combining it with other materials—might allow asteroid miners to avoid this non-appropriation issue entirely by using 3D printing to turn the harvested asteroid resources into an entirely new product to which the producers could claim title (Chatzipanagiotis 2016, p. 247).

Asteroid mining and fishing or mineral development in the high seas are not the same thing, however. It might, for example, be possible to retrieve an entire asteroid and break it up to extract and use its mineral content. As one scholar asks (Chatzipanagiotis, 2016, pp. 851–52):

If an asteroid-mining enterprise obtained control over a small asteroid in its entirety, with the intention of making use of all of its mineral content, would that be extraction of ‘asteroid resources’ or assertion of exclusive rights over the territory?

At this stage, there is not a definitive answer to this question, and that question may not be answered until there is a proposed activity that poses the test.

Even though there is some uncertainty about the interpretation and application of the Outer Space Treaty, the United States State Department has consistently maintained that the Outer Space Treaty allows for commercial extraction and ownership of resources (Schaefer 2017). According to the State Department, the Outer Space Treaty merely shapes the manner in which space resource utilization may be carried out; it does not preclude such activities, and the Treaty’s non-appropriation principle applies to space resources only when such resources are “in place” (Egan 2016). Once resources are removed from the celestial body, the prohibition on non-appropriation does not apply (Egan 2016). There may be another avenue to avoid application of the non-appropriation principle: while states may not appropriate celestial bodies, there is no prohibition on their appropriation by private parties (Tronchetti 2017). The existence of this latter loophole is questionable, given that under the Outer Space Treaty, and almost any existing or proposed approach to space activities, every private party must act under the authority and responsibility of a state.

Despite this ambiguity, the most common view of this issue is that the Outer Space Treaty does not prohibit the development and ownership of extracted minerals. The International Institute of Space Law, for example, takes the position that while the Outer Space Treaty does not create an express right to take and consume space resources, it also does not prohibit such action. Specifically, the Institute points out that Article I provides for the free exploration and use of outer-space celestial bodies without discrimination. While the Outer Space Treaty expressly states that the extraction of non-renewable natural space resources is encompassed in “free use,” (Tronchetti 2017), the concept of free use certainly implies that space resources can be utilized. As one commentator has observed, “it is widely (though not universally) accepted that commercial exploitation is lawful so long as it does not prevent any other entity from undertaking the same activity in space” (Force 2016, p. 267). Under this view, “natural resources ‘in place’ are still part of the territory and cannot be owned[,] but once the resource is removed and no longer ‘in place’, it may be extracted for non-scientific (i.e., commercial) purposes” (Force 2016, p. 268). The best and clearest interpretation of the Outer Space Treaty supports the extraction and ownership of minerals from celestial bodies. That said, more work is required to create commercial, institutional or legal structures that will allow a space mining company to get comfortable with undertaking that type of project.

Finally, there is some uncertainty concerning the requirement in Article VI that states authorize and supervise space activities. It may be that Article VI requires only light-touch regulation, and some commentators believe that this mandate exists only to ensure that activities are carried out in conformity with international legal obligations. Even light-touch regulation, however, would require some sort of mission authorization framework. The United States, for example, does not currently have a designated agency or process by which to authorize commercial space mining missions, although the Trump administration has taken actions (discussed further in Sect. [Recent Trump Administration Actions: Executive Order 13,914 and the Artemis Accords](#)) that could be viewed as an early step in encouraging the development of this type of regulatory system. Private companies and investors will look for a stable mission authorization system as part of their assessment of the viability of a space resource development enterprise.

Some argue that concern about the authorization and supervision requirement in Article VI is misplaced because this requirement is not self-executing (Dunstan and Szoka 2017). That is, the provision requires a launching state to enact domestic legislation that is binding and enforceable (Medellin v. Texas 2008). Under this view, because Article VI is not specific about its requirements, adherence to it is merely aspirational, subject to each nation’s implementation as it may see fit (Dunstan and Szoka 2017). Governments may be held absolutely liable for the actions of their citizens in space, although only for terrestrial damage and damage to airplanes in flight (Hertzfeld and van der Dunk 2005). Moreover, whether the provision is self-executing does not change that it imposes an international obligation on member states, the avoidance of which risks foreign retaliation and threatens the business case for space mining conducted under the legal framework developed by a country like the US or Luxembourg (Schaefer 2017).

While the Outer Space Treaty provides a potential basis for space resource extraction, there are issues that require further guidance and development. As discussed in Sect. 33.2.2, domestic laws may provide increased certainty where it is needed for companies and investors to engage in space mining.

33.2.1.2 Moon Treaty

The Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (commonly called the “Moon Treaty”) (United Nations Treaty Series 1979), addresses resource extraction from the Moon, and likely also applies to asteroids (Roth 2016). It declares that the Moon and other celestial bodies in the solar system, as well as their natural resources, are the “province of all mankind” (United Nations Treaty Series 1979, Art. 4). This language characterizes the bodies and their resources as being the “common heritage of all mankind” (United Nations Treaty Series 1979; Roth 2016), a term of art that is also used in UNCLOS, where it has been interpreted to require some manner of international profit sharing (as further discussed in Sect. 33.3.2). Indeed, some commentators have argued that the Moon Treaty was designed to preclude resource development (Simberg 2012):

The 1979 Moon Treaty ... was written in such a way as to effectively preclude resource extraction, since any entity attempting to utilize extraterrestrial resources would have to operate under an undefined “regime” whose primary purpose would have been the “equitable” distribution of the profits—to non-space-faring countries.

That view of the Moon Treaty may be overly harsh, but it is certainly true that the Moon Treaty adopts a conceptual framework for the use of outer space that is different from that expressed in the Outer Space Treaty.

The Moon Treaty has been signed by fewer than twenty countries and was not signed by the United States or other space-faring nations. As a result, some have regarded the Moon Treaty as obsolete (Simberg 2012, p. 844). This includes President Trump, who in April 2020 issued an Executive Order stating explicitly that “the United States does not consider the Moon [Treaty] to be an effective or necessary instrument to guide nation states regarding the promotion of commercial participation in the long-term exploration, scientific discovery, and use of the Moon, Mars, or other celestial bodies” and instructing the United States Secretary of State to “object to any attempt by any other state or international organization to treat the Moon [Treaty] as reflecting or otherwise expressing customary international law” (Exec. Order No. 13,914 2020). If the Moon Treaty had found favor among the international community, it could have presented a significant hurdle to private development of space resources. However, the Moon Treaty’s limited number of signatory states, coupled with the United States’ clear (and recent) rejection of its terms, suggests that there is unlikely to be a renewed international interest in the principles of the Moon Treaty in the foreseeable future.

33.2.2 *Domestic Law*

33.2.2.1 **United States**

United States Commercial Space Launch Competitiveness Act of 2015

In 2015, Congress passed the United States Commercial Space Launch Competitiveness Act, sometimes referred to as the Spurring Private Aerospace Competitiveness and Entrepreneurship Act of 2015. The Act was the consolidated outcome of four bills that expanded existing regulation of commercial space activity (U.S. Space Launch Act 2015; Dodge 2016). Most important, for space mining purposes, is Title IV, which establishes a basis for private ownership of extracted space resources (U.S. Space Launch Act 2015, Sects. 102–117). The other parts of the Act include the following:

Title I, the “Spurring Private Aerospace Competitiveness and Entrepreneurship Act” (“SPACE Act”), updates requirements for the commercial launch industry (U.S. Space Launch Act, Sects. 102–117).

Title II, the “Commercial Remote Sensing,” affirms congressional oversight of the commercial space industry and requires additional executive branch reports regarding the licensing process for private space-based remote sensing systems (U.S. Space Launch Act 2015, Sects. 201–202).

Title III, “Office of Space Commerce,” renames the “Office of Space Commercialization” the “Office of Space Commerce,” and clarifies its functions (U.S. Space Launch Act 2015, Sects. 301–302).

Title IV, the “Space Resource Exploration and Utilization Act,” creates private property rights over resources extracted from space (U.S. Space Launch Act 2015, Sects. 402–403). This portion of the Act directs the President to (1) facilitate the commercial exploration for and commercial recovery of space resources by United States citizens; (2) discourage government barriers to the development of such industries in a manner consistent with United States international obligations; and (3) promote the right of United States citizens to engage in such industries free from harmful interference (U.S. Space Launch Act 2015, Sect. 51,302(a)). Further, it requires the President to submit a report to Congress that identifies the authorities that will be responsible for overseeing space resource extraction missions (U.S. Space Launch Act 2015, Sect. 51,302(b)).

Most critically, this portion of the Act (U.S. Space Launch Act 2015, Sect. 51,303) establishes that:

[a] United States citizen engaged in commercial recovery of an asteroid resource or a space resource under this chapter shall be entitled to any asteroid resource or space resource obtained, including to possess, own, transport, use, and sell the asteroid resource or space resource obtained in accordance with applicable law, including the international obligations of the United States.

The Act defines “asteroid resource” as “a space resource found on or within a single asteroid” (U.S. Space Launch Act 2015, Sect. 51,303(1)). It defines “space resource” as “an abiotic resource in situ in outer space,” which includes water and minerals (U.S. Space Launch Act, Sect. 51,301(2)). And for the purposes of the Act, a “United States citizen” means an individual that is a citizen, an entity organized in the United States, or a foreign entity whose controlling interest is held by a person or entity in the first two categories, which could have the effect of encouraging United States-based investment in space mining ventures (U.S. Space Launch Act 2015, Sects. 51,301(3); 50,902(1)). The Act does not make clear how exactly a citizen should go about claiming rights to space resources, but the use of the word “obtained” seems to indicate a framework akin to the rule of capture in oil and gas law, or the doctrine of *pedis possessio* in mining law. *Pedis possessio* allows a miner to make a claim for a right to mine based on occupation of a potential mineral property. In the United States, for example (American Law of Mining 2020, Sect. 34.03[1]):

there must be an actual, physical occupation of the ground claimed, that a prospector must be diligently searching in good faith for a mineral discovery, and that other prospectors must be excluded, although those who make a forcible, fraudulent, or clandestine intrusion upon the possession of an earlier prospector can acquire no rights.

Finally, the statute clarifies that (U.S. Space Launch Act 2015, Sect. 51,301(1)):

[i]t is the sense of Congress that by the enactment of this Act, the United States does not thereby assert sovereignty or sovereign or exclusive rights or jurisdiction over, or the ownership of, any celestial body.

This provision is designed to ensure that the Act is consistent with the non-appropriation principle in the Outer Space Treaty.

Whether this statutory framework comports with the requirements of the Outer Space Treaty will rest on the resolution of the question of whether the Treaty’s prohibition on the appropriation of celestial bodies extends also to their extracted resources. As noted above, the International Institute of Space Law found that the Outer Space Treaty does not prohibit ownership of extract resources, and indeed implicitly authorizes ownership of space resources (International Institute of Space Law 2015). The Institute also recognizes that the harmonization of international and domestic laws in this area remains subject to development over time. (International Institute of Space Law 2015).

Luxembourg’s recent adoption of a similar, though more comprehensive, law (as described further in Sect. 33.2.2.2) may indicate that an international consensus around the legality of space resource extraction is beginning to take shape. The emerging understanding is that space resource extraction is consistent with the provisions of the Outer Space Treaty. This emerging understanding comes from the actions and interpretations of *developed* nations who currently possess an opportunity to extract such resources. Developing states might place a greater degree of weight on the requirement that such activities be carried out for the benefit of all countries (Force 2016). Moreover, industry responses to the legislation were uniformly positive. At the time, Planetary Resources co-founder Eric Anderson (Planetary Resources, 2015) stated:

This is the single greatest recognition of property rights in history. This legislation establishes the same supportive framework that created the great economies of history, and will encourage sustained development in space.

While the inter-relationship between the Outer Space Treaty and domestic laws remains a work in progress, Mr. Anderson’s statement provides some evidence that private investors and companies see domestic laws as adding certainty to the overall legal framework.

Recent Trump Administration Actions: Executive Order 13,914 and the Artemis Accords

If there was doubt that the United States was seeking to form an international consensus around its own interpretation of the Outer Space Treaty’s non-appropriation principle, it has likely been eliminated by the Trump administration’s recent executive actions on space resource development. In an Executive Order, President Trump proclaimed that “the United States does not view [outer space] as a global commons,” and he ordered the Secretary of State, working in cooperation with other relevant agency heads, to “encourage international support for the public and private recovery and use of resources in outer space” in accordance with this view (Office of the President of the United States 2020). Building this type of consensus is an important step in establishing international law or norms on this topic, and the United States—at least under the Trump administration—appears to view this as an important goal.

Towards that same end, the United States (acting through NASA) has very recently entered into what it calls the Artemis Accords with representatives of the national space agencies of Australia, Canada, Italy, Japan, Luxembourg, the United Arab Emirates, and the United Kingdom (National Aeronautics and Space Administration 2020). This agreement has been entered into in connection with NASA’s own Artemis program, which has targeted a return to the Moon by 2024 and from there further exploration on to Mars (Strickland 2020).

As a threshold matter, the Artemis Accords bring into focus the United States’ strategy to address the legal uncertainty caused by the conflicting interpretations of the Outer Space Treaty’s non-appropriation principle. As described in Sect. 33.2.1, would-be industry entrants currently face concerns over whether they are entitled to own the resources they extract, or whether another entity or nation might challenge their property rights as violating the non-appropriation principle. This uncertainty has been cited as a significant hurdle to increased investment and industry development. Some have considered whether this issue might be best addressed by widespread agreement on a new international regulatory regime; the Building Blocks proposed by the Hague Working Group (discussed in detail in Sect. 33.2.3) represent an attempt to outline what such a regime could and should look like if implemented. However, implementing this type of new system through an international treaty and obtaining buy-in from a sufficient number of signatory states would likely be a long

and arduous process of international diplomacy. By simply forging ahead with the Artemis Accords, the United States seems to be signaling that rather than wait for a new treaty to replace or supplement the Outer Space Treaty, it hopes to create an international consensus by entering into an agreement with other key space-faring nations that will memorialize the United States' view of how space activities should be regulated.

Like the Outer Space Treaty, the Artemis Accords provide a high-level framework addressing key principles for space resources extraction, while retaining a certain amount of ambiguity and flexibility (likely necessary in order to gain adoption by the current signatory parties). Notably, “[t]he principles set out in the [Artemis] Accords are intended to apply to civil space activities conducted by the civil space agencies of each Signatory” (National Aeronautics and Space Administration 2020, Sect. 1). The key points addressed include the following items.

- Of greatest interest to would-be space miners, the Artemis Accords create a system in which the parties agree that space resources can be extracted and used without violating the Outer Space Treaty, thereby further reinforcing the United States' interpretation of the Treaty's non-appropriation principal (National Aeronautics and Space Administration 2020, Sect. 10). Parties will also implement a system to create “safety zones” around each country's operations in order to avoid interference with one another's space activities (National Aeronautics and Space Administration 2020, Sect. 11). This is another concept that had been proposed as a recommended best practice, but which might arguably all foul of the Outer Space Treaty's non-appropriation principle.
- The Artemis Accords commit the parties to the Outer Space Treaty's principle of using space for only peaceful purposes, as well as to the principles of the Rescue Agreement and Registration Agreement (described further in Sect. 33.3.2) (National Aeronautics and Space Administration 2020, Sects. 3, 6–7).
- Parties have (National Aeronautics and Space Administration 2020, Sect. 5):

commit[ted] to us[ing] reasonable efforts to utilize current interoperability standards for space-based infrastructure, to establish such standards when current standards do not exist or are inadequate, and to follow such standards.
- Partners have agreed that they “intend to preserve” historically significant sites, such as the Apollo 11 lunar landing location, pursuant to standards to be agreed upon amongst the parties (as discussed in Sect. 33.3.3.2, and in the Hague Working Group's Building Blocks) (National Aeronautics and Space Administration 2020, Sect. 9).

The actual implementation of the Artemis Accords could create a clear tipping point for the future of international law regarding space resource utilization. It appears the United States hopes to build a consensus of space-faring powers who agree to the United States' vision for the controlling legal framework. If another country with a contrary interpretation of the Outer Space Treaty wants to challenge this interpretation, it will most likely do so as the Artemis Accords are first being put into

action and before the Accords gain widespread acceptance as customary international law.

Customary international law is the general practice of states, which is in turn generally accepted as law by states (Kundmueller 2002, p. 361). A rule of customary international law arises from a “general and consistent practice of states followed by them from a sense of legal obligation” (Scoville 2016, p. 1895). Customary international law evolves from generally accepted legal norms, manifest through the actual practices of the affected states (Scoville 2016, p. 1896). As a result, determining customary international law requires an empirical investigation of “the practices and views of the various states that comprise the international order” (Scoville 2016, p. 1896; Swainea 2015; Charlesworth 1998). In seeking customary international law, courts, arbitration panels, and other concerned stakeholders will look at how countries act, which norms are generally accepted, and the official pronouncements (including statutes, international agreements, and state judicial decisions) to see whether a norm or principle is subject to “mutual consensus” (Nanda et al. 2020).

The Outer Space Treaty, and even the more specific Artemis Accords, provide fairly high-level guidance to those engaged in the development of resources in outer space. Still, there will be a lot of specific actions and decisions that require a legal framework, and those principles will evolve over time. That means, of course, that the first actors in outer-space resource development will face some uncertainty about the legal basis for their decisions and activities. Also, the actions of private parties, without more, cannot create customary international law. Those decisions and strategies and structures will need to be accepted by the international community to achieve the mutual consensus necessary to be endorsed—and enforced—as customary international law.

Mission Authorization

As noted above, the Outer Space Treaty requires non-governmental space activities to be authorized and subject to continuing supervision by a state. The United States, therefore, must implement a regulatory framework that meets this obligation. However, at present, in the United States, no such regulatory framework for non-orbit space activities like mining exists. A comprehensive and manageable mission authorization regulatory system is the next step towards the United States’ goal of becoming a hub for commercial space mining ventures.

In the United States, comprehensive regulations exist for traditional space activities such as launch and re-entry, remote sensing, and satellite communications. Different federal agencies administer these regulatory frameworks, and the frameworks tend to be isolated within those agencies. For example, launches must be licensed with the Federal Aviation Administration (FAA) (Federal Aviation Administration 2020). Transmission of satellite communications via spectrum must be licensed with the Federal Communications Commission (FCC) (Telecommunications Act of 1996). Missions that involve remote sensing of Earth require a license from the National Oceanic and Atmospheric Administration (NOAA) (National and

Commercial Space Programs Act of 2010). The FCC and NOAA also impose rules pertaining to debris mitigation, of which the launching entity should be aware (47 C.F.R. 25.114; Schaefer 2017). The effect of this regulatory web is that many agencies must approve a mission, and the failure to secure any one of these approvals could thwart a mission. There is no single agency with the authority to sign off on a proposed mission (Richards 2017).

President Obama's administration proposed a mission authorization framework in April 2016 to fulfill the mandate of the U.S. Space Launch Act (Holdren 2016). The proposal recommended that Congress adopt an authorization and supervision framework modeled after the FAA's payload review process (Holdren 2016, p. 4). Review of a mission would require consideration of the United States' "international obligations, foreign policy and national security interests," and protection of "United States Government uses of outer space" (Holdren 2016, p. 6). The FAA would be charged with coordinating involved agencies to review missions on a case-by-case basis (Holdren 2016, p. 4). The proposal did not contemplate the development of substantive regulations surrounding these activities. Rather, it directed the FAA to develop procedural regulations, in coordination with other space-related agencies, to this end (Holdren 2016, p. 4).

There is some concern that this inter-agency approach to mission authorization may result in an unwieldy review process (Schaefer 2017). To ensure a timely review, some scholars propose approaching mission authorization with a presumption of mission approval (Schaefer 2017, p. 167). Law professor Matthew Schaefer explains that this presumption could take the form of a "foreseeable harm" requirement (Schaefer 2017, p. 168). That is, the agency would be required to find a foreseeable harm to a specified condition in order to disapprove a mission.

In 2016, then Representative and former NASA Administrator, Jim Bridenstine of Oklahoma, proposed an alternative authorization framework called the "American Space Renaissance Act" (H.R. Res. 4945 2016). The bill would create an Assistant Secretary for Commercial Space Transportation within the Office of Commercial Space Transportation of the FAA charged with issuing such regulations "as are necessary to provide for an enhanced review and determination process for payloads and associated activities after deployment ..." (H.R. Res. 4945 2016, Sect. 309(a)(2)). Further, it would require that a decision be issued within 60 days of submission, with automatic approval if the agency failed to reach a decision within that timeframe (H.R. Res. 4945 2016, Sect. 309(a)(2)(C)(i)).

While some form of enhanced payload review, administered by the FAA, seems to be the preferred strategy to meet the authorization and supervision requirements of the Outer Space Treaty, neither of these proposals has yet been adopted by the legislature. Thus, the process by which a space mining mission can gain authorization to launch remains uncertain.

In April 2018, the House of Representatives passed the American Space Commerce Free Enterprise Act, which would give the United States Commerce Department the authority to regulate asteroid miners' actions in outer space (Grush 2018). Under this proposed legislation, the Commerce Department would regulate

in-space activities and would assume NOAA's remote sensing regulatory responsibilities, while leaving in place the FAA's and FCC's respective regulatory authority over launch/re-entry and satellite transmissions (Grush 2018). Notably, the bill also sets out a 90-day time limit on the permit approval process, subject to only one possible 60-day extension (Foust 2018). Whether this bill, or anything similar to it, will ever be signed into law remains to be seen.

33.2.2.2 Luxembourg

Luxembourg has made a concerted effort to set itself apart as a jurisdiction that has legal and regulatory structures and economic incentives that are attractive to the space mining industry, in the hopes of attracting would-be space miners to set up shop within (and bring investment to) the country (Scoles 2017). The most important feature of this effort has been Luxembourg's adoption of a legal structure that explicitly recognizes a private right of ownership in materials extracted in outer space. To date, Luxembourg and the United States are the only countries to adopt this type of domestic statute.

Luxembourg's stated goal in this initiative has been to create a "legal and regulatory framework confirming certainty about the future ownership of minerals extracted in space from Near Earth Objects such as asteroids" (Scoles 2017 (quoting Luxembourg's Ministry of the Economy's official statement about the program)). Luxembourg's Deputy Prime Minister, Etienne Schneider, acknowledged that the law is designed to "reinforce [Luxembourg's] position as a European hub for the exploration and use of space resources" (Ministry of the Economy 2017a, b).

As part of its efforts, Luxembourg also set out to provide financial backing to industry participants in a variety of forms (Scoles 2017). At the time, Luxembourg entered into agreements with two key industry players, Deep Space Industries and Planetary Resources, to provide funding for their space mining ventures (Calandrelli 2016; Araxia Abrahamian 2017). Unfortunately, neither was able to achieve success in space mining, and both were ultimately bought out by other companies with no plans to pursue outer-space resource extraction (Foust 2019). Although Luxembourg was reported to have lost 12 million euros on its investment in Planetary Resources (Foust 2019), these financial setbacks do not appear to have cooled the country's desire to establish itself as a space mining hub.

Luxembourg's space mining law, passed in 2017, provides a relatively comprehensive—albeit not particularly onerous—system of regulations governing the space mining industry. Under Luxembourg's space mining law, the country's ministers of the economy and space activities are delegated the authority to regulate and oversee outer-space resource extraction activities (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 2, 15).

Luxembourg's space mining law places the following requirements on those seeking the government's authorization of a proposed mission:

- The applicant must be an entity created under Luxembourg law, or an entity created under the law of a European country that has its registered office in Luxembourg (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 4). To this end, an applicant must provide evidence proving that its central administration and registered office is located within Luxembourg (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 7(1)).
- Applicants must have a “robust internal governance scheme,” together with a “robust scheme of financial, technical, and statutory procedures and arrangements” relating to its proposed space mining mission (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 7(2)).
- Luxembourg requires a degree of transparency regarding the ownership of any applicant. Each applicant must disclose those persons or entities who own a ten percent or greater share in the applicant. If no single owner holds such an interest, then the applicant must disclose its 20 largest stakeholders (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 8(1)).
- Luxembourg’s space mining law imposes subjective standards on applicants’ management teams, requiring that they “be of sufficiently good repute” and “possess sufficient knowledge, skills and experience to perform their duties” (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 9(1)).
- Applications must include a mission risk assessment, including a description of how the applicant will be financially covered with respect to such risks (e.g., by insurance policies or self-insurance) (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 10).
- Applicants are required to undertake annual financial audits (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 11).
- Applicants must also pay an application fee, to be set by the relevant government ministers at an amount between 5,000 and 500,000 euros (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 13).

Under Luxembourg’s space mining law, an applicant who has received authorization for a space mining mission will be responsible for all damage caused by that mission, including damage related to work done in preparation for this mission (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 16). Further, the law sets up penalties—including fines and the potential for imprisonment—for those who violate the terms of the space mining law, either by failing to obtain a required mission authorization or by failing to comply with the terms of an authorization that has been granted (Law of July 20th 2017 on the Exploration and Use of Space Resources, Art. 18).

Following the enactment of Luxembourg’s space resources law in 2017, the country formed its own Luxembourg Space Agency (the “LSA”) in late 2018 (Ministry of the Economy 2018). Unlike NASA and other state-sponsored space agencies, the LSA does not intend to engage directly in space flight or research activities; rather, its focus will be on encouraging economic development of the space industry within Luxembourg (Ministry of the Economy 2018). As part of this mission—and seemingly undeterred by its setbacks with Planetary Resources and

Deep Space Industries—the government of Luxembourg has invested an undisclosed amount into a venture fund designed to provide financing for startup companies that are pursuing commercial space activities (Foust 2020).

33.2.2.3 Other Nations

Russia

While only the United States and Luxembourg have adopted domestic laws explicitly addressing the ownership of extracted space materials, other countries have adopted laws that relate to space mining operations more broadly. Russia is one such country, as its decree No. 5663-1 establishes a licensing system that addresses activities, including manufacturing, that occur in outer space.

This decree includes an explicit statement that (Roth 2016):

[t]he rights of jurisdiction and control over space objects, as well as of ownership thereof shall not affect the legal status of the area of outer space or the surface or subsoil of a celestial body occupied by it.

Although somewhat ambiguous, this statement’s reference to the “surface or subsoil of a celestial body” could be interpreted to prohibit ownership of extracted space resources (Roth 2016). This ambiguity would likely discourage private industry participants from basing their space resource extraction missions in Russia.

United Arab Emirates

Like Luxembourg, the United Arab Emirates (UAE) has made a recent push to establish itself as an attractive jurisdiction for commercial space activities. As a preliminary step towards this goal, in 2017 the UAE entered into a memorandum of understanding with Luxembourg that set up a five-year partnership, through which the respective governments agreed to share “information and expertise between Luxembourg and UAE space sectors in the areas of space science and technology, human capital development and space policy, law and regulation” and to “consult on questions of international governance of space to reach common positions in relevant international fora” (Ministry of the Economy 2017a, b).

In 2019, the UAE enacted a new law setting out a regulatory system for commercial space activities, including the extraction of resources from outer space (Government of the United Arab Emirates 2019). The law addresses a number of topics, including requiring permits for certain commercial space activities (including space resources extraction), setting out obligations related to the mitigation of potential harms caused by space debris, and prescribing rules for liability to be imposed upon operators participating in commercial space activities in the event those activities result in harm (Government of the United Arab Emirates 2019, Art. 14, 19, 20–24). However, in contrast to the laws of the United States and Luxembourg, the UAE’s space activities

law does not explicitly address property rights in and to any resources obtained from outer space (Government of the United Arab Emirates 2019). As a result, although the UAE's legal structure may attract investment from other types of commercial space activities, space mining entities in particular may prefer to establish their operations under the protections offered by the laws of the United States or Luxembourg.

Australia

In August 2019, Australia enacted the Space (Launches and Returns) Act 2018, which is intended to spur commercial space activities based in Australia by reducing the administrative burdens associated with obtaining the required licenses (Johnson et al. 2020). Under the Act, approval from the Australian Space Agency is required for all launches (and returns) of spacecraft from Australia, for launches (and returns) of spacecraft anywhere in the world by Australian nationals with ownership interests, and for the operation of a launch facility within Australia (Johnson et al. 2020). The Act is supported by regulatory rules that address licensing procedures, special provisions applicable to high-power rockets, and insurance requirements relating to space activities. Most notably, though, Australia's law does not address or purport to create any private property rights in any extracted space resources (Johnson et al. 2020).

33.2.3 *The Hague Space Resources Governance Working Group Building Blocks*

33.2.3.1 International Space Resources Governance Working Group

Since 2016, the Hague International Space Resources Governance Working Group (hereafter Working Group) has worked to help resolve the legal and regulatory uncertainties that have applied to, and likely hindered, outer-space resource extraction. The Working Group platform is a Consortium serviced by a Secretariat. The founding Consortium partner is the International Institute of Air and Space Law, Leiden Law School, Leiden University (the Netherlands). Members are major stakeholders from government, industry, universities, and research centers.

The Working Group's key mission is to "assess, on a global scale, the need for a regulatory framework for space resource activities and to prepare the basis for such regulatory framework" (Masson-Zwaan et al. 2016, p.164). To that end, in 2019, the Working Group formally adopted a set of "Building Blocks" that were designed as a regulatory framework for the development of resources in space (Hague Working Group 2019). These Building Blocks are intended to "create an enabling environment for space resource activities that takes into account all interests and benefits all countries and humankind" (Hague Working Group 2019, Sect. 1.1). The

Building Blocks are conceptually based upon existing international law, such as the principles that space resource development should be performed for the benefit and in the interests of all humankind and exclusively for peaceful purposes (Hague Working Group 2019, Sects. 4.1–4.3). The Building Blocks set out the following key principles:

- International responsibility for space resource activities and jurisdiction over space products
 - These Building Blocks provide that outer space resource development, and the products that result from such activities, should be authorized and regulated by individual States and intergovernmental organizations in a manner consistent with international legal principles. This reflects the Outer Space Treaty’s mandate that space activities must be supervised by a relevant State actor.
- Priority rights
 - This Building Block calls for the implementation of an international registry of extraction rights, through which space miners could register a priority right to explore and produce from a specified area for a given period of time. The circumstances of the priority right, including its size and the time it remained in effect, would be based upon the miner’s proposed activity.
- Resource rights
 - This Building Block attempts to address the critical open question regarding space resources development: Can an entity that extracts valuable materials from outer space obtain an ownership right in and to those materials? The Working Group suggests that a future international framework should address this issue by confirming, through domestic statutes or international agreements, that private ownership rights may indeed be obtained consistent with the non-appropriation principle of the Outer Space Treaty.
- Due regard for interests of all countries and humankind
 - This Building Block suggests that state actors should be required to give due regard to the interests of all countries and humankind. The concept of “due regard” appears to have been borrowed from the Article 87 of United Nations Convention on the Law of the Sea, which provides that the exercise of the freedom of the high seas must be conducted “with due regard for the interest of other States” (United Nations Treaty Series 1982, Art. 87(2)). According to the leading commentary on UNCLOS (Nandan and Rosenne 1995, Sects. 87.9(1)):

The standard of ‘due regard’ requires all States, in exercising their high seas freedoms, to be aware of and consider the interests of other States in using the high seas, and to refrain from activities that interfere with the exercise by other States of the freedom of the high seas.

By borrowing this concept from UNCLOS, the Working Group suggests that states should be entitled to free use of outer space, subject to the acknowledgment of the use of outer space by other states.

- Avoidance and mitigation of potentially harmful impacts resulting from space resource activities
 - This Building Block would implement a principle of preventing and addressing harms that could be caused by space resource extraction, including contamination, creation of space debris, or degradation of recognized outer-space heritage sites.
- Technical standards for, prior review of, and safety zones around space resource activities
 - The Working Group suggests that states should be required to implement a prior review of proposed space activities to confirm that they will be conducted safely and so as to minimize or avoid harmful impacts. Notably, the Working Group also recommends the development of a system for “safety zones,” which would allow individual operators to temporarily restrict access near the site of their operations for safety purposes without violating the principle of free access set out in the Outer Space Treaty.
- Monitoring and redressing harmful impacts resulting from space resource activities
 - This Building Block suggests that states should be required to monitor harmful impacts of the space resource activities they oversee and to respond if harmful impacts do indeed occur (including, but not limited to, considering termination or modification of the activity that caused the harm).
- Sharing of benefits arising out of the utilization of space resources
 - While the Working Group has advocated for some form of benefit sharing, it has been explicit that it is not calling for mandatory global revenue sharing similar to that implemented under UNCLOS (or under interpretations of the Moon Treaty based upon UNCLOS). The Building Block provides a list of examples of benefit-sharing mechanisms, including the fostering of technology, education, training, and capabilities in states interested in space resource development or the possible establishment of an international fund.
- Registration and sharing of information
 - The Working Group advocates that any new international framework should require states to share information regarding a number of matters, including by establishing a registry of priority space mining rights, providing notice of the location of any safety zones established around extraction activities, establishing a registry of outer-space objects (e.g., spacecraft), and maintenance of a

database of relevant information regarding space resource activities permitted by each state.

- Provision of assistance in case of distress
 - The Working Group suggests that states should be required to comply with the terms of Article V of the Outer Space Treaty and with the Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space (Rescue Agreement) (United Nations Treaty Series 1968) as they relate to participants in outer space resource extraction activities.
- Liability in case of damage resulting from space resource activities
 - This Building Block references existing treaties concerning damage in space (United Nations Treaty Series 1967, 1972) and suggests that operators should be encouraged to provide compensation for damage arising from their operations in space.
- Visits relating to space resource activities
 - This Building Block urges the affirmation of Article 12 of the Outer Space Treaty, which provides for reciprocal visitation rights by states to the sites of outer space operations conducted by other states.
- Institutional arrangements
 - This Building Block promotes the establishment of a registry of priority rights to extract space resources, as well as the development of a repository of information regarding items such as best practices applicable to space resource extraction and the location of sites of scientific interest and cultural heritage sites. Like the UNCLOS, this Building Block also contemplates the designation or creation of one or more international bodies. Unlike the International Seabed Authority (ISA), however, the principal purpose of this body or bodies would not be economic. That function of the ISA stems from the UNCLOS's global revenue-sharing mandate, which, as mentioned above, the Building Blocks do not adopt. Rather, the function of the international body or bodies contemplated by the Building Blocks would be to oversee (1) "[t]he consideration and promotion of best practices;" (2) the listing of certain "internationally endorsed" protected areas including natural sites, cultural heritage sites, and scientific interest sites; (3) "monitoring and review of the implementation of the international framework" contemplated by the Building Blocks; and (4) the international registry, database, and "any other mechanism that may be established for the implementation of the international framework."
- Settlement of disputes
 - This Building Block suggests that states should seek to resolve disputes through mechanisms such as the 2011 Permanent Court of Arbitration Optional Rules for Arbitration of Disputes Relating to Outer Space Activities.

- Monitoring and review
 - The Working Group recommends that once a new international framework is adopted, its implementation should be monitored and subject to further modification or amendment based upon the experience of the states and their outer-space operators.

These Building Blocks set forth a clear framework, consistent with the fundamental principles of existing international law, that could be used as the foundation for any new international treaty or multinational agreements that might be entered into in the future. However, the challenge in drafting and ultimately seeking international adoption of such a framework will almost certainly arise in negotiating and reaching agreement on the specific details attendant to the implementation of these broad principles.

33.3 Key Issues and Potential Solutions

33.3.1 Access and Property Rights

33.3.1.1 Introduction to Property Rights in Outer Space

Since the emergence of outer-space exploration, industrialization and commercialization, the legal framework that underpins these activities has struggled to keep pace. With the increase of private entities looking to explore, use and develop the resources that outer space has to offer, there are some key questions that need to be asked. If a company mines materials from an asteroid, does that company own the mined materials outright? How is legal ownership acquired? Who has the legal authority to confer such ownership rights?

Article I of the Outer Space Treaty 1967 states that all exploration and uses of outer space must be “for the benefit and in the interest of all mankind” (United Nations Treaty Series 1967, Art. I). Article II establishes that outer space, the Moon and celestial bodies are not subject to national appropriation by any state. Outer space is therefore established as *Res Communis* (United Nations Treaty Series 1967, Art. II). It has been suggested that the original purpose of *Res Communis* is to prevent outer space from being militarized and/or weaponized. Some consider that the Outer Space Treaty was designed to “temper the intensity of potential disputes certain to arise in future allocations of the spatial and material sources of outer space” (Zedalis and Wade 1978). Contrary to the concept of *Res Communis*, military purposes are inherently in the interest of one nation (Zedalis and Wade 1978).

Unfiltered and free access to outer space for all of mankind is the cornerstone of the Outer Space Treaty. Article VI establishes that the “activities of non-governmental entities in outer space ... shall require authorization and continuing supervision by the appropriate state party” (United Nations Treaty Series 1967, Art. VI). During

the emergence of space law, it was described as the “the law ... meant to regulate relations between States to determine their rights and duties resulting from all activities directed towards outer space and within it ...” (Lachs 1964, p. 33). Fundamentally, states are responsible for governmental and non-governmental activities. Whilst access for all in outer space is not disputed, the emergence of private entities as commercial actors has demonstrated the lack of clarity in relation to how such private entities are governed.

33.3.1.2 Property as a Concept

Property rights are unable to exist without law (Panesar 2000). Property as a concept is enforced by governments. States around the world maintain varying levels of property register, administer and charge land taxes and oversee development and regulations. Internationally this forms the basis of territorial sovereignty which at its very heart is based on acquired land and property rights. The same history of property rights and delineation does not, and arguably cannot, exist in outer space. It has been suggested that whilst creating a system of property rights in outer space, we have the opportunity to create a system not based on first possession.

The closest geographical comparison to outer space is airspace. It is important to note that there is no agreed demarcation or clear dividing line between airspace and outer space and this has been the source of much discourse (United Nations General Assembly Secretariat 1977). This has been established since the Paris Convention of 1919 (Convention Relating to the Regulation of Aerial Navigation 1930). However, legally airspace is considered to be part of the territory of the underlying state. Airspace is regulated by the underlying states’ laws or by international agreements (Marboe and Traummuller 2013). Despite this, there is no agreed legal definition of where airspace ends and outer space begins (Oduntan 2003; Cheng 1997; Chap. 14). The other notable area that we must look to for guidance is the high seas. The law of the high seas is well established and regulated by international law and not the laws of any one nation. The law of the high seas relies on a mutual understanding between nations and fixed boundaries. The political will to establish this on the same level for outer space and airspace is yet to be seen.

Outer space is arguably free from the notions of territorial sovereignty prevalent on Earth (St. John 2013). This is enshrined in Article II of the Outer Space Treaty which declares that outer space, the Moon, and other celestial bodies are not subject to national appropriation. This prevents all nations from claiming territorial or sovereignty rights over these areas of outer space and goes to the heart of what the international regime is trying to achieve: access.

There have, however, been attempts to claim territorial sovereignty of outer space. In 1976, eight equatorial states (Brazil, Colombia, Congo, Ecuador, Indonesia, Kenya, Uganda, and Zaire) argued that the geostationary orbit was a natural resource and that it was their inherent right to control sections of it (St. John 2013). These claims were unsuccessful, with the United Nations Committee on the Peaceful Uses of Outer Space (UNCOPUOS) declaring that geostationary orbit is a limited natural

resource “which must be used rationally, efficiently, economically and equitably” (United Nations Treaty Series 1994, Art. 44). The United States State Department has also rejected claims from private individuals attempting to claim parking fees from NASA for the parking of a research vessel on the basis that the asteroid was private property (United States District Court for the District of Nevada 2004).

33.3.1.3 Geostationary Orbit and Satellites

Geostationary orbit (GEO) is a good example of an international system of rights that can be looked to for guidance. This orbit is important because it is currently the most utilized orbit in outer space and has become vital to communication and navigation here on Earth. Satellites are now in abundance in GEO. The International Telecoms Union (ITU) has created an approach based on access and developed a procedure for this finite resource (United Nations Treaty Series 1994, Art. 44). It is accepted that the act of placing telecommunications satellites into orbit does not constitute an appropriation of that particular space. Entities that launch satellites into orbit do not acquire any property rights over the orbit. The current system is based upon temporary rights based on use akin to that of a license to use. The current regime does not have set rules on how long specific areas or slots can be occupied by a satellite. Therefore, provided that all of the ITU’s rules and regulations are followed, in theory this temporary right to occupy could last indefinitely. It is important to note that at the time of writing a dispute is yet to arise so therefore the system has not yet been tested, but with the increasingly overcrowded nature of the geostationary orbit, it is only a matter of time until these well-established principles are tested.

Similarly, there may be a need to create a registration system and similar temporary rights on the Moon. Some of the earliest mineral development is likely to occur on the Moon. One key requirement for developing a system of cataloging property rights is a system to map and designate those rights. It is possible to verify the location of a mining claim on the Moon using a standard lunar coordinate system. There are several alternative methods for developing a coordinate system for the Moon. For example, the coordinate system might reflect the center mass of the Moon (selenocentric) or the surface features of the Moon (selenographic) (Kettle 2017).

In 2006, the United States formed a working group to settle on a lunar coordinate system that could be used for an upcoming mission of the Lunar Reconnaissance Orbiter (LRO). The Working Group published a White Paper in 2008 setting out the rationale and basis for a standardized lunar coordinate system (Lunar Geodesy and Cartography Working Group 2008). The Working Group developed a selenocentric approach to setting up the coordinate system, using a latitude and longitude grid similar to the Earth-based system. Using Cartesian coordinates, the Moon’s equator provides the x-axis, that is, the 0° point for longitude (Lunar Geodesy and Cartography Working Group 2008). In picking the prime meridian for the Moon, this system used the “mean Earth direction,” i.e., the point in the Moon that, on average, points at the Earth (the averaging is necessary because the Moon wobbles a bit, and that point moves around) (Lunar Geodesy and Cartography Working Group 2008). With an

equator and a prime meridian determined, it is possible to provide a legal description of a parcel anywhere on the Moon. Using the lunar coordinate system set out in the LRO White Paper, for example, we can locate the Apollo 11 landing site at 0.67416° N, 23.47314° E. While the coordinate system for the Moon continues to be reviewed, the LRO White Paper provides a system sufficient to allow Moon miners to identify the areas where they will conduct their work.

33.3.2 Due Regard and International Benefit Sharing

The United Nations developed the Moon Treaty in 1979 (United Nations Treaty Series 1979). Only sixteen countries have entered into the Moon Treaty, and the parties do not include key industrialized nations like Russia, China, or the United States (Marboe 2016). Significantly, the Moon Treaty describes the Moon as “the common heritage of mankind,” and it may well be this concept that has prevented wider acceptance of the Treaty (United Nations Treaty Series 1979, Art. 11, Sect. 1). Many nations were reticent to enter into the Moon Treaty precisely because of the concern of most major space-faring states about adopting a legal framework that arguably requires mining companies to share technologies used for and proceeds from mining activities with every state on Earth (von der Dunk 2017).

Describing the Moon as the “common heritage of mankind” brings activities on the Moon in parallel with the activities in the deep sea, which are governed by the UNCLOS (United Nations Treaty Series 1982). Industrialized nations, in both instances, are concerned that their citizens will not be permitted to realize the full benefits of the substantial investment necessary to develop resources on either the Moon or the deep seabed (Marboe 2016, pp. 236–37). Notably, the Outer Space Treaty refers to outer space as “the province of all mankind,” but not as mankind’s “common heritage” (Pop 2016). Thus, the countries who are parties to the Outer Space Treaty, but not the Moon Treaty, have rejected the view that outer space should be treated in a manner analogous to the deep seabed. Still, some commentators use the phrase “the common heritage of mankind” when describing outer space, which is problematic (Pop 2016).

Article 87 of UNCLOS preserves the “freedom of the high seas.” The high seas—which are those areas of the sea outside of any state’s Exclusive Economic Zone—are “open to all States,” with this right of free access governed by UNCLOS and “by other rules of international law” (United Nations Treaty Series 1982, Art. 87(i)). As under the Outer Space Treaty, the high seas “shall be reserved for peaceful purposes” and the high seas are not subject to claims of sovereignty by any state (United Nations Treaty Series 1982, Arts. 88–89).

The treatment of resource development under UNCLOS is based on the notion that the deep seabed is the common heritage of mankind. It is described as the “Area,” which is defined as the part of the seabed that is “beyond the limits of jurisdiction” (United Nations Treaty Series 1982, Art. 1, Art. 136). As with celestial bodies under the Outer Space Treaty, a state cannot claim a sovereign right over or “appropriate” the

deep seabed “or its resources” (United Nations Treaty Series 1982, Art. 137(i)). That does not mean, however, that the resources of the deep seabed cannot be exploited (Fowler 2011; United Nations Treaty Series 1982, Art. 140).

The Convention provides for the Area to be administered by a United Nations organ, presently the ISA based in Jamaica. The ISA is charged with the orderly, safe and rational management of resources of the area in such a manner “as to foster healthy development of the world economy and balanced growth of international trade, and to promote international cooperation of the overall development of all countries, especially developing states...” (United Nations Treaty Series 1982, Art. 150). In a sense, the fundamental purpose of the ISA is economic, enabling the exploitation of the Area as a form of economic reserve, the last unclaimed mining territory on the planet. To perform this function, the ISA has a commercial arm, the Enterprise, which will act as a partner in joint ventures with licensed contractors exploiting the Area’s resources. Its purpose will be to hold and represent the economic interests of mankind in production-sharing contracts, which are expected to yield significant income. Annex III of the UNCLOS provides essential detail on how these contracts will be organized, including how the contractor will pay a royalty or a share of the proceeds to ISA. Additional provisions are designed to prevent any one state from obtaining a dominant position in the exploitation of any particular part of the Area.

These requirements—a joint venture arrangement with the ISA and the payment of something like a royalty—explain why some of the literature on space resources development is keen to point out that the Outer Space Treaty does not track the UNCLOS by describing space as the common heritage of mankind (Pop 2016).

Noyes (2011) summarizes the sliding scale of legal regimes governing the development of natural resources as follows:

The international community has developed several different types of legal regimes to govern natural resources. In general terms, these include:

- according states exclusive permanent sovereignty over natural resources, a system associated with territoriality;
- sharing resources, as in the cases of international rivers and migratory species;
- recognizing common property rights, as in the case of the high seas, where no one user has exclusive rights to resources and no one can exclude others from exploiting them, but capturing resources results in exclusive property rights; and
- recognizing property as the common heritage of mankind—or, to use a more contemporary phrase, the common heritage of humankind (CH)—whereby all manage resources and share in the rewards of exploiting them, even if they are not able to participate in that exploitation.

Countries like the United States and Luxembourg clearly see space resources as “common property right” in this formulation, and not the common heritage of mankind. Under the common property approach, resources developed from an asteroid can be owned as private property, with no obligation to share those resources or revenue from those resources, with every other country on the planet. Given how few countries signed the Moon Agreement, and the lack of space-faring nations

adopting the Moon Agreement, there is at present no firm basis for characterizing outer space as the common heritage of mankind.

33.3.3 Environmental Protection, and Conservation and Preservation of Historic Sites

33.3.3.1 Environmental Protection

As with terrestrial mining, a number of environmental concerns may attend space mining operations. Chief among these concerns is the increasing density of space debris. Additional environmental problems include nuclear contamination, other forms of contamination, and depletion of Earth's ozone layer.

These issues are not unique to the space mining industry, but apply broadly to all forms of space activities. However, space mining ventures may disproportionately exacerbate certain concerns relative to other space activities. For example, asteroid mining and processing of asteroid materials that takes place in near-Earth orbit could contribute to the problem of space debris to a greater degree than other space activities (Roth 2016). Space mining may also intensify environmental concerns in space and related to Earth's atmosphere just by virtue of increasing the scope of activities that humans undertake in space. Significantly, however, space mining may also reduce environmental burdens on Earth by, for example, enabling platinum development in space rather than on Earth (Emerging Technology 2018).

A high-level overview of the environmental issues and regulatory frameworks relevant to space mining is set out below.

Space Debris

The UNCOPUOS defines space debris as “all man-made objects, including fragments and elements thereof, in Earth orbit or re-entering the atmosphere, that are non-functional” (United Nations Office for Outer Space Affairs 2010). In other words, objects that have been sent into space and no longer serve a purpose are space debris.

Space debris is a significant source of concern to governments and non-government entities alike. At present, there are more nonfunctional than functional satellites orbiting Earth—specifically, 3,000 nonfunctional satellites and 2,000 active satellites (O’Callaghan 2020). Additionally, there are about 34,000 pieces of space debris bigger than 10 cm and 128 million pieces of space debris larger than 1 mm (O’Callaghan 2020). In 1978, NASA scientist Donald Kessler hypothesized that the density of space debris in low-Earth orbit could one day become great enough to lead to a chain reaction in which space debris continually collides with itself, rendering Earth's orbit unusable (O’Callaghan 2020).

Space debris poses significant risks for space activities of all kinds, including space mining ventures. While relatively rare at present, collisions with space debris can incapacitate active spacecraft (O’Callaghan 2020). Even particles as small as 1 cm

in size can cause significant property damage, and collisions with astronauts undertaking extra-vehicular missions can be fatal (Viikari 2015, p.722). Collisions can also release harmful substances such as radioactive material (Viikari 2015, p.722). Space debris poses a threat to Earth if it re-enters the atmosphere and crashes to Earth (Viikari 2015, p. 723).

The issue of space debris is not addressed directly in the Outer Space Treaty or the other treaties comprising international space law. However, certain of the Outer Space Treaty's provisions can be interpreted to impose some obligation to mitigate space debris. Most significant for these purposes is Article IX. First, Article IX requires that activities be conducted with "due regard to the corresponding interests of all other States parties to the Treaty" (United Nations Treaty Series 1967, Art. IX). As discussed above, a "due regard" standard requires that states "be aware of and consider the interests of other States ... and ... refrain from activities that interfere" with the interests of other states (Nandan and Rosenne 1995, Sect. 87.9(I)).

Second, Article IX (United Nations Treaty Series 1967 at Art. IX) requires that:

States ... conduct exploration of [the Moon and other celestial bodies] so as to avoid their harmful contamination and also adverse changes in the environment of the Earth resulting from the introduction of extraterrestrial matter, and, where necessary, shall adopt appropriate measures for this purpose.

Because space debris collisions can release harmful contaminants that may damage the Moon and other celestial bodies, Article IX could be leveraged as a hook to require mitigation of space debris so as to avoid this outcome. But, as one scholar put it, "the duty to avoid harmful contamination is general and aspirational, leaving indefinite the circumstances when active measures are to be taken" (Viikari 2015, p. 729 (internal quotation marks omitted)).

Finally, Article IX (United Nations Treaty Series 1967 at Art. IX) imposes a duty on parties to the Treaty to undertake international consultations if it:

has reason to believe that a [proposed] activity ... would cause potentially harmful interference with activities of other States parties in the peaceful exploration and use of outer space, including the Moon and other celestial bodies ...

A party that reasonably suspects that a planned activity of another party would cause potentially harmful interference may request consultation (United Nations Treaty Series 1967, Art. IX). Thus, Article IX could be leveraged to require international consultations if a proposed mission may result in the creation of space debris that compromises the ability of other parties to the Outer Space Treaty to peacefully explore and use outer space. Again, one scholar rightly points out the limitations of a consultation requirement; namely, that it lacks teeth and does not mandate denial of an ill-advised mission (Viikari 2015, pp. 730–31).

Additionally, under the Outer Space Treaty, parties "bear international responsibility for national activities in outer space," and must authorize and continually supervise non-governmental entities in outer space subject to their jurisdiction (United Nations Treaty Series 1967, Art. VI; United Nations Treaty Series 1972, Arts. II, III). A party that launches an object or procures the launching of an object is "liable

for damage to another State party to the Treaty or to its natural or juridical persons by such object or its component parts on the Earth ...” (United Nations Treaty Series 1967, Art. VII). The Registration Convention is relevant as well in that it could help identify the responsible owner of a piece of debris in the event of a collision (United States Treaty Series 1975, Art. II.1). There are, however, some limits to the utility of the Registration Convention when addressing space debris (Viikari 2015, pp. 737–39).

The Hague Building Blocks envision a binding international framework that addresses the environmental consequences of space mining with far more specificity than is set out under existing international law. Under the Building Blocks, the international framework should provide for the “[a]voidance and mitigation of potentially harmful impacts resulting from space resource activities” (Hague Working Group 2019, Sect. 10). Specifically, responsible nations and international organizations should be required to adopt measures to avoid and mitigate (Hague Working Group 2019, Sect. 10):

- (a) Risks to the safety of persons, the environment or property;
- (b) Damage to persons, the environment or property;
- (c) Adverse changes in the environment of the Earth, taking into account internationally agreed planetary protection policies;
- (d) Harmful contamination of celestial bodies, taking into account internationally agreed planetary protection policies;
- (e) Harmful contamination of outer space;
- (f) Harmful effects of the creation of space debris;
- (g) Harmful interference with other on-going space activities, including other space resource activities;
- (h) Changes to designated and internationally endorsed outer space natural or cultural heritage sites;
- (i) Adverse changes to designated and internationally endorsed outer space sites of scientific interest.

The Building Blocks expressly recognize space debris, and acknowledge the risks that can attend the accumulation of space debris (e.g., damage to persons, outer space and Earth environments, and property).

The Building Blocks also provide that the international framework should require nations and international organizations to implement a review and approval process designed to ensure that any activity avoids harmful impacts, and that the authorizing nation or international organization monitor and respond to any resultant harmful impact, including by considering whether the activity should be adjusted or terminated (Hague Working Group 2019, Sects. 11, 12).

To address the problem of space debris accumulation, the Scientific and Technical Sub-Committee of the UNCOPUOS has developed Space Debris Mitigation Guidelines (the COPUOS Guidelines) that are “the leading international arrangement to mitigate space debris” (Viikari 2015, p. 743 (internal quotation marks omitted)). The COPUOS Guidelines (United Nations Office for Outer Space Affairs 2010, Sect. 4)

consist of the following seven guidelines that are to be considered at all stages of the mission process:

1. Limit debris released during normal operations;
2. Minimize the potential for break-ups during operational phases;
3. Limit the probability of accidental collision in orbit;
4. Avoid intentional destruction and other harmful activities,
5. Minimize potential for post-mission break-ups resulting from stored energy;
6. Limit the long-term presence of spacecraft and launch vehicle orbital stages in the low-Earth orbit ... region after the end of their mission; and
7. Limit the long-term interference of spacecraft and launch vehicle orbital stages with the geosynchronous Earth orbit ... region after the end of their mission

Though non-binding, the COPUOS Guidelines direct member states and international organizations to voluntarily implement the Guidelines through domestic law mechanisms (United Nations Office for Outer Space Affairs 2010, Sect. 3). One scholar remarks that (Viikari 2015, p.743):

the fact that all major spacefaring states take part in the work of the [Scientific and Technical Sub-Committee] ... should facilitate the approval and implementation of the Guidelines on the national level.

Certain nations have also developed and implement debris mitigation practices (U.S. Government 2019).

Nuclear Contamination

Closely related to the issue of space debris are concerns about nuclear contamination. The primary dangers of nuclear contamination are that a spacecraft carrying a nuclear power source (NPS) may collide with another space object or debris, or that a spacecraft with an NPS on board may crash to Earth. This latter risk was highlighted when the Soviet satellite, Cosmos-954, which was carrying an NPS, crashed to Earth in Canada. This event is discussed in greater detail in Sect. 33.3.4. Concerns involving NPSs are not unique to space mining. And, it does not appear that space mining would necessarily increase the threat of nuclear contamination, except to the extent that space mining ventures may result in an increase of the presence of objects employing NPSs in space.

The Outer Space Treaty prohibits the placement into orbit of nuclear weapons, but it does not address the use of NPSs (United Nations Treaty Series 1967, Art. IV). However, the UN has adopted the Principles Relevant to the Use of Nuclear Power Sources in Outer Space (the Nuclear Power Source Principles), which set out, among other things, guidelines related to the safe use of NPSs in space, requirements for notifying other states of malfunctioning space objects that have NPSs on board that present a risk to Earth, and a framework for liability and compensation for damage caused by NPS-using space objects (United Nations General Assembly 1992, Principles 3, 5, 9). A notable limitation of the Nuclear Power Source Principles is that they apply only to NPSs used for purposes of electric power generation, and not to NPSs used for propulsion (United Nations General Assembly 1992):

Affirming that this set of Principles applies to nuclear power sources in outer space devoted to the generation of electric power on board space objects for non-propulsive purposes, which have characteristics generally comparable to those of systems used and missions performed at the time of the adoption of the Principles[.]

They are also non-binding (Viikari 2015, pp. 739–40).

The Hague Building Blocks do not directly address the use of NPS systems. But, Sect. 10 contains broad language that contemplates avoiding and mitigating the potential for harmful impacts associated with the use of NPSs. That section provides that (Hague Working Group 2019, Sect. 10):

the international framework should provide that States and international organizations responsible for space resource activities shall adopt appropriate measures with the aim of avoiding and mitigating potentially harmful impacts, including (a) Risks to the safety of persons, the environment or property; (b) Damage to persons, the environment or property.

Other Contamination

Other forms of contamination fall into two categories: forward contamination and back contamination (also called “backward contamination”). Forward contamination is “organic contamination of outer space” and “back contamination” refers to returning spacecraft bringing “unfamiliar contaminants” to Earth (Viikari 2015, p. 724).

A significant concern related to forward contamination is the possibility of compromising human understanding of outer-space environments. One scholar provides the following example to illustrate the concern (Roth 2016, pp. 865–66 (internal citations omitted)):

[I]n mid-November 2014, scientists at the [European Space Agency] announced that Philaehad discovered organic molecules on the surface of Comet 67P. [European Space Agency] researchers concluded that some of the molecules are of types never previously observed on a comet. Had a mining craft without proper sterilization protocols touched down on the comet, thereby contaminating the comet’s environment with organic material from Earth, the possibility of deriving scientific knowledge from the asteroid would have been forever lost.

As with all of the environmental concerns discussed herein, concerns with forward contamination are not unique to space mining. Indeed, the kind of contamination described above could be caused by any space activity that involves contact with celestial bodies, even research-oriented missions. That said, the probability of such contamination almost certainly increases with greater human presence in space.

As noted above, Article IX of the Outer Space Treaty requires that exploration of the Moon and other celestial bodies “avoid their harmful contamination ... and, where necessary, [States] shall adopt appropriate measures for this purpose” (United Nations Treaty Series 1967, Art. IX). The Building Blocks also call for the prevention of contamination. Specifically, they envision an international framework that requires states and international organizations to (Hague Working Group 2019, Sect. 10):

avoid[] and mitigate[] potentially harmful impacts, including ... [h]armful contamination of celestial bodies, taking into account internationally agreed planetary protection policies [and h]armful contamination of outer space

To guide compliance with the Outer Space Treaty, the Committee on Space Research has developed a Planetary Protection Policy (the COSPAR Planetary Protection Policy). The COSPAR Planetary Protection Policy is based on the following policy statement (Committee on Space Research Panel on Planetary Protection 2020, p.1):

The conduct of scientific investigations of possible extraterrestrial life forms, precursors, and remnants must not be jeopardized. In addition, the Earth must be protected from the potential hazard posed by extraterrestrial matter carried by a spacecraft returning from an interplanetary mission. Therefore, for certain space mission/target planet combinations, controls on contamination shall be imposed in accordance with issuances implementing this policy.

The COSPAR Planetary Protection Policy sets out five categories of missions, and recommends protective measures that increase in rigor based on the mission category. Categories I through IV are described based on (i) the degree to which the target body is of interest to understanding chemical evolution or the origin of life, and (ii) the likelihood of contamination (Committee on Space Research Panel on Planetary Protection 2020, pp. 1–2). Category V sets out requirements related to Earth-return missions (Committee on Space Research Panel on Planetary Protection 2020, pp. 2–3). Both NASA and the ESA implement the COSPAR Planetary Protection Policy (Office of Safety & Mission Assurance 2020; The European Space Agency 2020).

Damage to the Earth's Atmosphere

Another area of environmental concern related to increased space activities, including, but not limited to, space mining activities, is the disturbance of Earth's atmosphere by spacecraft launches. Rocket launches uniquely affect the stratosphere layer of Earth's atmosphere because they emit gases directly into the stratosphere (Ross and Vedda 2018, p. 3). The stratosphere lies directly above the troposphere, which is the layer of atmosphere closest to Earth's surface, and directly below the mesosphere (Ross and Vedda 2018, p. 4).

Regarding harm to the stratosphere (Ross and Vedda 2018, p. 5), researchers explain:

[t]he important emissions of concern with respect to global impacts are chlorine and alumina particles from solid rocket motors (SRMs) and soot particles (hereafter, black carbon or BC), mainly, though not exclusively, from kerosene fueled engines.

Black carbon particles that accumulate in the stratosphere form a “black umbrella” that warms the surrounding stratosphere and cools the Earth's surface by intercepting sunlight (Ross and Vedda 2018, p. 4). Alumina particles form a “white umbrella” that reflects sunlight back to space, further cooling the Earth's surface” (Ross and Vedda 2018, p. 4). At first blush, cooling may appear to have the beneficial impact of offsetting the rise in global temperatures due to climate change. But, the accumulation of black carbon and alumina particles in the stratosphere depletes the ozone layer of the atmosphere, and depletion of the ozone allows harmful ultraviolet radiation from the Sun to reach Earth's surface (Ross and Vedda 2018, p. 4; Shapiro 1995, p. 741). Scientists urge that more research is required to fully understand the impacts of rocket

emissions on Earth's atmosphere given the potentially serious consequences of those impacts (Ross and Vedda 2018, p. 9).

The Outer Space Treaty does not address harm to Earth's atmosphere caused by space activities that are initiated on Earth, and it would be difficult to leverage the general language in Article IX to require mitigation of rocket emissions because Article IX is focused on activities that take place in outer space (United Nations Treaty Series 1967, Art. IX). Because "outer space" is defined generically as the space beyond Earth's atmosphere, mitigating damage to the atmosphere appears to be outside the scope of the Outer Space Treaty. A commonly accepted definition of "space" holds that space begins at the Kármán Line, which is located 100 km or 62 miles above sea level (National Oceanic and Atmospheric Administration 2016). Notably, NASA and the United States military define space as beginning 12 miles below the Kármán Line (National Oceanic and Atmospheric Administration 2016).

Production and consumption of certain substances that damage the ozone are regulated at the international level by the Montreal Protocol on Substances That Deplete the Ozone Layer (the Montreal Protocol) (United Nations Treaty Series 1987). At present, the Montreal Protocol does not control emissions from rocket launches (World Meteorological Organization 2019). The 2018 Scientific Assessment of Ozone Depletion notes, rather, that "[r]ocket launches presently have a small effect on total stratospheric ozone (much less than 0.1%)" (World Meteorological Organization 2019, ES.50). However, it recognizes that "[s]pace industry developments indicate that rocket emissions may increase more significantly than reported in the previous Assessment" (World Meteorological Organization 2019, ES.50). Additionally, two researchers point out that, with the increasing frequency of rocket launches and international concern for global atmosphere, rocket launches could come under increased scrutiny under the Montreal Protocol in the future and space launch providers should plan for this possibility (Ross and Vedda 2018, p. 9). Regarding concerns related to ozone depletion caused, in part, by rocket launches in connection with space mining, certain countervailing considerations are worth mentioning. Specifically, if the Moon is used as a space base and refueling station—which would only be achievable by mining its resources—this could result in fewer rocket launches from Earth's surface, and, therefore, fewer emissions within the stratosphere.

33.3.3.2 Conservation and Preservation of Historic Sites

Related to mitigating the environmental issues that may attend space mining, the space mining industry may also eventually be affected by calls to withdraw some portion of the solar system from resource exploitation and to preserve certain sites of historical importance.

Recently, two researchers have argued that exploitation of space resources should be limited by a "one-eighth principle" (Elvis and Milligan 2019). Authors Martin Elvis and Tony Milligan describe the one-eighth principle as follows (2019, p. 575):

While economic growth remains exponential, we should regard as ours to use no more than one-eighth of the exploitable materials of the Solar System. And by “ours” we mean humanity’s as a whole, rather than any particular generation of humans or group of generations. The remaining seven-eighths of the exploitable Solar System should be left as space wilderness.

Elvis and Milligan explain that the primary problem that they seek to avoid is the depletion of the solar system’s resources (Elvis and Milligan 2019, p. 575). Accordingly, the wilderness designation would not necessarily prevent all forms of human impact, only human use (Elvis and Milligan 2019, p. 575).

Regarding historic site preservation, one group has appealed to the United Nations to declare that the Apollo 11 landing site and the Soviet Luna 2 spacecraft, which remains on the Moon’s surface 60 years after landing, deserve special recognition (Greenfieldboyce 2019). Relatedly, as a matter of voluntary practice, NASA has set out recommendations to protect United States government assets located on the Moon based on their historic and scientific value (National Aeronautics and Space Administration 2011).

An attempt by a nation to unilaterally protect areas of space might be inconsistent with the Outer Space Treaty which states that “there shall be free access to all areas of celestial bodies,” further suggesting that protection of certain celestial bodies in outer space would require international consensus (United Nations Treaty Series 1967, Arts. I, II).

The Hague Building Blocks contemplate the designation of internationally protected areas in space. Section 18 provides for (Hague Working Group 2019, Sect. 18.b):

[t]he establishment and maintenance of an international database ... for making publicly available ... iii. [t]he list of designated and internationally endorsed outer space natural and cultural heritage sites; and iv. [t]he list of designated and internationally endorsed sites of scientific interest ...

That section further contemplates the establishment of an international body charged with listing such sites (Hague Working Group 2019, Sect. 18.c.ii). The international framework envisioned under the Building Blocks would also require responsible states and international organizations to adopt measures to avoid and mitigate harm to these sites (Hague Working Group 2019, Sect. 10).

Efforts to protect portions of space are in some ways analogous to efforts to protect portions of Earth’s oceans that are not subject to any national jurisdiction. Over the past few years, UN delegates have been negotiating a new legally binding instrument to protect marine life in international waters (Heffernan 2018). The instrument would likely function as an extension of the UNCLOS (Heffernan 2018). The negotiations that were scheduled to continue during spring 2020 have been postponed due to COVID-19 (United Nations 2020). But, the terms of any successful binding instrument to ultimately come out of the negotiations will likely inform outer-space conservation efforts going forward.

Protections for historical and cultural sites located in international waters could also serve as an analogous legal framework to protect such sites in space. Both

the UNCLOS and the 2001 Convention for the Protection of Underwater Cultural Heritage (the Underwater Cultural Heritage Convention) include protections for sites of archaeological and historical value located in the Area, which, as noted above, is defined as “the seabed and ocean floor and subsoil thereof, beyond the limits of national jurisdiction” (United Nations Treaty Series 1982., Art. 1(1)). The UNCLOS states that (United Nations Treaty Series 1982, Art. 149):

[a]ll objects of an archaeological and historical nature found in the Area shall be preserved or disposed of for the benefit of mankind as a whole, particular regard being paid to the preferential rights of the State or country of origin, or the State of cultural origin, or the State of historical and archaeological origin.

Similarly, the Underwater Cultural Heritage Convention protects “traces of human existence having a cultural, historical or archaeological character which have been partially or totally under water, periodically or continuously, for at least 100 years,” including such traces that are located in the Area (United Nations Treaty Series 2001, Arts. 11, 12).

33.3.4 Liability for Damage

The recovery of resources in outer space will be a complex and difficult industrial process. There will be a risk of injury to people and property. This section discusses the allocation of liability for other accidents and incidents arising during the exploration for and exploitation of natural resources in outer space.

Article 6 of the Outer Space Treaty provides that state parties “bear international responsibility for national activities in outer space” (United Nations Treaty Series 1967, Art. VI). Further, all activities conducted in outer space, including activities by private parties, must take place under the authorization and supervision of a state party to the Outer Space Treaty (United Nations Treaty Series 1967, Art. VI). If the activity is conducted by an international organization, then the responsibility lies with the international organization and its constituent state parties (United Nations Treaty Series 1967, Art. VI). This liability is not limited to space junk that might fall to Earth—it extends to damage in the air, in outer space, or on any celestial body, including the Moon.

Article 7 of the Outer Space Treaty also provides for direct liability for the results of the launch of an object into outer space. If a government either launches an object itself, or procures the launch of an object (perhaps from a private sector contractor, for example), then the state party from which the object is launched is “internationally liable” for damage caused by that object to another government that is a party to the Outer Space Treaty, or the citizens or businesses of the injured government (United Nations Treaty Series 1967, Art. VII). If a bit of space junk crashes into Tanzania or strikes a Tanzanian aircraft, however, the Republic of Tanzania or its nationals cannot bring a claim against the launching state under the Outer Space Treaty, because

Tanzania is not a “State Party to the Treaty” (United Nations Treaty Series 1967, Art. VI).

The Outer Space Treaty, then, provides only broad principles related to the allocation of liability. These broad principles have been developed further in the Convention on International Liability for Damage Caused by Space Objects (Liability Convention) (United Nations Treaty Series 1972), the Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space (Rescue Agreement) (United Nations Treaty Series 1968), and the Convention on Registration of Objects Launched into Outer Space (Registration Agreement) (United Nations Treaty Series 1975).

The Liability Convention was created and opened for signature in 1972. It creates a liability framework for damage caused by spacecraft. It establishes a strict liability standard for accidents on the Earth’s surface and a negligence standard for accidents elsewhere (Roth 2016). The Liability Convention addresses the liability of “launching States,” that is, a state that launches or procures the launch of a space object, or the state from whose territory or facility a space object is launched (United Nations Treaty Series 1972, Art. 1). Article 2 of the Liability Convention provides that a launching state is “absolutely liable” for damage caused to the surface of the Earth or an aircraft in flight (United Nations Treaty Series 1972, Art. 2). The launching state may be relieved of its absolute liability where the claiming state (or those claiming under its jurisdiction) acted with gross negligence, or with the intent to cause damage. To assert this defense, the launching state must have acted in accordance with international law (United Nations Treaty Series 1972, Art. 6). Where the damage is caused “elsewhere than on the surface of the Earth,” then the launching state is liable only if the damage arises from the fault of the launching state or the fault of persons for whom the launching state is responsible (United Nations Treaty Series 1972, Art. 3).

The Liability Convention has a fairly complex approach to joint liability. If a state (State A) causes damage in outer space to another state (State B), and that incident causes collateral damage to a third state (State C), then State A and State B are jointly and severally liable to State C. If the damage occurs on the surface of the Earth or to an aircraft in flight, then the liability of State A and State B is absolute; if the damage occurs in outer space, then liability is based on the fault of the first two states (United Nations Treaty Series 1972, Art. 4). As between the two states that are jointly liable, their liability will be apportioned based on their fault, but the injured third state (State C) may choose to recover compensation from only one of the states responsible (United Nations Treaty Series 1972, Art. 4).

Given the potential of claims directly against the state, the FCC is considering requiring parties securing a license for a satellite to indemnify the United States government from potential claims under the Outer Space Treaty and the Liability Convention. Noting the provisions of those treaties that allow for claims directly against the state, the FCC states: “Under these circumstances, conditioning Commission authorization on indemnification of the U.S. government may be a reasonable step, given the absence of protections under international law of the protection from liability under U.S. law related to a licensing authority’s exercise of its discretionary functions” (Federal Communications Commission Report 2020a, b).

To date, the only settlement ever reached pursuant to the Liability Convention was the 1981 resolution of a dispute between the former Soviet Union and Canada regarding a Soviet nuclear-powered satellite that fell back to Earth over Canada, scattering radioactive debris in the Northwest Territories, Alberta, and Saskatchewan (Government of Canada and Government of the Union of Soviet Socialist Republics 1981, para. 2). Prior to the satellite's re-entry, the Soviet Union failed to forewarn Canada of the possibility that the satellite might fall onto its lands (Government of Canada and Government of the Union of Soviet Socialist Republics 1981, para. 4). The Soviet Union also failed to answer questions from Canada regarding the nature of the nuclear reactor that had been onboard the satellite, stating instead that the nuclear reactor had been designed to break up upon re-entry (Government of Canada and Government of the Union of Soviet Socialist Republics, para. 5). This complicated the Canadian clean-up and recovery effort, as the Canadians had incomplete information regarding the type of hazards they were facing. It was eventually determined that nearly all of the pieces of debris were radioactive, some at levels high enough to prove fatal (Government of Canada and Government of the Union of Soviet Socialist Republics, para. 10).

The satellite recovery efforts lasted ten months and cost Canada \$13,970,143.66 (Government of Canada and Government of the Union of Soviet Socialist Republics, para. 8). Of this amount, Canada submitted a claim to the USSR, pursuant to Article VIII of the Convention, for \$6,041,174.70, and the two countries ultimately settled the claim for \$3,000,000 (Government of Canada and Government of the Union of Soviet Socialist Republics, Art. II, para. 8).

The Rescue Agreement provides a framework for the return of objects and people who land outside their national territory upon re-entry to Earth (United Nations Treaty Series 1968). Under the Rescue Agreement, a contracting party agrees to provide notice to the launching authority and the UN Secretary General when that contracting party learns that the personnel of a spacecraft have had an accident, or are in distress, or have had to make an unscheduled or emergency landing (United Nations Treaty Series 1968, Art. 1). The contracting party provides this notice when it makes this discovery either within the territorial jurisdiction of that country, or where no country has territorial jurisdiction, such as on the high seas (United Nations Treaty Series 1968, Art. 1). If the launching authority is not clear, then the contracting party will make a public announcement "by all appropriate means at its disposal" (United Nations Treaty Series 1968, Art. 1). All other contracting parties are to lend assistance where the rescue arises outside the territorial jurisdiction of any country (United Nations Treaty Series 1968, Art. 3). Also, the contracting party that affects a rescue within its jurisdiction or outside any territorial jurisdiction must return the rescued personnel to the launching authority. For accidents within the area of a contracting party's territorial jurisdiction, that contracting party "shall immediately take all possible steps to rescue them and render them all necessary assistance" (United Nations Treaty Series 1968, Art. 2). The launching authority is required to provide assistance if doing so would be helpful (United Nations Treaty Series 1968, Art. 2).

The Registration Agreement requires signatories to register vehicles launched into space with a UN-operated registry (United Nations Treaty Series 1975, Art. 6). The registry can be used to identify the state or states that launched a space object, which would of course be relevant to the determination of liability under the Outer Space Treaty and the Liability Convention, as well as obligations under the Rescue Agreement (United Nations Treaty Series 1975, Art. 6).

33.3.5 *Dispute Resolution*

As in any human endeavor, there is a possibility that a disagreement may arise between states or individuals seeking to develop resources in outer space.

For example, a private company, operating under the authority and supervision of a country, may set up a mining operation on the Moon, extracting ice, using it to provide potable water and converting it to oxygen and hydrogen for human use and for fuel. What happens if another private company operating under the flag of a different country, comes along and starts mining in the same crater, and perhaps in the same seam of ice? How would a dispute between the two mining companies get resolved?

Or perhaps a private company surveys an asteroid and determines that it is mineral rich, and that the minerals in that asteroid can be extracted and used commercially. Then that same company comes back to that asteroid and does some sampling and analysis and testing, and develops a plan for the exploitation of that near-Earth object. Does that company have a claim if another private space mining company comes along and starts operating on that asteroid? How would that claim get resolved?

Finally, imagine that a company successfully mines an asteroid and is able to set up a commercial operation in near-Earth orbit. Maybe the company is manufacturing vehicle parts from metals in the asteroid, or is exploiting ice in an asteroid to create a refueling station in outer space. This commercial venture renders not just revenue but profit. As a company formed in Luxembourg, it takes advantage of Luxembourg law and retains that profit for itself. A developing country might bring a claim under the Outer Space Treaty or the Moon Treaty, and argue that the company owes a royalty to all mankind. How will this claim be adjudicated?

As discussed above, the Outer Space Treaty makes activities in outer space subject to international law, including the UN Charter (United Nations Treaty Series 1967, Art. 3). The Outer Space Treaty, however, does not include a dispute resolution process mechanism for such activities. The only treaty that even contemplates a tribunal to resolve disputes is the Liability Convention (United Nations Treaty Series 1972). As discussed in detail below, the Liability Convention contemplates the creation of a Claims Commission under limited circumstances, and such a commission has never been formed (Williams 2017, p. 995). Article IX of the Outer Space Treaty includes the requirement that activities in space avoid “harmful contamination and also adverse changes in the environment of the Earth resulting from the introduction of extraterrestrial matter” (United Nations Treaty Series 1967, Art. 9). Even this

regulatory requirement, however, does not include an enforcement mechanism, but rather requires “consultation” should there be some concern about a violation of this requirement. In the negotiation of the Outer Space Treaty, the United States wanted disputes in outer space to be resolved by the International Court of Justice, but the Soviet Union preferred to have disputes addressed through direct negotiations (Williams 2017, p. 999).

The UN Charter contains a broad request to states to resolve disputes that might threaten international security to “seek a solution by negotiation, enquiry, mediation, conciliation, arbitration, judicial settlement, resort to regional agencies or arrangements, or other peaceful means of their own choice” (United Nations Treaty Series 1945, Art. 33). The UN might ask countries to adopt one of these approaches if they do not do so willingly. The United Nations Security Council can also investigate such a dispute and recommend further action (United Nations Treaty Series, Arts. 34–38). Fundamentally, however, dispute resolution remains at the discretion of state actors.

The Liability Convention does take an additional step, and contemplates the formation of a Claims Commission to resolve claims of injury. But even the Claims Commission remains subject to the choice of the states involved in the dispute.

The Liability Convention creates a liability framework for damage caused by spacecraft. As noted above, it establishes a strict liability standard for accidents on the Earth’s surface and a negligence standard for accidents elsewhere (United Nations Treaty Series 1972, Arts. II, III). If a state seeks compensation under the Liability Convention, it presents its claim through diplomatic channels, or if the states do not have diplomatic relations, then through the Secretary General of the United Nations. If the claim has not been resolved after one year of diplomatic negotiations, then the Convention requires the parties to form a Claims Commission if either party requests a Claims Commission (United Nations Treaty Series 1972, Art. XIV):

If no settlement of a claim is arrived at through diplomatic negotiations as provided for in Article IX, within one year from the date on which the claimant State notifies the launching State that it has sub-mitted the documentation of its claim, the parties concerned shall establish a Claims Commission at the request of either party.

The Claims Commission proceeds in a manner similar to a three-member arbitration panel (United Nations Treaty Series 1972, Arts. XV–XVIII). The decision of the Claims Commission is binding only if both parties agree to be bound. Otherwise, the Claims Commission decision is advisory (United Nations Treaty Series 1972, Art. XIX).

The Hague Working Group Building Blocks set out a regulatory framework for the development of resources in space. The Building Blocks are based on international law, and, as noted, seek to “create an enabling environment for space resource activities that takes into account all interests and benefits all countries and humankind” (Hague Working Group 2019, Sect. 1.1). The Building Blocks, consistent with existing treaties, retain a fairly open-ended approach to dispute resolution (Hague Working Group 2019, Sect. 19):

The international framework should encourage recourse by States, international organizations and operators to the resolution of disputes through adjudicatory, non-adjudicatory or

hybrid mechanisms, for example by developing procedures for consultation or promoting the use of the 2011 Permanent Court of Arbitration Optional Rules for Arbitration of Disputes Relating to Outer Space Activities.

As with existing treaties, the Building Blocks leave the dispute resolution process at the discretion of state actors. Note that the Working Group does not recommend mandatory arbitration, but rather recommends “promoting” the Permanent Court of Arbitration Optional Rules for Arbitration of Disputes Relating to Outer Space Activities (the Optional Rules) (Hague Working Group 2019, Sect. 19). The Optional Rules are based on the 2010 UNCITRAL Arbitration Rules, amended to reflect the particular characteristics of disputes involving the use of outer space by states, international organizations and private entities; and to reflect the public international law applicable to the use of outer space (Permanent Court of Arbitration 2011, p. 4). The Optional Rules are not dramatically different from the UNCITRAL Rules, and are designed to provide considerable flexibility to promote the adoption and use of the rules (Kilgore 2018, p. 59). To date, the Optional Rules have not been used to adjudicate a space dispute (Kilgore 2018, p. 59).

One advantage of the Optional Rules is that they are intended to allow the use of the rules by private parties. Many of the treaties applicable to resource development in outer space are focused on disputes between sovereign states. While it is true that any activity in outer space must be conducted under the jurisdiction and supervision of a nation (United Nations Treaty Series 1967, Art. VII), it is possible, even likely, that disputes will arise between private parties conducting resource development on a moon, planet or asteroid (Permanent Court of Arbitration 2011, p. 4).

When thinking about potential disputes related to resource development in outer space, one fundamental issue will concern whether the Outer Space Treaty allows private parties to retain the minerals extracted from the Moon or an asteroid, and whether those parties will be required to share some of the value of their enterprise, as is required under the UNCLOS. The United States and Luxembourg have passed laws that allow companies operating under their jurisdiction to retain the benefits of resource development. There may be a legal challenge to those laws under the Outer Space Treaty once private companies start operating under that legal framework (Tronchetti 2017). Given that the Outer Space Treaty keeps the dispute resolution process open-ended, that claim could be brought in a local courthouse in any country, and be subject to all the related jurisdictional challenges that come with that course of action. The aggrieved country could bring its claim in the United States or Luxembourg. Finally, the country bringing the challenge could try to have the claim adjudicated in an international forum or through arbitration, but could only use these avenues with the consent of the United States or Luxembourg, as the case may be.

This dispute is unlikely to fall within the scope of the Vienna Convention on the Law of Treaties, which was adopted two years after the Outer Space Treaty (United Nations Treaty Series 1969). Luxembourg and the United States both signed the Outer Space Treaty in 1967. The Vienna Convention is not retroactive (United Nations Treaty Series 1969, Art. 4). Under the Vienna Convention, a state could

present the interpretation of a treaty to the International Court of Justice (United Nations Treaty Series 1969, Art. 66).

It will be even more difficult for private companies to have disputes resolved in a friendly forum. If a mining company operating under the Luxembourg statute sets up shop in Shackleton Crater and starts mining ice on the Moon, what recourse will it have if a company operating under the flag of another country jumps its claim? At present, the most likely venue for resolution of the claim will be in the courts of the country of origin of the claim jumper. It would be hard to show the necessary nexus to Luxembourg to be able to bring that claim in the courts of Luxembourg. And there is no binding venue for resolution of disputes under the Outer Space Treaty. The Moon Treaty allows for consultation if a state believes that another state is not acting in a manner consistent with the provisions of the Treaty, with the possibility of referral to the Secretary General of the UN for assistance (United Nations Treaty Series 1979, Art. 15). There is no binding dispute resolution mechanism under the Moon Treaty. The Hague Building Blocks recommend creating a registration system to acknowledge priority rights to develop resources, but no such registration regime exists at present (Hague Working Group 2019, Sect. 14). Also, as noted above, the Building Blocks do not recommend developing a binding dispute resolution process, but rather suggests that a space resource development regime should encourage state actors and private parties to use the Optional Rules. While there are knotty jurisdictional issues to be resolved when addressing the enforcement of a contract or mining right arising in outer space, there may be even more interesting questions to be considered in asserting jurisdiction over a criminal matter (Kean 2020).

33.4 Conclusion

By its very nature, the field of space resource extraction moves at a fast pace, and law and public policy have to date struggled to keep up. The Outer Space Treaty entered into more than 50 years ago provided the initial framework for answering some of the pervasive questions faced by would-be space miners, but left other critical questions—for example, “Am I entitled to own the resources I extract?”—without a clear answer. And although the Moon Treaty attempted to resolve these issues in favor of an international benefit-sharing regime, it was—likely for that very reason—never signed by the major space-faring nations.

The United States has recently, once again, rejected the Moon Treaty—both implicitly through its adoption of a domestic law that recognizes private ownership in resources extracted from outer space, and explicitly through an Executive Order recently issued by President Trump. Rather than seek a new international treaty to address these issues, the United States has chosen to bypass the long and arduous treaty process in favor of entering into the Artemis Accords with a chosen selection of key space allies. The Artemis Accords represent the United States’ attempt to form international consensus around its own view of the laws of space resource extraction,

in the hope that enough countries will accept the Artemis Accords that it will come to be viewed as customary international law.

As a result, it is possible that the Artemis Accords might one day be looked upon as the tipping point that provided the requisite level of certainty to draw hesitant industry participants and investors fully into the field. But even these new domestic laws and multilateral agreements leave certain questions unanswered, for example in the areas of dispute resolution between entities acting in outer space, with respect to the preservation of historic sites in space, and regarding the potential establishment of a mining claims registration system. As nations continue to advance private commercial activities in space, space mining law and policy are likely to continue to evolve to address these and other issues.

References

- 114th Congress of the United States. 2016. American Space Renaissance Act, H.R. Res. 4945
- Araxia Abrahamian, A. 2017. How a tax haven is leading the race to privatise space. *The Guardian*. <https://www.theguardian.com/news/2017/sep/15/luxembourg-tax-haven-privatise-space>.
- Calandrelli, E. 2016. Deep Space Industries partners with Luxembourg to test asteroid mining technologies. TechCrunch. <https://techcrunch.com/2016/05/05/deep-space-industries-partners-with-luxembourg-to-test-asteroid-mining-technologies/>.
- Canadian Reporting Standards. 2014. Canadian Institute of Mining, Metallurgy and Petroleum, Definition Standards for Mineral Resources and Mineral Reserves. https://mrmr.cim.org/media/1068/cim_definition_standards_2014.pdf.
- Charlesworth, H. 1998. The unbearable lightness of customary international law. In *Proceedings of the ASIL annual meeting*, vol. 92, 44. Washington, DC.
- Chatzipanagiotis, M. 2016. 3D printing using material from celestial bodies—a method to circumvent the non-appropriation principle? In *Proceedings of the international institute of space law*, eds. Blount, P.J., Masson-Zwaan, T., Moro-Aguilar, R., and Schrogl, K. The Hague, Netherlands: Eleven Int'l.
- Cheng, B. 1997. *Studies in international space law*. Oxford, UK: Clarendon Press.
- Committee on Space Research Panel on Planetary Protection. 2020. COSPAR Policy on Planetary Protection. *Space Research Today*, No. 208, 10–22
- Convention Relating to the Regulation of Aerial Navigation. 1930. *J. Air L. & Com* 1:94, Evanston, IL. <https://scholar.smu.edu/jalc/vol1/iss1/6>.
- Dodge, M. 2016. The U.S. Commercial Space Launch Competitiveness Act of 2015: Moving U.S. Space Activities Forward, 29 No. 3. *Air & Space Law*. Wolters Kluwer. Alphen aan den Rijn, Netherlands.
- Egan, B. 2016. The next fifty years of the outer space treaty. In *Galloway symposium on critical issues in space law*. Washington, DC. <https://2009-2017.state.gov/s/l/releases/remarks/264963.htm>.
- Elvis, M., and T. Milligan. 2019. How much of the Solar System should we leave as wilderness? *Acta Astronautica* 162:574–580.
- Emerging Technology from the arXiv. 2018. Asteroid mining might actually be better for the environment. *MIT Technology Review*. <https://www.technologyreview.com/2018/10/19/139664/asteroid-mining-might-actually-be-better-for-the-environment/>.
- The European Space Agency (last visited 2020). 2020. Planetary protection. The European Space Agency. https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Exploration/ExoMars/Planetary_protection.

- Federal Aviation Administration. 2020. Streamlined Launch and Reentry License Requirements (Final Rule). https://www.faa.gov/space/streamlined_licensing_process/media/SLR2_Final_Rule_450.pdf.
- Federal Communications Commission. 2020a. 47 C.F.R. § 25.114.
- Federal Communications Commission. 2020b. Report and Order and Further Notice of Proposed Rulemaking, In the Matter of Mitigation of Orbital Debris in the New Space Age. IB Docket No. 18-313, FCC 20-54
- Foust, J. 2018. House passes commercial space regulatory bill. SpaceNews. <http://spacenews.com/house-passes-commercial-space-regulatory-bill/>.
- Foust, J. 2019. Luxembourg expands its space resources vision. SpaceNews. <https://spacenews.com/luxembourg-expands-its-space-resources-vision/>.
- Foust, J. 2020. Luxembourg establishes space industry venture fund. SpaceNews. <https://spacenews.com/luxembourg-establishes-space-industry-venture-fund/>.
- Force, M.K. 2016. The paradox of United States' position on regulation of space resource extraction. In *Proceedings of the international institute of space law*, eds. Blount, P.J., Masson-Zwaan, T., Moro-Aguilar, R., and Schrogl, K. The Hague, Netherlands: Eleven Int'l.
- Fowler, R. 2011. Cracks in the ice: the need for review of the legal status of the arctic continental shelf. In *International mining and oil and gas law, development, and investment*. Rocky Mountain Mineral Law Foundation Special Institute.
- Testimony of Joanne Irene Gabrynowicz Before the Subcommittee on Space of the Committee on Science, Space, and Technology. 2014. United States House of Representatives. <https://science.house.gov/sites/republicans.science.house.gov/files/documents/Gabrynowicz%20Final%20Testimony%20H.R.%205063.pdf>.
- Gilbert, A., and M. Bazilian. 2020. The Era of Commercial Space Mining Begins. Payne Institute Commentary Series 9/23/2020. <https://payneinstitute.mines.edu/the-era-of-commercial-space-mining-begins/>.
- Girones, E.O., A. Pugachevsky, and G. Walser. 2009. *Mineral rights cadastre: promoting transparent access to mineral rights*. World Bank Extractive Industries for Development Series #4, Washington, DC.
- Testimony of Mike Gold Before the Subcommittee on Space, Science, and Competitiveness of the Committee on Science, Space, and Technology. 2017. United States Senate. <https://www.hsdl.org/?view&did=807259>.
- Government of Canada, Government of the Union of Soviet Socialist Republics. 1981. Cosmos 954 Satellite Claims Resolution. Space Law. https://www.jaxa.jp/library/space_law/chapter_3/3-2-2-1_e.html.
- Government of the United Arab Emirates. 2019. Federal Law No. 12 of 2019 on the Regulation of the Space Sector. <https://www.moj.gov.ae/assets/2020/Federal%20Law%20No%2012%20of%202019%20on%20THE%20REGULATION%20OF%20THE%20SPACE%20SECTOR.pdf.aspx>.
- Greenfieldboyce, N. 2019. How Do You Preserve History On The Moon? National Public Radio. <https://www.npr.org/2019/02/21/696129505/how-do-you-preserve-history-on-the-moon>.
- Grush, L. 2018. House bill would regulate bold commercial space missions—but not very closely. The Verge. <https://www.theverge.com/2018/4/24/17272338/hr-2809-american-space-commerce-free-enterprise-act-regulation>.
- The Hague International Space Resources Governance Working Group. 2019. Building Blocks for the Development of an International Framework on Space Resource Activities. <https://www.universiteitleiden.nl/binaries/content/assets/rechtsgeleerdheid/instituut-voor-publiekrecht/lucht-en-ruimterecht/space-resources/bb-thissrwg-cover.pdf>.
- Heffernan, O. 2018. U.N. Makes a Bold Move to Protect Marine Life on the High Seas. *Scientific American*. <https://www.scientificamerican.com/article/u-n-makes-a-bold-move-to-protect-marine-life-on-the-high-seas/>.
- Hertzfeld, H.R., and F.G. von der Dunk. 2005. Bringing space law into the commercial world: property rights without sovereignty. *Chi. J. Int'l L.* 6:81. Chicago, IL.

- Letter from John Holdren, Dir. & Asst. to President for Science and Technology to Sen. Thune and Rep. Smith (2016). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/csla_report_4-4-16_final.pdf.
- International Institute of Space Law. 2015. *Position paper on space resource mining*. Paris: France.
- Johnson, M., M. Brady, and A. Duffy. 2020. Space mining: the Australian perspective. Hogan Lovells US LLP. https://www.hoganlovells.com/~media/hogan-lovells/pdf/2020-pdfs/2020_05_28_space_global_mining_newsletter_australia.pdf?la=en.
- Kean, S. 2020. How not to deal with murder in space. Slate. <https://slate.com/technology/2020/07/arctic-t3-murder-space.html>.
- Kettle, S. 2017. Planetary coordinate systems: the moon (Blog Post). ESRI. <https://community.esri.com/groups/coordinate-reference-systems/blog/2017/10/16/planetary-coordinate-reference-systems-the-moon>.
- Kilgore, S. 2018. Arbitration rules for disputes arising from outer space activity. *The Federal Lawyer*, 58–63.
- Kundmueller, M.M. 2002. The application of customary international law in U.S. courts: custom, convention, or pseudo-legislation?. *J. Legis.* 28:359, Notre Dame, IN.
- Lachs, M. 1964. Recueil des cours, vol. III. The International Law of Outer Space. Académie De Droit International.
- Law of July 20th 2017 on the Exploration and Use of Space Resources. 2017. The Government of the Grand Duchy of Luxembourg. https://space-agency.public.lu/en/agency/legal-framework/law_space_resources_english_translation.html.
- Lunar Geodesy and Cartography Working Group. 2008. A Standardized Lunar Coordinate System for the Lunar Reconnaissance Orbiter and Lunar Datasets, LRO and LGCWG White Paper, Version 5. National Aeronautics and Space Administration, Goddard Space Flight Center. <https://lunar.gsfc.nasa.gov/library/LunCoordWhitePaper-10-08.pdf>.
- Marboe, I. 2016. The end of the concept of “Common Heritage of Mankind”? The view of State Parties to the Moon Agreement. In *Proceedings of the international institute of space law*, eds. Blount, P.J., Masson-Zwaan, T., Moro-Aguilar, R., and Schrogl, K. The Hague, Netherlands: Eleven Int’l.
- Marboe, I., and K. Traunmuller. 2013. *The legal framework of the use of outer space technologies: an introduction*. NPOC Space Law Austria. Vienna, Austria.
- Masson-Zwaan, T., R. Lefeber, G. Reibaldi, and M. Stewart. 2016. The Hague space resources governance working group—a progress report. In *Proceedings of the international institute of space law*, eds. Blount, P.J., Masson-Zwaan, T., Moro-Aguilar, R., and Schrogl, K. The Hague, Netherlands: Eleven Int’l.
- Messer, S. 2015. Contracting for a feasibility study. In *Mining agreements: contracting for goods & services*. Rocky Mountain Mineral Law Foundation Special Institute, Vancouver.
- Ministry of the Economy. 2017a. Luxembourg and the United Arab Emirates to cooperate on space activities with particular focus on the exploration and utilization of space resources. The Government of the Grand Duchy of Luxembourg. <https://space-agency.public.lu/dam-assets/press-release/2017a/2017a-10-10-press-release-mou-space.pdf>.
- Ministry of the Economy. 2017b. Luxembourg is the first European nation to offer a legal framework for space resources utilization. The Government of the Grand Duchy of Luxembourg. <http://www.spaceref.com/news/viewpr.html?pid=51198>.
- Ministry of the Economy. 2018. Luxembourg launches business-focused national space agency. The Government of the Grand Duchy of Luxembourg. <https://space-agency.public.lu/dam-assets/press-release/2018/2018-09-12-Press-release-Launch-Lux-Space-Agency.pdf>.
- Nanda, V., et al. 2020. Sources of customary international law. *Litigation of International Disputes in U.S. Courts* 2, § 9:2.
- Nandan, S.N., and S. Rosenne, eds. 1995. *United nations convention on the law of the sea 1982: a commentary*, vol. III. The Hague, Netherlands, London, UK, Boston, MA: Martinus Nijhoff Publishers.

- National Aeronautics and Space Administration. 2011. Recommendations to Space-Faring Entities: How to Protect and Preserve the History and Scientific Value of U.S. Government Lunar Artifacts. National Aeronautics and Space Administration. https://www.nasa.gov/sites/default/files/617743main_NASA-USG_LUNAR_HISTORIC_SITES_RevA-508.pdf.
- National Aeronautics and Space Administration. 2020. The Artemis Accords: Principles for Cooperation in the Civil Exploration and Use of the Moon, Mars, Comets, and Asteroids for Peaceful Purposes. <https://www.nasa.gov/specials/artemis-accords/img/Artemis-Accords-signed-13Oct2020.pdf>.
- National Oceanic and Atmospheric Administration. 2016. Where is space? National Oceanic and Atmospheric Administration. <https://www.nesdis.noaa.gov/content/where-space>.
- Noyes, J.E. 2011. The common heritage of mankind: past, present, and future. *Den. J. Int'l L. Pol'y* 40:447–471. Denver, CO.
- O'Callaghan, J. (last visited 2020). 2020. What is space junk and why is it a problem? Natural History Museum. <https://www.nhm.ac.uk/discover/what-is-space-junk-and-why-is-it-a-problem.html>.
- Oduntan, G. 2003. The never ending dispute: legal theories on the spatial demarcation boundary plane between airspace and outer space. *Hertfordshire Law Journal* 1(2).
- Office of the President of the United States. 2020. Encouraging International Support for the Recovery and Use of Space Resource, Exec. Order No. 13,914, 85 Fed. Reg. 20,381.
- Office of Safety & Mission Assurance (last visited 2020). 2020. Planetary Protection. National Aeronautics and Space Administration. <https://sma.nasa.gov/sma-disciplines/planetary-protection>.
- Panesar, S. 2000. Theories of private property in modern property law. *Denning Law Journal* 15:113–138, Buckingham, UK.
- Permanent Court of Arbitration. 2011. Optional Rules for Arbitration of Disputes Relating to Outer Space Activities. The Hague, Netherlands. <https://docs.pca-cpa.org/2016/01/Permanent-Court-of-Arbitration-Optional-Rules-for-Arbitration-of-Disputes-Relating-to-Outer-Space-Activities.pdf>.
- Planetary Resources. 2015. President Obama Signs Bill Recognizing Asteroid Resource Property Rights Into Law. <http://www.planetaryresources.com/2015/11/president-obama-signs-bill-recognizing-asteroid-resource-property-rights-into-law/>.
- Pop, V. 2016. Is outer space proper the 'common heritage of mankind'? In *Proceedings of the international institute of space law*, eds. Blount, P.J., Masson-Zwaan, T., Moro-Aguilar, R., and Schrogl, K, 239–246. The Hague, Netherlands: Eleven Int'l.
- Quinn, A. 2008. The new age of space law: the outer space treaty and the weaponization of space. *Minn. J. Int'l L.* 17:475, Minneapolis, MN.
- Statement of Robert (Bob) Richards Before the United States Senate Committee on Commerce, Science, and Transportation Subcommittee on Science, Space, and Competitiveness. 2017. United States Senate. https://www.commerce.senate.gov/public/_cache/files/fe68e195-f1ac-43a5-9ae8-10a3435749df/0F70D3F36637F5FA4283870BCD0CDE87.dr.-bob-richards-testimony.pdf.
- Rocky Mountain Mineral Law Foundation. 2020. *American law of mining*, 2d Ed. (LexisNexis Matthew Bender).
- Ross, M., and J. Vedda .2018. The Policy and Science of Rocket Emissions. Center for Space Policy and Strategy. https://aerospace.org/sites/default/files/2018-05/RocketEmissions_0.pdf.
- Roth, S. 2016. Developing a law of asteroids: constants, variables, and alternatives. *Colum. J. Transnat'l L.* 54:827, New York, NY.
- Schaefer, M. 2017. The contours of permissionless innovation in the outer space domain. *U. Pa. J. Int'l L.* 39:103, Philadelphia, PA.
- Schmitt, H. 2006. *Return to the moon: exploration, enterprise, and energy in the human settlement of space*. New York: Praxis Publishing.
- Scoles, S. 2017. Luxembourg's Bid to Become the Silicon Valley of Space Mining. *Wired*. <https://www.wired.com/2017/01/luxembourg-setting-silicon-valley-space-mining/>.
- Scoville, R.M. 2016. Finding customary international law. *Iowa L. Rev.* 101:1893, Iowa City, IA.

- Shapiro, L. 1995. The need for international agreements concerning the ozone depleting effects of chemical rocket propulsion. *S. Cal. Interdisc. L.J.* 4:739–774, Los Angeles, CA.
- Simberg, R. 2012. Homesteading the Final Frontier: A Practical Proposal for Securing Property Rights in Space. Competitive Enterprise Institute, Issue Analysis.
- St. John, D. 2013. The Bogotá declaration and the curious case of geostationary orbit. *Denver Journal of International Law and Policy*. <http://djilp.org/3494/the-bogota-declaration-and-the-curious-case-of-geostationary-orbit/>.
- Stimers, P., and A. Dinegar. 2020. The Artemis Accords: a shared framework for space exploration. *The Space Review*. <https://www.thespacereview.com/article/4019/1>.
- Strickland, A. 2020. Eight nations sign NASA's Artemis Accords that guide cooperative exploration of the moon. CNN. <https://www.cnn.com/2020/10/14/world/artemis-accords-nations-sign-nasa-scn-trnd/index.html>.
- Swainea, E.T. 2015. Identifying customary international law: first thoughts on the ILC's first steps. *Am. J. Intl. L. Unbound* 108:184.
- Taylor, K. 2019. Fictions of the final frontier: why the United States SPACE act of 2015 is illegal. *Emory Int'l L. Rev.* 33:653, Atlanta, GA.
- Tennen, L.I. 2010. Towards a new regime for exploitation of outer space mineral resources. *Neb. L. Rev.* 88:794, Lincoln, NE.
- Tronchetti, F. 2017. Legal aspects of space resource utilization. In *Handbook of space law*, eds. F.G. von der Dunk, and Tronchetti, F. Cheltenham, UK: Edward Elgar Publishing.
- United Nations (last visited 2020). 2020. Intergovernmental Conference on an international legally binding instrument under the United Nations Convention on the Law of the Sea on the conservation and sustainable use of marine biological diversity of areas beyond national jurisdiction (General Assembly resolution 72/249). United Nations. <https://www.un.org/bbnj/>.
- United Nations General Assembly Secretariat. 1977. The question of the definition and/or the delimitation of outer space. United Nations, A/AC.105/C.2/7/Add.1. <https://digitallibrary.un.org/record/741816?ln=en>.
- United Nations General Assembly. 1992. Principles Relevant to the Use of Nuclear Power Sources in Outer Space. United Nations, A/RES/47/68. <https://digitallibrary.un.org/record/159141?ln=en>.
- United Nations Office for Outer Space Affairs. 2010. Space Debris Mitigation Guidelines of the Committee on the Peaceful Uses of Outer Space. United Nations, Vienna, Austria.
- United Nations Treaty Series. 1945. Charter of the United Nations, 1 U.N.T.S. XVI.
- United Nations Treaty Series. 1967. Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, 610 U.N.T.S. 205.
- United Nations Treaty Series. 1968. Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space, 672 U.N.T.S. 119.
- United Nations Treaty Series. 1969. Vienna Convention on the Law of Treaties, 1155 U.N.T.S. 331.
- United Nations Treaty Series. 1972. United Nations Convention on International Liability for Damage Caused by Space Objects, 961 U.N.T.S. 187.
- United Nations Treaty Series. 1975. Convention on Registration of Objects Launched into Outer Space, 1023 U.N.T.S. 15.
- United Nations Treaty Series. 1979. Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, 1363 U.N.T.S. 3.
- United Nations Treaty Series. 1982. United Nations Convention on the Law of the Sea, 1833 U.N.T.S. 3.
- United Nations Treaty Series. 1987. Montreal Protocol on Substances that Deplete the Ozone Layer, 1522 U.N.T.S. 3.
- United Nations Treaty Series. 1994. Constitution and Convention of the International Telecommunication Union, 1825 U.N.T.S. 331.
- United Nations Treaty Series. 2001. Convention on the protection of underwater cultural heritage, 2562 U.N.T.S. 3.
- United States Code. 2020a. 47 U.S.C. §§ 701–757. (Telecommunications Act of 1996).

- United States Code. 2020b. 51 U.S.C. § 50903 (Federal Aviation Administration Act).
- United States Code. 2020c. 51 U.S.C. §§ 51301–51303, U.S. Commercial Space Launch Competitiveness Act (2015), Pub. L. No. 114-90, 129 Stat. 704.
- United States Code. 2020d. 51 U.S.C. § 60121 National and Commercial Space Programs Act of 2010.
- United States District Court for the District of Nevada. 2004. *Nemitz v. United States*, CV-N030599-HDM (RAM), 2004 WL 3167042.
- United States Supreme Court. 2008. *Medellin v. Texas*, 552 U.S. 491.
- Viikari, L. 2015. Environmental aspects of space activities. In *Handbook of space law*, ed. F.G. von der Dunk and F. Tronchetti, 717–768. Cheltenham, UK: Edward Elgar Publishing.
- Vilhena Filho, C. 2005. Security of Tenure in Brazil. Rocky Mountain Mineral Law Foundation Special Institute.
- von der Dunk, F.G. 2001. Dispute resolution mechanisms for space? A few legal considerations. In *Proceedings of the forty-fourth colloquium on the law of outer space*. Toulouse, France.
- von der Dunk, F.G. 2017. International space law. In *Handbook of space law*, ed. F. von der Dunk and F. Tronchetti, 29–126. Cheltenham, UK: Edward Elgar Publishing.
- Williams, M. 2017. Dispute resolution regarding space activities. In *Handbook of space law*, ed. F.G. von der Dunk and F. Tronchetti, 995–1046. Cheltenham, UK: Edward Elgar Publishing.
- World Meteorological Organization. 2019. Scientific Assessment of Ozone Depletion: 2018. Global Ozone Research and Monitoring Project-Report No.58. <https://ozone.unep.org/sites/default/files/2019-05/SAP-2018-Assessment-report.pdf>.
- Written Testimony of James E. Dunstan & Berin Szoka Before the Subcommittee on Space, Science, and Competitiveness of the Committee on Science, Space, and Technology. 2017. United States Senate. <https://2009-2017.state.gov/s//releases/remarks/264963.htm>.
- Written Testimony of Matthew Schaefer before the Subcommittee on Space, Science, and Competitiveness of the Committee on Science, Space, and Technology. 2017. United States Senate. <https://www.hsdl.org/?view&did=807259>.
- Zedalis, R., and C. Wade. 1978. Anti-satellite weapons and the outer space treaty of 1967. *Cal W. Intl. L. J.* 8:454, San Diego, CA.

Chapter 34

Legal Considerations for Space Resources



Austin C. Murnane

Abstract Recent developments in the United States and other nations provide a legal basis for the acquisition and use of space resources by private entities. However, there remain questions and controversies regarding the laws and rules that will govern the use of resources in space, especially for commercial purposes. Uncertainty about the legality of space-resource utilization could discourage investment in space exploration and limit development of resource utilization capabilities. In order to address this issue directly, we consider throughout this chapter the following hypothetical situation: a for-profit commercial venture (the Venture) is technically equipped and adequately funded to extract and use space resources for commercial purposes. With this venture in mind, we consider legal factors that may impact space-resource operations as the industry grows and develops.

34.1 Executive Summary

National governments can be expected to exert supervision and control over space-resource operations like our hypothetical Venture, just as they have supervised private and public space flight since the dawn of the space age. Such supervision is a requirement of international law and consistent with the long-standing practice of states

Austin C. Murnane—Austin C. Murnane is a senior legal counsel for Blue Origin, LLC. He is a member of the International Institute of Space Law, the American Institute of Aeronautics & Astronautics, and other aerospace organizations. He holds a Bachelor of Science degree from the U.S. Naval Academy, a Juris Doctor from Fordham University School of Law, a Master of Science Degree from Colorado School of Mines, and he is a Ph.D. candidate in the Space Resources Program of the Colorado School of Mines. The author thanks Dr. Andrea Harrington and Dr. Fabio Tronchetti for their tutelage and review of this work. Any remaining errors are the author's responsibility, as are any opinions and views expressed herein, which do not reflect the opinions or views of Blue Origin, LLC, its affiliates, or any person besides the author.

A. C. Murnane (✉)
Blue Origin, LLC, 21601 76th Ave S, Kent, WA 98032, USA
e-mail: amurnane@blueorigin.com

regarding the protection and control of their nationals traveling abroad.¹ However, unique legal and environmental characteristics of the space environment will have important ramifications for a space-resource venture.² Although the Venture will find its actions in space governed by at least one nation state on Earth, it will not enjoy the same “real property” rights that a state might grant within its sovereign territory on Earth. Nevertheless, the continuing jurisdiction of an Earth-based state over the Venture’s spacecraft and astronauts will offer important protections that could make space-resource acquisition and use economically feasible, in a manner analogous to state protections for fishing and mining in international waters.

International law and practical considerations will drive states to implement various laws and rules governing space-resource activities. These include regulations to prevent harm or interference among space-resource operations and other actors in space. Other rules, some of which already exist, will limit or control a space-resource venture’s impact on the natural environment of a planetary body and prevent it from returning harmful material to Earth. All states will almost certainly enforce long-standing treaty obligations regarding the rescue and return of astronauts. Several states have also imposed rules for the protection of historical sites in outer space. Finally, it is possible that future developments in international law will create an over-arching regime for space-resource governance, which could largely supplant existing state laws and rules. Advocacy for such a regime has arisen in certain non-governmental and academic contexts.

There appears to be a small but growing multinational consensus in favor of using space-resource to facilitate space exploration, and involving commercial entities like our hypothetical Venture in that process. Authorities in several spacefaring states generally express enthusiasm about the possible uses of space resources by official and commercial entities. Several states clearly favor using national laws and bilateral or multilateral agreements to enable the growth of this industry. Others have signaled support for a more international approach.

The following analysis will discuss the laws and rules that will impact the rights, responsibilities, and risks of a space-resource venture. The analysis will rely on illustrative examples of domestic laws that implement international principles. The most frequent domestic examples will come from the United States, due to the United States’ uniquely robust development of space-resource law and policy as well as the author’s experience and expertise.

34.2 Principles of Space Resources Law

Space-resource operations like our hypothetical Venture will be subject to a variety of controls that arise from international and domestic air and space law as well as other legal authorities. Many of these controls will exist because of the related but distinct

¹ See, e.g., 2015 U.S. SPACE Act; 2017 Luxembourg Space Law; 2019 UAE Space Law.

² See, e.g., OSI Open Letter (2020).

legal concepts of national responsibility, jurisdiction, and sovereignty, all three of which are addressed in detail below. These concepts are foundational aspects of law and they are implicated throughout the four widely accepted international treaties on outer space, which are known as the Outer Space Treaty,³ the Rescue Agreement,⁴ the Liability Convention,⁵ and the Registration Convention.⁶ In these treaties we find specifically space-related legal principles that are especially important to space resources, including the “freedom of use,” “non-appropriation,” “due regard,” and “non-interference” principles, among many others. Wherever our Venture is based, it is likely that the laws and regulations governing its operations will depend upon the state’s interpretation and implementation of these international principles. It is also possible that an international regime for space-resource governance could arise, especially if a more controversial treaty known as the Moon Agreement,⁷ which has achieved ratification in eighteen nations,⁸ were to gain wider acceptance.

We must also keep in mind that national and international laws related to space resources are few in number and often vaguely articulated. The history of space-resource acquisition and use is also limited, and the record of commercial transactions involving space resources is sparse as well. In legal terms, the shortage of references to space resources in treaties and legislation means that there is very little “positive law” on this subject.⁹ Likewise, the near-absence of mineral resource utilization in the relatively brief history of space exploration means that there are few commonly accepted practices of the kind that might establish “customary law.”¹⁰ Finally, the shortage of published court decisions directly related to space-resource utilization makes it difficult to identify legal precedents that lawyers and courts might cite in support of future rulings.¹¹ In summary, there remains a great deal of ambiguity surrounding many of the legal principles applicable to space resources, including the provisions of laws and treaties that specifically discuss the subject. Therefore, the decisions of a relatively small number of authorities and the actions of a few operators could have a profound effect upon the development of space-resource laws and practices over the next several decades.

³ Full title: Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies (1967).

⁴ Full title: Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space (1968).

⁵ Full title: Convention on International Liability for Damage Caused by Space Objects (1972).

⁶ Full title: Convention on Registration of Objects Launched into Outer Space (1975).

⁷ Full title: Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (1979).

⁸ As of 2020, the state parties to the Moon Agreement are Armenia, Australia, Austria, Belgium, Chile, Kazakhstan, Kuwait, Lebanon, Mexico, Morocco, Netherlands, Pakistan, Peru, Philippines, Saudi Arabia, Turkey, Uruguay, and Venezuela. See *Treaties Status* (2020).

⁹ See “positive law,” Black’s (2019).

¹⁰ See “customary law,” Black’s.

¹¹ See “citable” Black’s.

34.2.1 *National Responsibility, Jurisdiction, and Sovereignty*

A space-resource venture must always be mindful of its legal rights and responsibilities because all spacecraft, including privately owned and operated spacecraft, will be subject to the laws of at least one state on Earth throughout their space-flight missions: from launch to orbit, through orbital transfers, through rendezvous or landing upon other planetary bodies, through resource acquisition, use, and/or sale, and all the way through re-entry, if applicable. Positive international law (i.e., law arising from binding treaties like the Outer Space Treaty) requires spacefaring states to supervise spacecraft and their operators even when they are far from Earth. Although it may seem odd, for instance, that US laws or regulations would apply to a privately owned spacecraft on the far side of the Moon, this is consistent with the standard practice of governments on Earth, which routinely maintain jurisdiction over their nationals, aircraft, and vessels when they travel internationally.¹² This legal authority over persons should not be confused with a government's territorial jurisdiction, which is limited to a nation's borders. Even though Earth's states do not claim territory in space, they maintain continuous legal authority and responsibility over their spacefaring nationals. These are the legal foundations on which controls and protections for space-resource operations arise.

34.2.1.1 **The Extension of Terrestrial Law into Outer Space**

The laws of Earth extend into outer space because all spacefaring states, plus most non-spacefaring states,¹³ have agreed to supervise all of their public and private activities in space. The states made these agreements in the aforementioned Outer Space Treaty, Liability Convention, and Registration Convention, which hold state parties (nations that sign and ratify the treaties) responsible for their "national activities" in space, regardless of whether those activities are conducted by an official space program, like NASA, or a private entity like our hypothetical space-resource Venture.¹⁴ Under these treaties, state parties can be liable to each other for damages that their spacecraft, whether public or privately owned or operated, might cause.¹⁵ State parties have also pledged to provide information about each of their space objects in a publicly available registry, and to retain jurisdiction and control over their registered spacecraft, and any personnel thereof, whether they are flying in

¹² See *S.S. Lotus* (1927) at 19.

¹³ As of 2020, 110 out of 195 UN member states have ratified the Outer Space Treaty. The largest non-ratifier, in terms of population, is Ethiopia, which has signed but not ratified the Treaty. The largest non-ratifier, in terms of gross domestic product, is Iran, which has also signed but not ratified. *Treaties Status* (2020).

¹⁴ Outer Space Treaty (1967) art. VI.

¹⁵ Outer Space Treaty (1967) art. VII; Liability Convention (1972) arts. II—V. Note that the treaties generally hold the "launching State" liable for damages caused by space objects.

space or landed upon a celestial body.¹⁶ Published information on space activities includes the locations or orbital parameters of spacecraft and the nature and results of their operations.¹⁷ National governments comply with these treaty requirements by establishing domestic laws to regulate private and public space missions.¹⁸ Therefore, our hypothetical Venture, its spacecraft, and its astronauts (if any) should expect continuous monitoring and supervision from at least one Earth-based government at all times in space.

There may be some question, however, about which particular government, or group of governments, may exert authority over a proposed space mission. The Outer Space Treaty makes it relatively clear that the state that registers a spacecraft must “retain jurisdiction and control” over the spacecraft and its personnel. The same treaty requires that the “appropriate” state, presumably the state of registration, must authorize and continuously supervise their national activities in space, even when the spacecraft or personnel conducting those activities belong to non-governmental entities like our Venture.¹⁹ The Outer Space Treaty, Liability Convention, and Registration Convention all indicate that the state of registration should be a “launching State,” but here the potential for confusion or disagreement can arise, especially for certain private spaceflight operations, because the various treaties define “launching State” as the state that “launches or procures the launching” of the spacecraft²⁰ or a state from whose “territory or facility” the spacecraft is launched.²¹

If multiple countries satisfy one or more of the “launching State” definitions for a single space mission, there could be disagreement as to which one of them should register the spacecraft and thereby exercise jurisdiction and supervision. For instance, the venture could be a United States company that has built its own lunar prospecting rover. However, the same venture might engage a French company to carry this rover to the Moon aboard a French spacecraft. The French spacecraft, in turn, might initially launch from Earth’s surface atop a rocket built and operated by a Russian space agency, and launched from Kazakhstan. In such a scenario, several nations might be considered “launching States,” but only one should be the state of registration of the rover.

The Registration Convention requires that, if there are multiple launching States, they must jointly determine which one of them shall register the spacecraft and thereby exercise jurisdiction and control over it.²² The Registration Convention does not provide further guidance as to how launching states should make this determination. The UN General Assembly has recommended that where there are “joint launches” of space objects, there should be separate registration of “each space

¹⁶ Outer Space Treaty art. VIII.

¹⁷ Outer Space Treaty art. XI; Registration Convention (1975) arts. II, IV, V.

¹⁸ See, e.g., U.S. Code Title 51, National and Commercial Space Programs.

¹⁹ Outer Space Treaty arts. VI, VIII.

²⁰ Outer Space Treaty (1967) art. VII; Liability Convention art. I; Registration Convention art. I.

²¹ Liability Convention art. I; Registration Convention art. I.

²² Registration Convention art. II.

object” launched.²³ In such cases, the state responsible for operating each space object should list that object on its national registry. This has been the practice, for example, in the case of the Beresheet lunar landing mission of 2019, in which the United States registered the launch vehicle²⁴ and Israel registered the lunar lander.²⁵

A space-resource venture should also be aware that there has been significant inconsistency in the registration decisions made by launching states, however defined. This includes different registration practices for “joint launches” as well as other circumstances.²⁶ Due to the lack of uniform state practice in this area, space-resource operations should take care to determine, well in advance of launch, the applicable laws and policies of any nation that could possibly consider itself to be the venture’s “launching State.” This review is especially important due to the different positions that various states have adopted (see Sects. 34.3.1–34.3.18) and may adopt in the future regarding the legality of space-resource utilization. Therefore, the review should include, at a minimum: (1) the state where the venture is incorporated, (2) the state from which the venture’s spacecraft or payload is controlled, (3) the state from which it is launched, and (4) any state claiming responsibility or control over transfer vehicles, landers, communications stations, or other support functions.

34.2.1.2 Jurisdictional Principles Applicable to Space-Resource Activities

When a state’s government prescribes, adjudicates, and enforces laws for spacecraft, their crews, and their activities, the government is exercising jurisdiction²⁷ in space, and it is usually doing so based on what is known as the “nationality principle.” As we shall see below, the nationality principle, along with the related “passive personality principle,” is more likely to apply to our space-resource venture than the “territoriality principle,” which is a more traditional basis for jurisdiction on Earth. Other reasons for a government to exercise jurisdiction over space-resource activities may include the “protective principle,” sometimes known as the “impact territoriality principle,” as well as a rarely invoked “universal principle” of jurisdiction.²⁸ These principles provide opportunities and limitations for governments to wield authority in outer space, with important implications for space-resource operations.

The nationality principle refers to the power that a state claims over its nationals, which might include citizens, residents, or other persons associated with the state, no matter where they travel. It applies to natural persons (individual human beings) and legal or juridical persons (corporations) as well as aircraft and vessels.²⁹ This

²³ G.A. Res. 62/101 (2008).

²⁴ U.S. Registry (2020).

²⁵ Israel Registry (2019).

²⁶ Schmidt-Tedd and Soucek (2020).

²⁷ See, e.g., Restatement: Foreign Rel. (2018) § 401.

²⁸ See Benchbook on Int’l L. (1972) II.A-2; *see also* Gorove (1972) 313–14.

²⁹ See, e.g., ILC Report (2006) Annex V 231–232 ¶ 14.

principle of jurisdiction enables enforcement of civil and criminal responsibilities that states create for their nationals.³⁰ Under this principle, the state of registry for a space-resource venture would prescribe rules for the conduct of the venture's operations on an alien planet, and enforce those rules using the state's courts and agencies, if necessary. For crewed missions, the state claiming jurisdiction over the venture's spacecraft under the nationality principle would, by extension, claim jurisdiction over all persons aboard, regardless of their individual nationalities. Again, this follows the terrestrial practice by which a state asserts jurisdiction over all passengers and crew, no matter their nationalities, as long as such persons are aboard the state's national or "flagged" vessels and aircraft.

Similarly, the passive personality principle allows a state to protect its nationals and their rights, even when they travel outside of the state's territory.³¹ This principle is the basis, for example, on which United States laws criminalize actions harming US nationals abroad. Such laws may apply whether the US victims are located in foreign countries, sailing in international waters, or in aircraft flying over international waters, and even when the perpetrators are not US nationals.³² The passive personality principle can also provide private causes of action for a state's nationals to bring civil lawsuits against foreign persons for actions abroad.³³ A state would be invoking this principle if it took action to defend a space-resource venture that suffers harm in space, or if the state enabled the venture itself to sue the perpetrator of any such harm.

The territoriality principle, on the other hand, applies jurisdiction in essentially the opposite manner. Whereas nationality- and passive personality-based jurisdictions represent a state's responsibility for its nationals, regardless of where they are, territorial jurisdiction is a state's responsibility over its territory and whoever might be located therein³⁴—with some exceptions.³⁵ Perhaps most importantly for those involved in resource prospecting and mining, territorial jurisdiction includes the state's power to grant and recognize private ownership of land or territory, otherwise known as real property or real estate.³⁶ A state typically wields the sole power to issue or endorse deeds, which identify the holders of title (ownership) for parcels of land.³⁷ The state may also specify in a deed whether the title holder owns any resources located above or below the land's surface. For instance, the state may declare that a person owns a parcel of land, but *not* certain resources beneath the

³⁰ Restatement: Foreign Rel. § 402 n.4.

³¹ ILC Report (2006) Annex V, 231 ¶ 15.

³² Restatement: Foreign Relations (2018) § 402 n.4.

³³ See, e.g., Terrorism exception, 28 U.S.C. § 1605A (2008).

³⁴ Benchbook on Int'l L. (1972) II.A-2.

³⁵ Certain persons, especially diplomats, may enjoy immunity from territorial jurisdiction. See, e.g., the Diplomatic Relations Act (1978), 22 U.S.C. § 254.

³⁶ See, e.g., U.S. Const. art. IV § 3 ("The Congress shall have power to dispose of... the Territory or other Property belonging to the United States.").

³⁷ See, e.g., Bagnell (1839).

surface.³⁸ All of these legal powers derive from a state's territorial jurisdiction or sovereignty, the power the state holds over its sovereign territory.³⁹

For a space-resource venture like ours, the most important feature of territoriality is that it does *not* exist in outer space. The prevailing legal interpretation is that, by joining the Outer Space Treaty, the spacefaring states have foresworn territorial jurisdiction in space, specifically agreeing that, "Outer space, including the Moon and other celestial bodies, is not subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means."⁴⁰ Therefore, no state asserts territorial powers in space. A space-resource operation cannot ask its government to decide who owns various parcels of land on the Moon another celestial body,⁴¹ nor can a state declare that anyone holds exclusive rights to the resources on or below any extraterrestrial parcel. Section 34.2.2 discusses further implications of this prohibition for space-resources operations.

Finally, it is possible that a state might act under the protective or universal principles of jurisdiction to exert control over space-resource activities. The protective principle is also known as the "impact territoriality" principle because it is concerned with protecting the state itself and its interests from harmful impacts originating elsewhere.⁴² Typical terrestrial examples include counterfeiting of a state's currency or espionage against its government, both of which are threats that a state may act to suppress outside of its territory, regardless of who commits them.⁴³ In a space-resource context, a state may act under this principle to prevent or strictly regulate the delivery of space resources to Earth in order to mitigate the risk of harmful biological contamination.⁴⁴ Section 34.2.2.2 provides further discussion of such regulations.

It is also possible, though less likely, that a state might invoke the universal principle of jurisdiction in a space-resource context. This is the principle by which states sometimes address "certain offenses of universal concern, such as piracy, slavery, forced labor, trafficking in persons" and others.⁴⁵ It is conceivable that a state might cite this principle when acting to combat piracy against space-resource operations like our hypothetical Venture—presumably cyberpiracy involving hacking, spoofing, and jamming of communications.

³⁸ See, e.g., Coal and mineral rights (1993), 42 U.S.C. § 299.

³⁹ Curry (2007).

⁴⁰ Outer Space Treaty (1967) art. II.

⁴¹ Note, however, that the term "celestial body" lacks a clear definition in international law, providing a possible ambiguity regarding the territorial prohibition in Article II of the Outer Space Treaty. See Lyall and Larsen (2018) 182.

⁴² Benchbook on Int'l L. (1972) II.A-3.

⁴³ Restatement: Foreign Rel. (2018) § 402 comment i.

⁴⁴ See, e.g., NPD 8020.7G (2013).

⁴⁵ Restatement: Foreign Rel. § 402 comment j.

34.2.2 *Rights and Responsibilities of Space-Resource Operations*

The Outer Space Treaty provides that the Moon and other celestial bodies shall be “free for exploration and use by all States”⁴⁶ The prevailing consensus among space-law scholars is that space-resource mining is one of the activities that falls under the “freedom of use” guaranteed in this foundational treaty but, like all other space activities, it will be governed by law.⁴⁷ The use of space, along with space exploration, is an activity that shall be “carried out for the benefit and in the interests of all countries,” according to the Treaty, which also provides that space exploration and use “shall be the province of all mankind.”⁴⁸

In much the same way that treaties and laws on Earth recognize competing legal interests, space law incorporates several competing interests in space exploration. These include freedom of use, freedom of access, scientific investigation,⁴⁹ peace and security,⁵⁰ non-proliferation of weapons of mass destruction,⁵¹ astronaut safety,⁵² responsibility for damages,⁵³ national supervision,⁵⁴ avoidance of harmful contamination, and protection of Earth.⁵⁵ Just as courts and governments have always worked to resolve disagreements over liberty, safety, and prosperity, our Venture should expect that the same institutions will weigh competing interests when controversies arise regarding space resources. The following section therefore describes the most prominent rights and responsibilities related to space resources and consider how states might implement them.

34.2.2.1 **Space-Resource Rights and Protections**

As we noted earlier, states have generally given up the right to claim sovereignty over territory on the Moon or other celestial bodies under Article II of the Outer Space

⁴⁶ Outer Space Treaty (1967) art. I.

⁴⁷ Hobe et al. 41 (2016).

⁴⁸ Outer Space Treaty art. I. A common misconception is that the Treaty declares space and its regions to be “the province of all mankind.” On the contrary, the Treaty declares “the *exploration and use of* outer space, including the Moon and other celestial bodies” to be humankind’s province. By employing “province” to describe activities (exploration and use), the Treaty relies on a less common usage, in which “province” refers to a “proper or appropriate function or scope” instead of a physical area or region. Merriam-Webster (2021).

⁴⁹ Outer Space Treaty art. I.

⁵⁰ Outer Space Treaty art. III.

⁵¹ Outer Space Treaty art. IV.

⁵² Outer Space Treaty art. V; Rescue Agreement (1968).

⁵³ Outer Space Treaty arts. VI and VII; Liability Convention (1972).

⁵⁴ Outer Space Treaty art. VI and VIII; Registration Convention (1975).

⁵⁵ Outer Space Treaty art. IX.

Treaty.⁵⁶ A space-resource venture should therefore not expect any government to grant or sell to the venture the exclusive rights to a parcel of land that would be available on Earth. This unavailability of real property rights from Earth states has led some scholars to suggest that it is legally impossible for a commercial entity like our hypothetical Venture to acquire resources as property in space.⁵⁷ However, a state can recognize and protect the Venture's right to acquire, keep, and use *moveable* property in space, including resources that the Venture physically collects from the unowned lands of the Moon and other bodies. Indeed, several nations have already explicitly recognized these rights, and more appear ready to follow. The laws providing for such activity include the 2015 U.S. Commercial Space Launch Competitiveness Act (the 2015 U.S. SPACE Act),⁵⁸ Luxembourg's Law of 20 July 2017 on the Exploration and Use of Space Resources (the 2017 Luxembourg Space Law),⁵⁹ the 2019 United Arab Emirates Law on the Regulation of the Space Sector (the 2019 UAE Space Law),⁶⁰ and Japan's 2021 Act on Promotion of Business Activities Related to the Exploration and Development of Space Resources (the 2021 Japan Space Law).⁶¹

All four of these nations have joined the Artemis Accords,⁶² thereby affirming "that the extraction of space resources does not inherently constitute national appropriation under Article II of the Outer Space Treaty."⁶³ Australia, Bahrain, Brazil, Canada, Colombia, France, Israel, Italy, Mexico, New Zealand, Nigeria, Poland, Republic of Korea, Romania, Rwanda, Saudi Arabia, Singapore, Ukraine, and the United Kingdom have also joined the Accords.

Each of the space-resource laws listed above provides a legal basis for a state to regulate the acquisition of space resources without claiming the territory from which those resources are collected. In so doing, these laws can be described as analogous to long-standing legal principles regarding collection of resources in unowned areas, like fishing in international waters—an analogy pointed out by officials in the United States and Luxembourg, among others.⁶⁴

For example, the U.S. Fishermen's Protective Act of 1967 requires the U.S. Secretary of State to take necessary steps to protect fishing vessels and their crews if they are seized by foreign countries due to their fishing activities.⁶⁵ Likewise, the U.S. Deep Seabed Hard Mineral Resources Act of 1980 (the U.S. Deep Seabed Act) protects the

⁵⁶ Outer Space Treaty (1967) art. II.

⁵⁷ See, e.g., Markoff 83 (1970); see also Christol (1984).

⁵⁸ 2015 U.S. SPACE Act.

⁵⁹ 2017 Luxembourg Space Law.

⁶⁰ 2019 UAE Space Law.

⁶¹ 2021 Japan Space Law.

⁶² The Artemis Accords are a series of bilateral agreements establishing shared principles, guidelines, and best practices for space exploration with the intention of advancing the Artemis Program, a United States-led multilateral effort to conduct sustainable human exploration of the Moon. Artemis Accords (2020) § 1.

⁶³ Artemis Accords § 10.2.

⁶⁴ See, e.g., Hofacker (2020), Amos (2016).

⁶⁵ Fishermen's Protective Act, 22 U.S.C. § 1972 (1967).

rights of US citizens to collect hard minerals from the ocean floor beneath international waters for commercial purposes.⁶⁶ Japan, France, Italy, the United Kingdom, and the former Federal Republic of Germany (West Germany) have passed similar deep seabed mining laws.⁶⁷ The U.S. Deep Seabed Act also provides a regulatory basis for the issuance of licenses and prohibitions to control this activity.⁶⁸ In the Act itself, the United States clarifies its authority by distinguishing among the jurisdictional principles we discussed in Sect. 34.2.1.2: the United States “exercises its jurisdiction over [its] citizens and vessels” at sea, which is in line with the nationality principle. However, the United States “does not thereby assert sovereignty” over “any areas or resources in the deep seabed,” which is to say that there is no claim of territorial jurisdiction or, as the Act describes it, “extraterritorial sovereignty.”⁶⁹

In these examples and others, we see a state’s power extending beyond its borders—not to claim sovereignty over new territory, but to protect and control the actions of the state’s nationals as they acquire moveable property in an unclaimed space. In this spirit, the 2015 U.S. SPACE Act recognizes the rights of US citizens to “possess, own, transport, use, and sell” resources obtained in outer space.⁷⁰ The same law requires the president of the United States to promote the rights of US citizens to engage in commercial space-resource activities in a manner similar to the U.S. Fishermen’s Protective Act.⁷¹ The 2015 U.S. SPACE Act also contains language strikingly similar to that of the U.S. Deep Seabed Act, whereby the US Congress asserted that its recognition of resource collection rights for US citizens does *not* constitute an assertion by the United States of “sovereignty or sovereign or exclusive rights or jurisdiction over . . . any celestial body.”⁷² Again, no extraterritorial or extraterrestrial claims of sovereignty are made. Instead, a state here asserts only its jurisdiction over its citizens in outer space, in accordance with the nationality and passive personality principles, and pursuant to the Outer Space Treaty, Liability Convention, and Registration Convention.

It should be noted, however, that whereas the “freedom of fishing” in international waters is specifically recognized in the UN Convention on the Law of the Sea (UNCLOS),⁷³ the foundational treaties of space law do *not* explicitly mention space resources or the collection thereof—either positively or negatively. As we have discussed, Article II prohibits “national appropriation” of outer space, including claims of sovereignty, but it stops short of prohibiting or even discussing private acquisition of resources.⁷⁴ This omission does not reflect a lack of imagination or foresight regarding the future of space exploration. On the contrary, the Outer Space

⁶⁶ Deep Seabed Act (1980) 30 U.S.C. § 1401 et seq.

⁶⁷ Tronchetti (2009) 111.

⁶⁸ 30 U.S.C. §§ 1411–12.

⁶⁹ *Id.* § 1402(a).

⁷⁰ *Id.* § 51,303.

⁷¹ 51 U.S.C. § 51,302.

⁷² 2015 U.S. SPACE Act.

⁷³ UNCLOS (1982) art. 87(1)(e).

⁷⁴ Outer Space Treaty (1967) art. II.

Treaty's drafters, working in the 1960s, did contemplate a range of future activities beyond the scientific activities of official space agencies. Article VI, for example, addresses space activities carried on by "non-governmental entities."⁷⁵ Article IX, discussed in more detail below, addresses the possibility of "an activity or experiment planned by [a State Party] or its nationals" in space.⁷⁶ The Treaty also repeatedly asserts and promotes the rights of all countries to explore and "use" outer space, including the Moon and other celestial bodies.⁷⁷

The Treaty does not limit or define the "use" of outer space, so long as the users' purposes are "peaceful."⁷⁸ Nowhere does the Treaty limit the use of outer space to strictly scientific or official missions, nor does it derogate commercial uses of space, celestial bodies, or their resources. This omission was not an oversight. The Treaty's drafters had seen and considered a draft version of Article II, proposed by the International Institute of Space Law (IISL), which would have contained a more comprehensive ban: "Celestial bodies *or regions on them* shall not be subject to national or *private* appropriation, by claim of sovereignty, by means of use or occupation or by other means."⁷⁹ No prohibition against private appropriation appears in the final version of the Treaty.

In a 2016 paper, leading international space-law scholars associated with the IISL reviewed the applicable treaties as well as other authoritative sources and determined that space-resource utilization falls within the "freedom of use" guaranteed in Article I of the Outer Space Treaty, so long as other requirements of international law are met.⁸⁰ However, controversy persists, particularly among space-law scholars, as to whether states should be allowed under international space law to unilaterally authorize their nationals or anyone else to acquire resources in space through national legislation.⁸¹

Although jurists and others will continue to disagree as to whether states *should* enable commercial use of space resources through national legislation, the fact remains that several states have done so and others are following their lead. It remains to be seen whether the commercial use of space resources will be economically feasible without a sovereign government granting deeds of title to tracts of land on the Moon or elsewhere. However, certain provisions of national and international law, along with certain characteristics of the space environment, may reduce economic risks and thereby facilitate resource acquisition in this territorial vacuum.

As we have discussed, the foundational space treaties hold launching states liable for damages that their spacecraft cause. This liability includes harms that spacecraft or their component parts may cause to individual persons as well as corporations in outer

⁷⁵ Outer Space Treaty art. VI.

⁷⁶ Outer Space Treaty art. IX.

⁷⁷ Outer Space Treaty art. I, III, IX, X, XI, and XIII.

⁷⁸ Outer Space Treaty art. IV, IX, and XI.

⁷⁹ IISL Draft (1965) (emphasis added).

⁸⁰ Hobe et al. 41.

⁸¹ See, e.g., Tronchetti 32, 211–25 (2009).

space, including on the Moon and other celestial bodies.⁸² Launching and registration states are required to authorize and continuously supervise their space operators, both public and private, in order to ensure compliance with treaty obligations.⁸³ Therefore, these treaty provisions offer some protection to space-resource operations because they require governments to prevent their supervisees in outer space from breaking into a venture's spacecraft to steal harvested resources. At the very least, the treaties would hold the thieves' governments liable for any such theft.

The venture should be aware, however, that the international liability regime only creates obligations between states, meaning that the launching state of a spacecraft that commits a theft or other damage need only compensate the *government* of the victims, and not the victims themselves, for the harm done.⁸⁴ If the venture's own government refuses to pass along the payment from the wrongdoer's government, the venture may find itself uncompensated for any damages. The venture can still take some solace in the reduced risk of such thefts ever happening thanks to the deterrent effect created by these treaties, the continuous supervision requirements they impose, and the practical limitations of the space environment, which are discussed further below. Although reliance on treaty-based protections may provide limited comfort to a venture considering the high costs and risks of space operations, this degree of uncertainty is the nature of law, and life, in all unclaimed realms, from the deep seabed to the high seas, where the relatively remote possibility of pirate attack does not prevent widespread global commerce.

The example of international piracy raises an important environmental distinction that should give additional comfort to space-resources operators. Piracy and other crimes of theft and violence are feasible in international waters due to environmental factors that are entirely different from the space-resources context. Whereas a handful of persons, armed with a few weapons in a single-engine boat, stand a chance of successfully committing a lucrative act of piracy in some waters, the prospect of a similar criminal enterprise on the Moon or another planetary body is not a serious consideration in the near term. Instead, the most likely threats to space operations involve cybercrimes.⁸⁵ States can be expected to take action to protect their nationals from such threats, as they already do. The U.S. Space Command, for example, is already tasked with promoting security and stability in space and protecting the interests of the United States and its allies.⁸⁶ In the future, this mandate could include space-resource operations, especially because US law obliges the president of the United States, the commander-in-chief of the US military, to promote US citizens' rights to acquire and use space resources for commercial purposes.⁸⁷ Just as navies, coast guards, and other military and law enforcement agencies protect terrestrial commerce in the air, on land, at sea, and in cyberspace, it can be expected that

⁸² Outer Space Treaty art. VII; Liability Convention art. III.

⁸³ Outer Space Treaty arts. VI and VIII; Liability Convention; Registration Convention.

⁸⁴ Outer Space Treaty art. VII; Liability Convention art. VIII.

⁸⁵ Hitches (2020).

⁸⁶ Space Command (2021).

⁸⁷ 2015 U.S. Space Act § 51,302(a)(3).

security forces' mandates will expand to cover commercial activity in outer space as hazards and threats arise. For this reason, cyber defense is already a primary area of focus for the U.S. Space Force and Space Command.⁸⁸

34.2.2.2 Possible Restrictions on Space-Resource Operations

It is impossible to predict with certainty the restrictions that governments might impose upon space-resource activity, but international and national laws already offer helpful guidance. The Outer Space Treaty and other widely accepted space treaties impose several obligations on state parties, leading to the creation of laws and rules affecting space-resource operations like our hypothetical Venture. National governments could impose other restrictions for a variety of reasons. The purposes of these restrictions include, but are not limited to, the prevention of harm and interference with others, planetary protection and defense, rescue and return of astronauts in distress, and the preservation of historical sites. It also remains possible, though unlikely, that states will belatedly embrace the Moon Agreement or accept its call for an international regime, which could profoundly affect space-resource development.

Preventing Harm and Interference Among Operators

International law provides several protections against harmful activity and interference in space. States enact laws and regulations in keeping with these principles, which can be expected to help space-resource operations by enabling them to operate safely, while also limiting the operations' freedom of action in order to protect others. The creation of safety zones appears to be the first step that governments will take to mitigate these risks.

First, a space-resource venture should expect that regulations and laws will command it to do no harm to others in space. This requirement will arise, in part, because Article IX of the Outer Space Treaty requires state parties to act at all times with due regard to the corresponding interests of other states. The same Article of the Treaty specifically requires states to engage in international consultations prior to any "potentially harmful interference with the activities" of other states. Furthermore, as we have discussed, the Outer Space Treaty, Liability Convention and Registration Convention contain other provisions to hold launching states liable for damages that their space activities cause to others. For these reasons, governments would impose restrictions on the venture to prevent it from harming other nations' astronauts, damaging their equipment, or interfering with their activities. Obvious examples might include controls to prevent a venture's rover from crashing into others, or

⁸⁸ Hitches.

controls to ensure that a venture safely operate its nuclear power source.⁸⁹ These would be logical extensions of safety requirements on Earth.⁹⁰

Interference, on the other hand, could raise thornier legal issues than harm. If a spacecraft is damaged or an astronaut is injured, harm has clearly occurred, but if activity in space does not physically damage persons or property, and only frustrates the plans, hopes, or aspirations of others, such activity might constitute harmful interference—or it might not. The Outer Space Treaty does not define what kind of “harmful interference” creates an obligation among state parties to consult, nor does the Treaty provide any guidance about the consultations required or how they might resolve any disputes. It does not clarify, for instance, whether a venture’s mining activities might constitute harmful interference if they disturb the readings of another person’s seismometer or melt ice that another person hoped to examine in its frozen state. Furthermore, the Treaty does not clarify whether it would constitute harmful interference if a rival miner harvested resources that the venture had planned to mine, perhaps after an expensive prospecting effort.

On Earth, prospectors and miners can avoid interference, or at least resolve their disputes about it, thanks to mining laws established under the territorial jurisdiction of sovereign governments. These laws clarify the ownership and control of mining sites by allowing prospectors to stake exclusive claims to resources in the ground. They also establish requirements and limits to prevent wide-ranging claims that would deny opportunities to other miners and leave resources inefficiently unused. For instance, on lands owned by the United States, the General Mining Act of 1872 provides that a new mining claim to a vein or lode of precious metals may be no longer than 1,500 feet (457 m) long and 600 feet (183 m) wide. In order to make such a claim, a prospector must first confirm that a vein or lode of precious metals is located within the geographic boundaries of the claim.⁹¹ Other federal and local laws and regulations provide additional requirements, further clarity, and additional mitigation of the risk of interference and disputes.⁹²

The first prospectors on the Moon and other planetary bodies will probably arrive in a place that lacks a general territorial authority prescribing limits and rules for mining claims. In this one respect, the situation will be somewhat similar to California in the mid-nineteenth century, when prospectors and miners found themselves in an almost complete legal vacuum because the Mexican government had ceded the territory to the United States, which failed to establish institutions of governance and law for several years. Fortunately for our hypothetical Venture and its fellow prospectors, the analogy of a “wild west” goes no farther than this absence of a territorial claims authority. Unlike California’s “’49ers,” who lived temporarily in something approaching a state of nature without laws or institutions to enforce them,⁹³

⁸⁹ See, e.g., RFP (2020).

⁹⁰ See, e.g., U.S. Atomic Energy Act (1954) 42 U.S.C. § 2011 et seq.; U.S. Nuclear Waste Policy Act (1982) 42 U.S.C. § 10,101 et seq.

⁹¹ U.S. Gen. Mining Act (1872) 30 U.S.C. § 23.

⁹² See generally Colo. Rev. Stat., Tit. 34 Art. 43 (2018).

⁹³ McDowell (2002) 2.

space-resources operators will fall under a constant system of government supervision from the very beginning of their operations. Relying on the historical principles of extraterritorial jurisdiction discussed in Sect. 34.2.1.2, and acting pursuant to the control and supervisory requirements of international space law, which we described in Sect. 34.2.1.1, Earth's states will impose regulations and restrictions to promote peace, safety, cooperation, and other interests, which are discussed in this and the following sub-sections. These states may also impose territorial limits on their nationals' mining activities, which may effectively constrain space-resource operations in the same way that mining laws limit the scope of claims on Earth. The difference, of course, will be that limits imposed by one state on its own miners will not be binding on other states and their miners.

Then again, it is possible that prospectors and miners from several states will take a leading role in establishing, under the watchful supervision of their Earth-based governments, customs and common practices for staking claims, just as the miners of California did prior to the imposition of mining laws. If space-resource operations were to follow the example of the California miners, they would create a *de facto* regime that strictly limits the scope of each mining operation, preserves opportunities for new entrants into the field, and ensures efficient use of resources.⁹⁴ Environmental realities on the ground can be expected to influence the rules and practices these miners and their states adopt.

The Hague International Space Resources Governance Working Group (ISRGGWG), an international forum representing stakeholders from academia, industry, governments, and NGOs, proposed in 2019 that states and international organizations should be permitted to establish "safety zone[s], or other area-based safety measure[s]" around space-resource activity sites in order to "assure safety and to any harmful interference with that space-resource activity."⁹⁵ Since 2020, 23 states have adopted the ISRGGWG's safety zone recommendation by joining the Artemis Accords, which include a section on "Deconfliction of Space Activities." In this section, the Accords' signatories state their intention to cooperatively develop "safety zones" that might prevent issues of harmful interference, while reaffirming their commitments to non-interference under the Outer Space Treaty. The Accords describe safety zones in flexible and modest terms, emphasizing that the dimensions of the safety zones will be limited by reasonable scientific and engineering principles, and change over time as the need for such zones changes or ceases to exist. Notably, the signatories do not make any claims to exclusive use of the zones. Instead, they commit only to share information about the safety zones that they believe to be necessary around their own operations, and to respect the safety zones of others—but only to the extent they are reasonable.⁹⁶ If the proposed safety zones are implemented, they may function similarly to the 200-km "keep-out sphere" that the United States,

⁹⁴ McDowell 4.

⁹⁵ Building Blocks (2019) 11.3.

⁹⁶ Artemis Accords (2020) § 11.6–11.

Russia, and other states voluntarily observe during spaceflight in the vicinity of the International Space Station (ISS).⁹⁷

Space-resource operations should therefore expect, if their state is an Artemis signatory, that authorities will not allow a venture to encroach upon the safety zones of others. A venture should likewise expect its government will require information about its activities in order to inform the state's determinations regarding whether and how to establish safety zones.

One of many issues yet to be resolved is how the United States, China, and their nationals and partners will avoid harmful interference with each other's space-resource operations on the Moon and other planetary bodies. Current US law, known as the Wolf Amendment, strictly controls cooperation between NASA and the Chinese government as well as Chinese-owned companies.⁹⁸ Chinese entities do not participate in the NASA-led Artemis Program or most other US space activities. The China National Space Administration (CNSA) has its own Chinese Lunar Exploration Program (CLEP), which has already launched five robotic missions to the Moon, three of which have landed, and one of which has returned samples of lunar resources to Earth. The stated goals of CLEP include the use and development of resources on the Moon. In this effort, CNSA has already collaborated with Germany, Sweden, and the Netherlands on spacecraft development. It has also made agreements with Russia, Turkey, Ethiopia, and Pakistan for future projects.⁹⁹ The latest of these agreements calls for the establishment of an International Lunar Research Station (ILRS), to be built on or around the Moon by China, Russia, and other international partners.¹⁰⁰

China and its partners may develop non-interference practices that are separate and distinct from those described in the Artemis Accords. Such a scenario may lead to difficulties for space-resource operations, especially if different states' non-interference rules conflict with each other. On the other hand, China, its partners, and other states may adopt non-interference practices that are substantially similar to those recommended by the Hague ISRGWG and implemented by the Artemis Accords. China may also reach agreement with the United States on voluntary safety measures, just as Russia and other nations have done in the context of ISS operations.

Planetary Protection and Defense

By "planetary protection," we refer to rules intended to protect other planets, not including Earth. We distinguish these from "planetary defense" rules, which are intended to preserve Earth, its environment, and its life forms, especially humans. Here we draw a distinction from certain sources, which use the terms planetary protection and defense interchangeably. International space law is relatively vague

⁹⁷ Koons et al. (2010) § 2.1 "Scope of the ISS Visiting Vehicle Requirements".

⁹⁸ Wolf Amendment (2011).

⁹⁹ Chunlai et al. (2019).

¹⁰⁰ CNSA (2021).

regarding whether and to what extent space operations like our hypothetical Venture must limit their activities in order to preserve the environments of other planets. However, any operation in space may still find itself subject to regulations and controls that will function like environmental protection laws on Earth. The strictest of these rules will probably apply during missions to planetary bodies or regions where there is a higher possibility of alien life or evidence thereof.

Planetary protection and defense rules derive from the Outer Space Treaty, Article IX of which requires state parties to avoid “harmful contamination” of celestial bodies during “exploration” and “adopt appropriate measures for this purpose.” The same article calls for the prevention of “adverse changes in the environment of Earth” caused by extraterrestrial matter.¹⁰¹ The Treaty does not further define or describe “harmful contamination” of other celestial bodies, but from the context it appears that the Treaty is primarily concerned with the prevention of interference with other state parties’ activities in space. It follows logically, for example, that contamination or destruction of a site containing evidence of extraterrestrial life might give rise to allegations of interference with others’ search for such evidence. This risk has led spacefaring states to adopt protection and mitigation measures, as discussed later in this section.

In the preceding section, we discussed Article IX’s requirement that state parties undertake “consultations” with other states if they believe that their activities on the Moon or elsewhere would cause “potentially harmful interference” with others’ activities.¹⁰² The same article requires state parties to act with due regard for the corresponding interests of others. This article, along with rest of the Treaty, explicitly concerns itself with potential risks to astronauts, spacecraft, and their activities—not the risk of altering natural environments. Where the Treaty mentions the danger of “adverse changes” in an environment, it limits that concern to Earth. The Treaty’s focus on protecting other spacefaring nations and their activities appears in other articles containing prohibitions against nuclear weapons and other weapons of mass destruction into space.¹⁰³ The Treaty also contains numerous provisions for the prevention of harm and damage to spacecraft and astronauts, as do other binding instruments of space law.¹⁰⁴ Besides the single use of “contamination” in Article IX, the Treaty does not address actions that might alter the natural environment of a celestial body.

Nevertheless, space-resource operations like our hypothetical Venture may find themselves subject to legal restrictions that limit and control their effects upon the natural environments of planets, moons, and other bodies in space. Spacefaring states, exercising supervision and jurisdiction over their nationals, will probably continue to create and enforce rules derived from the Policy on Planetary Protection published

¹⁰¹ Outer Space Treaty (1967) art. IX.

¹⁰² *Id.*

¹⁰³ Outer Space Treaty art. IV.

¹⁰⁴ Outer Space Treaty arts. VI, VII, VII; Liability Convention (1972); Registration Convention (1975).

by the Committee on Space Research (COSPAR), an international organization dedicated to promoting scientific research. COSPAR's Policy is designed to protect scientific investigations of possible extraterrestrial "life forms, precursors, and remnants," and to protect Earth from the potential hazards of invasive extraterrestrial life.¹⁰⁵ This policy adheres closely to Article IX because it discourages actions that might interfere with the scientific search for extraterrestrial life, and seeks to protect the Earth's environment from alien hazards. Neither of these objectives necessarily require space operations to preserve all extraterrestrial environments per se. Instead, spacefaring states and their nationals must avoid certain destructive or contaminant activities on or around certain planetary bodies or regions where opportunities for scientific research into alien life or its precursors seem most promising.

In furtherance of these goals, the Policy divides space missions into five categories (numbered I through V), largely based on the probability that the mission might encounter alien life or related evidence. For each category, the Policy establishes requirements to prevent spacecraft or crew from compromising future scientific research into life and its origins. Although the Policy provides categorical guidance for certain well-known destinations like the Moon, Mars, and Europa, it leaves the vast majority of possible mission destinations undefined and encourages members to use "the best multidisciplinary scientific advice," preferably from "Member National Scientific Institutions," to make category determinations for other targets.¹⁰⁶ In some cases, states have made their own determinations as to which planetary protection categories apply to certain types of missions. Therefore, a space-resource venture must be aware of not only the COSPAR recommendations for planetary protection, but also any rules imposed by the venture's supervising government.

Category I missions are the easiest, from a compliance perspective. COSPAR recommends no planetary protection requirements for such missions because their target bodies are not of "direct interest" to scientists for studying life or the origins/evolution thereof.¹⁰⁷ A space-resource venture aimed at extracting or using resources on or around such a target body would not have to worry about keeping its mining or prospecting equipment out of certain areas, sanitizing its vessel, or even keeping track of any biological material carried from Earth on its vessel. According to COSPAR, Category I destinations would include Jupiter's moon Io as well as undifferentiated metamorphosed asteroids.¹⁰⁸

The United States, through NASA, also regards certain areas of Earth's Moon as "Category I-L" destinations for space missions,¹⁰⁹ despite the COSPAR Policy's description of the Moon as a Category II target.¹¹⁰ NASA imposes no planetary

¹⁰⁵ COSPAR Policy (2020) 13.

¹⁰⁶ COSPAR Policy (2020) 13. COSPAR's Policy also advises that COSPAR itself may, upon request, consider providing advice on category determination in consultation with Member National Scientific Institutions and International Scientific Unions.

¹⁰⁷ COSPAR Policy 13–14.

¹⁰⁸ COSPAR Policy 16.

¹⁰⁹ NID 8715.128 (2020) 4.

¹¹⁰ COSPAR Policy 16.

protection requirements for a venture operating in these Category I-L areas. However, space-resource operations interested in prospecting or mining the permanently shadowed regions (PSRs) at the lunar poles should be aware that NASA's Category I-L does *not* extend to PSRs. Instead, NASA classifies PSRs, along with the entire lunar surface north of 86°N latitude and south of 79°S latitude, as "Category II-L" regions.¹¹¹ A venture would be required to provide NASA's Office of Planetary Protection with an inventory of biological materials that might be delivered from Earth to the II-L region by the spacecraft.¹¹²

As we see from the above example, Category II missions generally impose limited compliance requirements on a space-resource venture. According to COSPAR, Category II missions involve a target body where there is "significant" scientific interest regarding evidence of life or chemical evolutionary processes, *but* the chance of the spacecraft contaminating such evidence and compromising future investigations is "remote."¹¹³ For example, if a venture were to deploy a prospecting rover to a body like Venus, which holds significant interest for exobiologists, but where there is only a remote chance of that rover contaminating a scientifically important area, it would probably be a Category II mission. For such a mission, the venture need only submit documentation of a "short planetary protection plan," including intended and potential targets, strategies, and a post-encounter/end-of-mission report providing location data for any areas impacted by the rover.¹¹⁴ However, for certain Category II bodies like Jupiter's moon Ganymede or Saturn's moon Titan, COSPAR's Policy would add a requirement that the spacecraft operator provide documented analysis showing that its mission poses less than a one-in-ten-thousand chance of introducing a viable organism from Earth into any liquid-water environments on those bodies.¹¹⁵

The COSPAR Policy advises that missions to "small bodies" of the solar system generally fall under Categories I or II, described above. An asteroid mining venture could therefore expect relatively light planetary protection and defense compliance requirements, unless the target asteroid exhibited liquid-water saturation, organic materials, and/or conditions of temperature, radiation, and meteor bombardment that might increase the possibility of life developing. The Policy anticipates "[f]urther elaboration" of its requirements for small bodies, so would-be asteroid miners should continue to monitor developments at COSPAR and their national governments.¹¹⁶

Category III missions have higher compliance requirements, which apply to fly-by and orbital missions, not landers. These requirements involve target bodies where there is scientific interest in possible evidence of life or evolution *and* a significant chance of contamination.¹¹⁷ If a venture intends to fly by or orbit such a target, it will need to provide additional documentation beyond that which is required for Category

¹¹¹ NID 8715.128 4.

¹¹² NID 8715.128 4.

¹¹³ COSPAR Policy 14.

¹¹⁴ COSPAR Policy (2020) 14.

¹¹⁵ COSPAR Policy 16.

¹¹⁶ COSPAR Policy 21.

¹¹⁷ COSPAR Policy 14.

II missions, and it will need to implement additional procedures including trajectory biasing, cleanrooms, and possibly bioburden analysis and reduction techniques.¹¹⁸ These procedures are designed to prevent the spacecraft from impacting the surface or, if it does impact, prevent contamination. The COSPAR Policy states that these measures should reduce the probability of contamination to no more than a one-in-one-thousand chance during the 50 years after the mission arrives, but COSPAR leaves it to member states to make these calculations as they see fit.¹¹⁹ The Policy points to the United States' *Viking* mission of the late 1970s as an example of the sterilization protocols that operators should adopt, with some modifications.¹²⁰ According to COSPAR, Category III mission targets include Mars and Jupiter's moon Europa.

Category IV missions involve planned landings upon the same bodies, including Mars and Europa, for which Category III provided fly-by recommendations. The guidance for landers is once again intended to protect areas where there is significant interest in life or its origins/evolution *and* significant contamination risk.¹²¹ If a space-resource venture were to attempt to land in such an environment, it could expect to be subject to even more detailed documentation requirements including bioburden analysis, a probability of contamination analysis, organics inventory, and various other procedures involving cleanrooms, sterilization, and bioshields.¹²² There should be less than a one-in-one-thousand chance that the mission contaminates the planetary body within 50 years of its arrival, as calculated by the member state in its best judgment.¹²³ For missions in or near certain designated "special regions" on Mars, which are generally warmer and wetter than the rest of the planet and therefore more hospitable to life, the sterilization requirements will be even higher.¹²⁴ The COSPAR Policy also briefly addresses the possibility of a human mission to Mars, acknowledging that a "comprehensive planetary protection protocol" for such missions needs to be developed in the future.¹²⁵

Category V missions involve all spacecraft returning to the Earth–Moon system.¹²⁶ Space operators like our Venture should be aware that Category V requirements will generally apply, with some exceptions, to any mission returning to the Earth *or* the Moon.¹²⁷ The Moon falls within the protective bubble of Category V in order to avoid creating a need for planetary defense barriers between the Earth and the Moon.¹²⁸ As long as the Moon remains free of biological contaminants from the rest of outer space, missions can travel freely from the Earth to the Moon and

¹¹⁸ COSPAR Policy 14.

¹¹⁹ COSPAR Policy 16.

¹²⁰ COSPAR Policy 14, 16.

¹²¹ COSPAR Policy 14.

¹²² COSPAR Policy 14.

¹²³ COSPAR Policy 16.

¹²⁴ COSPAR Policy (2020) 17–18.

¹²⁵ COSPAR Policy 19.

¹²⁶ COSPAR Policy 14.

¹²⁷ COSPAR Policy 14.

¹²⁸ COSPAR Policy 14.

back without costly and inconvenient planetary defense safeguards. Similarly, other solar system bodies deemed by scientists to have “no indigenous life forms” can be assigned to the subcategory of “unrestricted Earth return,” and therefore avoid planetary defense requirements. Venus is one such body.¹²⁹

Other target bodies would fall under “restricted Earth return,” and thereby trigger the highest degree of protection if a venture attempted to return material or equipment from such a body into the Earth–Moon system.¹³⁰ Such a venture would need to take stringent safety precautions to prevent a destructive impact on the Earth or the Moon, and to maintain strict containment protocols for any extraterrestrial material and any equipment that contacted such material throughout the return phase of the mission.¹³¹ The highest containment and sterilization protocols would remain in effect after the return of the spacecraft and any samples to Earth, especially if a “nonterrestrial replicating entity” (i.e., a reproductive alien life form) is found.¹³²

In addition to the requirements described above, COSPAR also recommends that member states share other information regarding their planetary missions, including data about spacecraft, their locations and operations as well as bioburdens, sterilization activities, and related matters.¹³³ As with much else in spaceflight, a venture should expect little privacy or trade secrecy regarding its operations in and around planetary bodies.

New assessments and regulations from state authorities will have a significant impact on the operations of space-resource operations going forward. In the United States, for example, government policy calls for the development of a new “forward contamination risk assessment framework,” to analyze and systematize the need for planetary protection in future missions.¹³⁴ This framework may lead to the creation or revision of guidelines and requirements for missions, and could cause changes in the payload review process by which US government agencies determine, among other things, whether a venture’s proposed mission should be approved for launch and flight.¹³⁵ Similar analyses and new requirements may follow for planetary defense of Earth.¹³⁶ Regulators in the United States have acknowledged that the private sector should be involved in the drafting of these policies, and are soliciting relevant feedback from the commercial spaceflight industry. Space-resource operators may therefore find that they are able to provide comments on the necessity and feasibility of planetary protection and defense rules through the National Science and Technology Council of the US government’s Office of Science and Technology Policy.¹³⁷

¹²⁹ COSPAR Policy 16.

¹³⁰ COSPAR Policy 14.

¹³¹ COSPAR Policy 14–15.

¹³² COSPAR Policy 15.

¹³³ COSPAR Policy 15.

¹³⁴ Nat’l Space Council (2020).

¹³⁵ Nat’l Space Council 5.

¹³⁶ Nat’l Space Council 6.

¹³⁷ Nat’l Space Council (2020) 6.

Rescue and Return of Astronauts

Like planetary protection and defense, the obligation to assist astronauts in distress will probably endure as a stable and predictable obligation under international and national law. The state parties have agreed, in the Outer Space Treaty as well as the Rescue Agreement, to protect each other's astronauts from harm and to render aid and assistance to them when possible.¹³⁸ These international obligations could lead a government to require a space-resource venture to assist astronauts in distress when they are reasonably able to do so. Such a requirement would be a logical extension of similar laws on Earth, which require mariners to assist individuals at sea who are in danger.¹³⁹

Preservation of Historic Sites

Rules protecting historic sites from disturbance exist, but they are largely limited to the national laws of spacefaring states because there is not yet any international law on the subject. The Outer Space Treaty's prohibition against appropriation of territory in space prevents nations from drawing internationally binding boundaries to protect the territory around historic landing and impact sites, as they might do around a national monument or park. However, spacefaring nations can, and have already, created self-limiting laws and regulations to prevent their national space agencies and private space operations under their control from disturbing historic landing sites. For instance, the United States classifies "Apollo landing and other historic sites" on the Moon under the same Category II-L as the PSRs described above, because of these sites' "historical and scientific value."¹⁴⁰ US law requires the NASA Administrator to require that contractors and partners of the US space program adhere to NASA's best practices for protecting and preserving lunar artifacts. The Administrator may waive this requirement, but only to allow activities in the historical sites that have "significant historical, archaeological, anthropological, scientific, or engineering value."¹⁴¹ Furthermore, the United States and its partner nations under the Artemis Accords have pledged to "preserve outer space heritage," including historically significant landing sites and "other evidence of activity."¹⁴² Although the Accords leave this category of "historically significant" sites undefined, a space-resource venture should expect a heightened level of scrutiny and control if its planned activities bring it into the vicinity of any previous mission's area of operations.

¹³⁸ Outer Space Treaty (1967) art. V; Rescue Agreement (1968).

¹³⁹ See, e.g., Marine Casualty Assistance, 46 U.S.C. §§ 2303–04.

¹⁴⁰ NID 8715.128 (2020) 4.

¹⁴¹ One Small Step Act (2020).

¹⁴² Artemis Accords (2020) § 9.1.

The Moon Agreement and an International Regime

A decade after the Outer Space Treaty was signed, certain members of the UN Committee on the Peaceful Use of Outer Space sought to remedy what they saw as deficiencies in the Treaty by proposing what became the Moon Agreement of 1979. This agreement would prohibit the private acquisition of property upon celestial bodies, except where such acquisition was subject to the control of a to-be-created international regime with effective sovereignty over all celestial bodies in the solar system or, perhaps, the universe.¹⁴³ In stark contrast to the vague principles espoused in the Outer Space Treaty, the Moon Agreement was more explicit about property and commerce. The Agreement provided that, until an international legal regime is established to govern the exploitation of lunar resources, “[n]either the surface nor the subsurface of the [M]oon, nor any part thereof or natural resources in place, shall become property of any State, international intergovernmental or nongovernmental organization, national organization or non-governmental entity or of any natural person.”¹⁴⁴

The Moon Agreement is frequently cited as an attempt to entirely prohibit the acquisition and use of space resources, but there remains a possible interpretation of the Agreement that would make its prohibition more narrow. The Agreement states that it seeks to prevent “natural resources in place” from becoming property. By specifying only that natural resources “in place” should not be owned, the Moon Agreement’s drafters might have only meant that nobody should declare ownership of all resources connected with a given parcel of land. Such a prohibition would be a matter of real property law, where grants of title can convey exclusive ownership of land as well as resources attached to it. Such a prohibition on real property would amount to no more than a restatement of Article II of the Outer Space Treaty, which effectively prohibits the acquisition of real property in space by prohibiting claims of national sovereignty—a.k.a., the type of jurisdiction necessary for states to establish real property rights for their nationals. If read so narrowly, the Moon Agreement would not prohibit a space-resource venture from picking up resources where it finds them, in land claimed by no one, thus turning them into moveable property through the act of collection just as fishing operations in international waters do, as discussed in Sect. 34.2.2.1. This interpretation, however, is controversial.¹⁴⁵

Regardless of what the Moon Agreement was meant to prohibit, the vast majority of nations, and all independently spacefaring nations, declined to ratify it.¹⁴⁶ Leaders in the United States, for example, were concerned that the Moon Agreement might create a prohibition against space-resource utilization, and voted against ratification

¹⁴³ Moon Agreement (1979) art. 11; see also Gangale (2009) 67–88.

¹⁴⁴ Moon Agreement art. 11 ¶ 3.

¹⁴⁵ Tronchetti (2009) 55.

¹⁴⁶ Treaties Status (2020).

because they did not want to give up the rights of use that their nationals might someday enjoy.¹⁴⁷ Only eighteen states have signed and ratified the Agreement.¹⁴⁸

In the absence of official support for the Moon Agreement, several individuals and non-governmental organizations have expressed support for a new international effort to establish the kind of regime contemplated in the Agreement.¹⁴⁹ These include the Hague ISRWG, which has worked since 2016 to assess the need for an international space-resource governance framework and lay the groundwork for such framework. The Working Group has published a set of “Building Blocks” for the development of an international framework on space-resource activities, which it encourages nations, international organizations, and non-governmental organizations to consider and use.¹⁵⁰ These Building Blocks generally discuss the sorts of issues that a future framework of space-resource rules might address, without establishing what those rules should be.

The Outer Space Institute, an organization of multi-disciplinary academics based in Canada, has taken a stronger stance, publishing the Vancouver Recommendations on Space Mining, which criticized “unilateral” national legislation on space resources as “inadequate” and recommended that nations enter into multilateral talks to establish an international regime for space mining.¹⁵¹ The Institute made similar arguments in an International Open Letter on Space Mining to the President of the UN General Assembly.¹⁵²

None of these efforts appear likely to generate laws or rules affecting space-resource operations in the near term. The governments adopting national space-resource policy have almost uniformly gone the other way, establishing their own national space-resource laws or participating in efforts to develop and use space resources with other nations. However, a space-resource venture should remain aware of developments in these non-governmental and intergovernmental fora. If an international regime were to receive support and consensus among state authorities, it could dramatically impact the space-resource industry.

One example of the kind of international regime that has been proposed for space-resources operations is the International Seabed Authority, established under UNCLOS.¹⁵³ Under this international framework, which has existed for several decades but not yet been enforced, the Authority would effectively control all territory and resources in the deep seabed beneath international waters. Commercial operators wishing to mine resources in this area would need to seek approval from the Authority, pay fees up front, deliver a share of any resources collected, hand over prospecting data on areas identified as potential mining sites (half of which the Authority would claim for itself), and license their technology to the Authority.

¹⁴⁷ Moon Treaty Hearings (1980) 83–85.

¹⁴⁸ Treaties Status.

¹⁴⁹ See, e.g., Tronchetti 233–90.

¹⁵⁰ Building Blocks (2019).

¹⁵¹ Vancouver Recommendations (2020).

¹⁵² OSI Open Letter (2020).

¹⁵³ See, e.g., Kerrest (2019).

The Authority would use the funds, resources, data, and technology provided by the other miners to establish a UN-owned and operated mining company, known as The Enterprise, which would compete with the miners for use of these resources.¹⁵⁴

34.3 National Legislation, Regulation, and Policy

Many countries have near-term plans for planetary exploration, including long-established spacefaring powers like the United States, Russia, and China, as well as newer spacefaring nations like India and the United Arab Emirates. Major differences between current plans and previous space activities include a greater extent of commercial involvement and more specific interest in using space resources. The official space agencies leading these efforts have made it clear that a major focus of initial efforts will be the use of lunar resources to facilitate further exploration of the Moon and other destinations in the solar system. In support of these efforts, numerous commercial entities have already been engaged or signed agreements to participate in the new phase of space exploration.¹⁵⁵

The United States, Luxembourg, the United Arab Emirates, and Japan have taken legislative action to facilitate space-resource use by commercial entities. Other nations have signed agreements or made collaborative plans with these countries, reflecting either explicit or tacit agreement with their approach to space resources. On the other hand, some nations have publicly avoided the subject or expressed concerns about space-resource utilization. Most objections have come from individual state officials making informal remarks, but some nations have issued formal statements of concern in international fora like the United Nations. Regardless of whether they have taken an official position for or against space-resource utilization, most spacefaring nations are preparing for space-resource development either directly or through partnerships. Additionally, several nations that have not yet developed independent space launch capabilities are signing agreements for future partnerships with established space powers. These nations could exert diplomatic influence over international space law even before they become major participants in the space-resource industry.

¹⁵⁴ UNCLOS (1982) arts. 140–170.

¹⁵⁵ See, e.g., CLPS (2021) listing “Astrobotic Technology, Blue Origin, Ceres Robotics, Deep Space Systems, Draper, Firefly Aerospace, Intuitive Machines, Lockheed Martin Space, Masten Space Systems, Moon Express, Orbit Beyond, Sierra Nevada Corporation, SpaceX, [and] Tyvak Nano-Satellite Systems” as providers for the NASA Commercial Lunar Payload Services program.

34.3.1 *Australia*

Australia is one of the original eight signatories to the Artemis Accords.¹⁵⁶ By joining the Accords, Australia apparently signaled its agreement to the Accords' assertion that "the extraction of space resources does not inherently constitute national appropriation under Article II of the Outer Space Treaty."¹⁵⁷ If the Accords' assertion does indeed reflect Australia's position regarding space resources, the country's accession is particularly noteworthy because it is also one of the eighteen state parties to the Moon Agreement of 1979, which would arguably limit or prohibit the commercial use of space resources until an international regime is established to govern such activity.¹⁵⁸ It is not clear whether, by joining the Accords, Australia intends to withdraw from the Moon Agreement or reinterpret its obligations in such a way as to allow for participation in the Artemis Program.

34.3.2 *Belgium*

Belgium has ratified the Moon Agreement. In January 2019, Belgium signed a joint declaration with Luxembourg, committing to collaboration on the development of an international framework for the exploration and use of space resources.¹⁵⁹ Later that year, Belgium's delegation to the UN Committee on the Peaceful Uses of Outer Space (COPUOS) submitted a proposal along with Greece, recommending that the COPUOS Legal Subcommittee establish a working group to explore the establishment of an international regime to govern space-resource activities.¹⁶⁰ The recommended working group has not yet been created as of this writing.

34.3.3 *Brazil*

Brazil's Minister of Science, Technology, and Innovation has signed a Statement of Intent with NASA indicating Brazil's intent to join the Artemis Accords.¹⁶¹

¹⁵⁶ NASA, Press Release (Oct. 13, 2020).

¹⁵⁷ Artemis Accords (2020) § 10.2.

¹⁵⁸ See Sect. 34.2.2.2, above.

¹⁵⁹ Ministry of the Economy–Luxembourg, Press Release (Jan. 23, 2019).

¹⁶⁰ Belgium/Greece (2019).

¹⁶¹ NASA, Press Release (Dec. 14, 2020).

34.3.4 *Canada*

Canada is one of the original eight signatories to the Artemis Accords.¹⁶² The country has also committed to provide a robotic arm for the Gateway, a space station proposed to facilitate the Artemis Program, and has reached an agreement with NASA to send Canadian astronauts on Artemis missions to the Moon.¹⁶³

34.3.5 *China*

Although China has not established a firm position on the legality of space-resource utilization, the country's actions through its China National Space Administration (CNSA) and Chinese Lunar Exploration Program (CLEP) indicate that it will be a major participant in the industry. Through CLEP's ongoing *Chang'e* missions, which have already collected resources from the Moon for scientific study, the country plans to use and develop lunar resources while building a base on the Moon's south pole.¹⁶⁴ China has signed a Memorandum of Understanding with Russia, announcing its intent to cooperate on the construction of an International Lunar Research Station on or around the Moon.¹⁶⁵

The Chinese government is also allowing its private space sector to enter the space-resource industry. A private Chinese company, Origin Space, based in Shenzhen, has raised funds and contracted with manufacturers to develop technology for asteroid mining.¹⁶⁶

China is not involved in the Artemis Program or the Artemis Accords, due in no small part to legal and diplomatic issues exemplified by the Wolf Amendment, a US law that places strict controls on collaboration between US and Chinese space activities. However, the country is collaborating with several international partners on its space-resource activities. These include Germany, Sweden, and the Netherlands, all of which contributed equipment to the *Chang'e* 4 mission to the Moon, as well as Russia, Turkey, Ethiopia, and Pakistan, all of which have signed agreements to work with China on future lunar or asteroid exploration missions.¹⁶⁷

¹⁶² NASA, Press Release (Oct. 13, 2020).

¹⁶³ Foust, (Dec. 16, 2020a).

¹⁶⁴ Chunlai et al. (2019).

¹⁶⁵ CNSA (2021).

¹⁶⁶ Jones (2020).

¹⁶⁷ Chunlai et al.

34.3.6 *France*

Although France is home to a robust space industry, including major commercial spaceflight companies, the country has not established a firm position on space-resource utilization. France has not joined the Artemis Accords, nor has it ratified the Moon Agreement. The country did sign the Moon Agreement, but unlike the Outer Space Treaty and other foundational treaties of international space law, France ultimately decided not to ratify the Agreement.¹⁶⁸

34.3.7 *Germany*

Germany, which also boasts a robust space industry, has similarly not established a firm position on space-resource utilization. The country is not an Artemis Accords signatory, nor did it sign or ratify the Moon Agreement. However, the country's official space administration has signed a memorandum of understanding with the Luxembourg Space Agency to cooperate in future efforts to explore and use space resources.¹⁶⁹ The country has also participated in China's CLEP by providing equipment that flew to the Moon on the *Chang'e 4* lander.¹⁷⁰

34.3.8 *Greece*

Greece's delegation to the UN COPUOS submitted a proposal in 2019 along with Belgium, recommending that the COPUOS Legal Subcommittee establish a working group to explore the establishment of an international regime to govern space-resource activities.¹⁷¹ The recommended working group has not yet been created as of this writing.

Greece, unlike Belgium, has neither signed nor ratified the Moon Agreement.¹⁷²

34.3.9 *India*

Although India has not yet established a firm position on space-resource utilization, the nation's official space agency, the Indian Space Research Organization (ISRO)

¹⁶⁸ Treaties Status (2020).

¹⁶⁹ Ministry of the Economy–Luxembourg, Press Release (Oct. 23, 2019).

¹⁷⁰ Wimmer-Schweingruber et al. (2016).

¹⁷¹ Belgium/Greece (2019).

¹⁷² Treaties Status.

has played a crucial role in detecting resources, especially water, on the Moon.¹⁷³ The Chairman of ISRO expressed the Organization's ambition to lead the international effort to mine helium-3 from the Moon and return it to Earth for energy uses.¹⁷⁴

34.3.10 Italy

Italy is one of the original eight signatories to the Artemis Accords.¹⁷⁵

34.3.11 Japan

The Japanese government has passed a bill similar to the space-resource laws in the United States, Luxembourg, and UAE.¹⁷⁶ The country is also one of the original eight signatories to the Artemis Accords.¹⁷⁷ The Japan Aerospace Exploration Agency (JAXA) has already successfully conducted two missions to recover asteroid materials from outer space and bring them to Earth for scientific uses.

34.3.12 Luxembourg

The Grand Duchy passed the 2017 Luxembourg Space Law, which provides that space resources “are capable of being owned,” and which sets out requirements for companies wishing to engage in the use of such resources.¹⁷⁸ The nation passed this law shortly after establishing the Luxembourg Space Agency, a commerce-focused government organization dedicated to leading Luxembourg's effort to become a European hub for space-resource utilization. Luxembourg has partnered with the European Space Agency (ESA) to establish a European Space Resources Innovation Center (ESRIC), which is designed to support scientific research and commercial opportunities.¹⁷⁹ Luxembourg has also signed memoranda of understanding and joint declarations regarding space-resource utilization with at least seven other nations, including Belgium, China, the Czech Republic, Germany, Poland, the United Arab Emirates, and the United States.

¹⁷³ JPL (2018).

¹⁷⁴ Kotoky (2018).

¹⁷⁵ NASA, Press Release (Oct. 13, 2020).

¹⁷⁶ *The Japan Times* (2020).

¹⁷⁷ NASA, Press Release (Oct. 13, 2020).

¹⁷⁸ 2017 Luxembourg Space Law.

¹⁷⁹ The Luxembourg Government, Press Release (Nov. 18, 2020).

Luxembourg is also one of the original eight signatories to the Artemis Accords.¹⁸⁰

34.3.13 *Netherlands*

The Netherlands is one of eighteen nations that ratified the Moon Agreement of 1979.¹⁸¹ This ratification might indicate that the country supports prohibitions against ownership of space resources, at least until the establishment of an international regime. However, the country is also a participant in China's CLEP, having contributed equipment to the *Chang'e 4* lunar lander.¹⁸²

The Netherlands also hosts, through the University of Leiden, the Hague ISRGWG, discussed in Sect. 34.2.2.2, above.

34.3.14 *Russia*

Russia has not yet established a firm position on space-resource utilization, though it is actively discussing plans to begin such activity. High-ranking Russian government officials reached out to Luxembourg in January 2019, hoping to enter into a “framework agreement on cooperation in the use of mining exploration in space.”¹⁸³ However, in June of the same year, Russia submitted a working paper to the UN COPUOS, in which it decried “unprecedented legislative action” in “some jurisdictions regarding space resources” which “create a major legal ambiguity” regarding “the fundamental norm prohibiting national appropriation of outer space.”¹⁸⁴ At the time, the only nations that had enacted legislation on space-resource rights were Luxembourg, to which Russia had just made overtures, and the United States. In the same 2019 COPUOS session, the Russian delegation advocated the establishment of a legally binding regime to govern the development of space resources, but in doing so may have implicitly conceded that international law, as it stands, lacks a general prohibition against commercial use of resources in outer space.¹⁸⁵ Russia also signed an agreement with China in 2019 to provide support for CLEP, including research and development of space-resource utilization capabilities.¹⁸⁶

The following year, Russia's top space officials apparently objected to the United States' plans for commercial use of space resources through the Artemis Program.

¹⁸⁰ NASA, Press Release (Oct. 13, 2020).

¹⁸¹ Treaties Status (2020).

¹⁸² Radboud University (2018).

¹⁸³ Soldatkin (2019).

¹⁸⁴ Russian-UN (2019).

¹⁸⁵ Listner 34–35 (2019).

¹⁸⁶ Chunlai et al. (2019).

When a US Executive Order reaffirmed the United States' commitment to space-resource utilization,¹⁸⁷ a deputy director general of the Russian official space agency, Roscosmos, accused the United States of attempting to “expropriate outer space” and making “aggressive plans to actually seize territories of other planets.” The same official warned that “everyone remembers the outcome” of examples where a country “start[s] to seize territories for its own benefit.”¹⁸⁸ Roscosmos Director General Dmitry Rogozin echoed those comments, describing “any attempts to privatize the Moon” as “illegal” and “counter to international law.”¹⁸⁹ Later in the same year, Rogozin criticized the United States' Artemis Program as “too U.S.-centric,” but left open the possibility of participating through cooperating on efforts to build Artemis's Gateway space station.¹⁹⁰

Despite Russia's expression of concerns about resource exploitation and its calls for international regulation of the activity, the country has not signed or ratified the Moon Agreement, which would address both issues. In 2021, Russia signed a memorandum of understanding with China reflecting the two countries' plans to build an International Lunar Research Station on or around the Moon, and to engage in acquisition and use of lunar resources.¹⁹¹

34.3.15 *Ukraine*

Ukraine is the ninth signatory to the Artemis Accords.¹⁹²

34.3.16 *United Arab Emirates*

The 2019 UAE Space Law provides a legal framework for the acquisition and use of resources in space.¹⁹³ This law followed the establishment of the country's National Space Policy in 2016, which included an objective of “becoming an integral international partner in the study and exploration of the Solar System including exploration of available resources there.”¹⁹⁴ Through its UAE Space Agency and other relevant government authorities, the country seeks to become a “regional and global hub” for a variety of space activities, including “[e]xploration, mining, extraction and

¹⁸⁷ Exec. Order (2020).

¹⁸⁸ Roscosmos, Press Release (Apr. 7, 2020).

¹⁸⁹ Tass (May 25, 2020).

¹⁹⁰ Foust (Oct. 12, 2020b).

¹⁹¹ CNSA (2021).

¹⁹² US Embassy in Ukraine, Press Release (Nov. 17, 2020).

¹⁹³ 2019 UAE Space Law.

¹⁹⁴ UAE Space Policy (2016) 44.

utilization of resources in space.”¹⁹⁵ The UAE hopes to achieve these goals not only through its official space agency, but also by promoting “creative entrepreneurship and commercial space projects,” which will include commercial missions involving “the exploration and exploitation of resources in space.”¹⁹⁶

The UAE has signed a memorandum of understanding with Luxembourg regarding cooperation in space activities, with a particular focus on the exploration and utilization of space resources.¹⁹⁷

The UAE is also one of the original eight signatories to the Artemis Accords.¹⁹⁸

34.3.17 United Kingdom

The United Kingdom is one of the original eight signatories to the Artemis Accords.¹⁹⁹

34.3.18 United States

NASA’s decision to solicit and accept offers from contractors for the purchase of lunar regolith samples illustrates a consistent US government policy to authorize and encourage private entities to acquire and use space resources.²⁰⁰ Other examples include the 2015 U.S. SPACE Act, a bipartisan law enacted during the administration of President Barack Obama, and similar national space policies implemented throughout the administration of President Donald Trump.²⁰¹ President Joseph Biden’s administration has indicated that it will largely continue the space policies of its predecessors, including the Artemis Program.²⁰²

The agencies that implement these policies in the United States include NASA itself as well as other federal authorities that would control or influence a space-resource venture’s ability to conduct its missions. These include the U.S. Federal Aviation Administration (the FAA), from which a venture must obtain a license for any space launch, and the Federal Communications Commission (the FCC), which controls licenses for use of the radio spectrum. Likewise, the Department of Commerce (the DoC) can influence or control space operations through a variety of regulations governing, among other things, the use of cameras and sensors in

¹⁹⁵ *Id.* 44–45.

¹⁹⁶ UAE Space Policy (2016) 48.

¹⁹⁷ Press Release, Ministry of the Economy – Luxembourg (Oct. 10, 2017).

¹⁹⁸ NASA, Press Release (Oct. 13, 2020).

¹⁹⁹ *Id.*

²⁰⁰ Bridenstine (2020).

²⁰¹ *See, e.g.*, U.S. Space Policy (2020) 3, 5, 13.

²⁰² Davenport (2021).

space. These authorities consult other US government agencies, including NASA, the Department of Defense (the DoD), the State Department, and the U.S. Geological Survey, when deciding whether to grant or withhold licenses for space operations. For instance, the State Department may intervene with another agency to block a planned launch if it determines that a venture's proposed activity poses a risk to the United States' foreign policy interests.²⁰³ Therefore, it is theoretically possible that a future administration could reinterpret international space law to include certain prohibitions against space-resource commerce and therefore deny space-related licenses for resource activity.

Nevertheless, the 2015 U.S. SPACE Act provides a "sense of Congress" that a US citizen's right to "possess, own, transport, use, and sell asteroid resources or space resources" is not a violation of the United States' international obligations under Article II of the Outer Space Treaty.²⁰⁴ The FAA has also been encouraged, through federal law, to issue launch licenses in such a manner as to "encourage, facilitate, and promote" commercial spaceflight.²⁰⁵ This legislative guidance should provide some comfort to a space-resource venture, but if the venture is still concerned about whether its space-resource utilization plans will be approved by regulators, it may request a payload review from the FAA prior to or independently of any launch license application.²⁰⁶

Through the Artemis Program and Artemis Accords, as well as laws, directives, and policies, the US government has committed to facilitating commercial use of resources in space and seeking international support and participation in such efforts in order to enable exploration of the solar system. These efforts have led 23 nations to sign the Artemis Accords, with others expressing intent to join. The European Space Agency, though not an Artemis Accords signatory, has agreed to collaborate with NASA on building the Gateway space station to facilitate Artemis missions.²⁰⁷

34.4 Conclusion

A space-resource venture could be reasonably pleased by recent developments in law and policy because several spacefaring states are actively encouraging the development of a commercial space-resource industry, and other states have signed agreements to participate in such efforts. Nevertheless, legal requirements will have a constant and thorough influence on space-resource operations. Indeed, a mining operation on the Moon, though it would be located more than 230,000 miles (380,000 km) away from its government's territory, might find itself under stricter controls than miners operating within a few minutes' drive of a government office. Such is the

²⁰³ 51 U.S.C. § 50905(b)(2)(B).

²⁰⁴ 2015 U.S. SPACE Act.

²⁰⁵ 51 U.S.C. § 50903.

²⁰⁶ Payload Review, 14 C.F.R. § 415.57(a).

²⁰⁷ ESA, Press Release (Dec. 27, 2020).

nature of spaceflight, where the distances from authorities are vast, but the state's duty to supervise is omnipresent.

As we have seen, many of the laws, policies, and practices governing space-resource utilization are yet to be determined. There are advocates for the creation of an international regime to govern all such activity. However, the current trend indicates that states will continue to craft the first space-resource rules independently or in consultation with their space exploration partners, while adhering to the principles outlined in space-law treaties. Due to technological advances that have rapidly democratized access to space, the number of states involved in space resources is already far larger than those involved in the original "Moon race," during which the Outer Space Treaty was signed. States as diverse as China and Ukraine, India and Italy, the UAE and Ethiopia have all taken steps to involve themselves in the use of space resources.

These states and the commercial entities they supervise will establish rules and norms, balancing competing interests like safety, freedom, prosperity, equality, efficiency, and sustainability just as societies have done on Earth since time immemorial. Many of these rules and norms will depend upon scientific and technical developments that are impossible to predict. For instance, if a venture manages to locate a proven reserve of valuable and feasibly extractable resources, it could have a dramatic effect on rules and policies for future prospecting and mining. Likewise, the discovery of extraterrestrial life or evidence thereof could have serious implications in several fields of human society, not least of which would be space-resource exploration.

Of course, all of this will remain largely theoretical until a space-resource operation like our hypothetical Venture manages to acquire some lunar or asteroid material and sell or use it under the color of law. As noted above, there is an acute shortage of positive laws, customs, and legal precedents surrounding the acquisition and use of resources in space. Scholars and commentators continue to try to fill in the gaps with proposals and theories for the governance of such activity, and authorities have offered various forms of regulation and encouragement. In order for any of these legal possibilities to become reality, someone needs to get up there and start digging.

References

- Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, Dec 5, 1979, 610 U.N.T.S. 205 (Moon Agreement).
- Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space, Apr 22, 1968, 19 U.S.T. 7570, 672 U.N.T.S. 119 (Rescue Agreement).
- Amos, Jonathan. 2016. *Luxembourg to support space mining*, BBC, Feb 3, 2016.
- Bagnell v. Broderick*, 38 U.S. 436 (1839).
- Bernhard, Schmidt-Tedd and Alexander Soucek. 2020. *Registration of space objects*. Oxford Research Encyclopedia—Planetary Science.
- Black's Law Dictionary (11th ed. 2019).
- Bridenstine, Jim. 2020. *Space resources are the key to safe and sustainable lunar exploration*. NASA Administrator Jim Bridenstine, Sep. 10, 2020. <https://blogs.nasa.gov/bridenstine/2020/09/10/space-resources-are-the-key-to-safe-and-sustainable-lunar-exploration/>.

- Building Blocks for the Development of an International Framework on Space Resource Activities (Int'l Space Res. Governance Working Grp., Nov 2019) (Building Blocks).
- Christol, Carl Q. 1984. Article 2 of the 1967 principles treaty revisited. *Annals of Air and Space Law* 9:217–44 (Nicolas Mateesco Matte ed.).
- Chunlai, et al. 2019. *China's present and future exploration program*. Science, July 19, 2019a.
- Claims—How Located, Colo. Rev. Stat., Tit. 34 Art. 43 (2018).
- Comm. on the Peaceful Uses of Outer Space, Legal Subcomm. Working Paper by Belgium and Greece, Apr 1–12, 2019, U.N. Doc. A/AC.105/C.2/L.311 (Belgium/Greece).
- Convention on International Liability for Damage Caused by Space Objects, Mar 29, 1972, 24 U.S.T. 2389, 861 U.N.T.S. 187 (Liability Convention).
- Convention on Registration of Objects Launched into Outer Space, Jan 14, 1975, 28 U.S.T. 695, 1023 U.N.T.S. 15 (Registration Convention).
- Curry, Steven. 2007. *Sovereignty and jurisdiction*. Encyclopedia of Law & Society: American and Global Perspectives (David S. Clark, ed.).
- Davenport, Christian. 2021. *The Biden Administration has set out to dismantle Trump's legacy, except in one area: space*. The Washington Post, Mar 2, 2021.
- Deep Seabed Hard Mineral Resources Act, codified at 30 U.S.C. § 1401 et seq. (1980).
- Dep't. of Energy & NASA, Request for Proposal No. TBD, Fission Surface Power Project (2020). https://beta.sam.gov/opp/24744267c3794298bd7ffce0db947b62/view?keywords=INL-21-002&sort=-relevance&index=&is_active=true&page=1 (RFP).
- Dep't of Def. & Full-Year Continuing Appropriations Act, Pub. L. 112–10 § 1340(a) (2011) (Wolf Amendment).
- Diplomatic Relations Act, codified at 22 U.S.C. § 254 (1978).
- Draft resolution of the IISL concerning the legal status of celestial bodies. In *Proceedings, 8th colloquium on the law of outer space* (Am. Inst. of Aeronautics & Astronautics 1965) (IISL Draft).
- Duties related to marine casualty assistance and information, 46 U.S.C. § 2303.
- Exec. Order on Encouraging International Support for the Recovery and Use of Space Resources (Apr 6, 2020).
- Fishermen's Protective Act, codified at 22 U.S.C. § 1972 (1967).
- Foust, Jeff. 2020a. *Canadian astronaut to fly on first crewed Artemis mission*. SpaceNews, Dec 16, 2020.
- Foust, Jeff. 2020b. *Russia skeptical about participating in lunar Gateway*. SpaceNews, Oct. 12, 2020.
- G.A. Res. 62/101 (Jan. 10, 2008).
- Gangale, Thomas. 2009. *The development of outer space: sovereignty and property rights in international law* (Praeger 2009).
- Gorove, Stephen. 1972. Criminal jurisdiction in outer space. *Int'l Lawyer* 6 (2): 313–323.
- Hitches, Theresa. 2020. *Cyber Attacks most likely space threat: Maj. Gen. Whiting*. Breaking Defense, Sep 16, 2020.
- Hobe, Stephan, et al. 2016. Background Paper, Does International Space Law Either Permit or Prohibit the Taking of Resources in Outer Space and on Celestial Bodies, and How is this Relevant for National Actors? What is the Context, and What are the Contours and Limits of this Permission or Prohibition? (IISL Inst. Of Studies 2016).
- Hofacker, Cat. 2020. *NASA plans to pay companies to extract tablespoons of lunar regolith*. Aerospace America, Sep. 10, 2020b.
- Int'l Law Comm'n, Report to the Gen. Assembly, U.N. Doc. A/61/10 (2006) (ILC Report).
- Jones, Andrew. 2020. *Chinese space resource utilization firm Origin Space signs deal for space telescope*. SpaceNews, Apr 23, 2020.
- Act on the Promotion of Business Activities for the Exploration and Development of Space Resources, Act No. 82 of 2021 (Japan) (2021 Japan Space Law).
- Jurisdictional, Preliminary, and Procedural Concerns*. Benchbook on International Law (Diane Marie Amann ed., Am. Soc'y Int'l L. 2014) (Benchbook on Int'l Law).

- Kerrest, Armel. 2019. *Contribution of the deep seabed mining legal regime to space resource activities*. IISL European Centre for Space Law Symposium.
- Kminek, et al. 2017. Planetary Protection Policy, (COSPAR Dec 2017) (COSPAR Policy).
- Koons, Diane S., et al. 2010. *Risk mitigation approach to commercial resupply to the international space station* (NASA 2010).
- Kotoky, Anurag. 2018. *ISRO in quest to find a trillion-dollar nuclear fuel on the Moon*. The Hindustan Times, Jun 27, 2018.
- Law of 20 July 2017 on the Exploration and Use of Space Resources, Official Gazette of the Grand Duchy of Luxembourg, No. 674 (2017 Luxembourg Space Law).
- Listner, Michael J. 2019. *The Précis Issue XXXIII* (Jun 28, 2019) (on file with author).
- Lyll, Francis and Paul B. Larsen. 2018. *Space law: a treatise*, 2nd ed. Routledge.
- Markoff, Marco G. 1970. *Space resources and the scope of the prohibition in article II of the 1967 treaty*. In *Proceedings, 13th colloquium on the law of outer space* (Am. Inst. of Aeronautics & Astronautics).
- McDowell, Andrea G. 2002. From commons to claims: property rights in the California gold rush. *Yale JL & Humanities* 14: 1.
- Merriam-Webster.com Dictionary*, s.v. “province”. <https://www.merriam-webster.com/dictionary/province>. Accessed 16 Apr 2021
- Moon treaty: hearings before the subcommittee on science, technology, and space of the committee on commerce, science, and transportation*, 96th Congress (1980) (Moon Treaty Hearings).
- NASA, Commercial Lunar Payload Services. <https://www.nasa.gov/content/commercial-lunar-payload-services> (CLPS). Accessed 16 Apr 2021
- NASA et al. 2020. Artemis Accords.
- NASA Interim Directive 8715.128, Planetary Protection Categorization for Robotic and Crewed Missions to Earth’s Moon (Jul 9, 2020) (NID 8715.128).
- NASA Jet Propulsion Laboratory, *Ice Confirmed at the Moon’s Poles* (Aug 20, 2018). <https://www.jpl.nasa.gov/news/ice-confirmed-at-the-moons-poles> (JPL). Accessed 2 May 2021
- NASA Policy Directive 8020.7G, Biological Contamination Control for Outbound and Inbound Planetary Spacecraft (May 17, 2013) (NPD 8020.7G).
- National Space Council, National Strategy for Planetary Protection (Dec 2020).
- National Space Policy of the United Arab Emirates (Sep 2016) (UAE Space Policy).
- National Space Policy of the United States of America (Dec 9, 2020) (U.S. Space Policy).
- One Small Step to Protect Human Heritage in Space Act of 2020, Pub. L. 116–275 (One Small Step Act).
- Open Letter to Tijjani Muhammad-Bande, President, UN General Assembly, Re: Multilateral Agreement on Space Resource Utilization (Outer Space Inst. Aug 2020). <http://www.outerspaceinstitute.ca/docs/InternationalOpenLetterOnSpaceMining.pdf> (OSI Open Letter).
- Payload Review, 14 C.F.R. § 415.57.
- Permanent Mission of Israel to the United Nations (Vienna), “Information furnished in conformity with General Assembly resolution 1721 B (XVI) by States launching objects into orbit or beyond: Registration information on Beresheet (international designator 2019–009B),” (Aug 6, 2019) A/AC.105/INF/436 (Israel Registry).
- Permanent Mission of the Russian Federation to the United Nations (Vienna), “Survey of the problem of discretion exercised by States in interpreting basic legal principles and norms related to safety and security in outer space” UN COPUOS, 62nd Session (Feb 4, 2019) A/AC.105/L.319 (Russia-UN).
- Permanent Mission of the United States of America to the United Nations (Vienna), “Information furnished in conformity with the Convention on Registration of Objects Launched into Outer Space: Registration information for November 2018 to May 2019,” (Oct 8, 2020) ST/SG/SER.E/924 (U.S. Registry).
- Press Release, CNSA, China and Russia sign a Memorandum of Understanding Regarding Cooperation for the Construction of the International Lunar Research Station (Mar 9, 2021).
- Press Release, ESA, Positive signs for Europe as ESA goes forward to the Moon (Dec 27, 2020).

- Press Release, Luxembourg Government, Luxembourg teams up with ESA to create a unique ‘European Space Resources Innovation Center’ to be established in the Grand Duchy, (Nov 18, 2020).
- Press Release, Ministry of the Economy–Luxembourg, German Aerospace Center and LSA to co-operate in research with space research utilization focus (Oct 23, 2019).
- Press Release, Ministry of the Economy–Luxembourg, Luxembourg and the United Arab Emirates to cooperate on space activities with particular focus on the exploration and utilization of space resources (Oct 10, 2017).
- Press Release, Ministry of the Economy–Luxembourg, The Grand Duchy of Luxembourg and Belgium join forces to develop the exploration and utilization of space resources (Jan 23, 2019).
- Press Release, NASA, NASA Administrator Signs Statement of Intent with Brazil on Artemis Cooperation (Dec 14, 2020).
- Press Release, NASA, NASA, International Partners Advance Cooperation with First Signings of Artemis Accords (Oct 13, 2020).
- Press Release, Roscosmos, Plans to seize other planets’ territories damage cooperation, (Apr 7, 2020).
- Press Release, U.S. Embassy in Ukraine, Ukraine becomes the 9th Country to sign the Artemis Accords (Nov 17, 2020).
- Radboud University Astrophysics, Netherlands–China Low-Frequency Explorer (NCLE) onboard Chang’e 4, 2018. <https://www.ru.nl/astrophysics/radboud-radio-lab/projects/netherlands-china-low-frequency-explorer-ncle/>.
- Reservation of coal and mineral rights, 42 U.S.C. § 299 (1993).
- Restatement (Fourth) of the Foreign Relations Law of the United States (Am. L. Inst. 2018).
- Soldatkin, Vladimir. 2019. *Russia wants to join Luxembourg in space mining*, Reuters, Mar 6, 2019.
- S.S. *Lotus*, 1927 P.C.I.J. (ser. A) No. 10.
- Tass. 2020. *Russia will not accept attempts to privatize the Moon, says Roscosmos CEO*, May 25, 2020.
- Terrorism exception to the jurisdictional immunity of a foreign state, 28 U.S.C. § 1605A (2008).
- Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, Jan 27, 1967, 18 U.S.T. 2410, 610 U.N.T.S. 205 (Outer Space Treaty).
- Tronchetti, Fabio. 2009. *The exploitation of natural resources of the moon and other celestial bodies: a proposal for a legal regime* (Martinus Nijhoff).
- United Arab Emirates Fed. L. No. (12) of 2019 On the Regulation of the Space Sector, Art. 18 (2019 UAE Space Law).
- United Nations Convention on the Law of the Sea, Dec 10, 1982, 1833 U.N.T.S. 397 (UNCLOS).
- United Nations Office for Outer Space Affairs, “Status of International Agreements Relating to Activities in Outer Space as at 1 January 2020”. <https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/status/index.html> (Treaties Status).
- United States Atomic Energy Act, codified at 42 U.S.C. § 2011 et seq. (1954).
- United States Commercial Space Launch Competitiveness Act of 2015, Pub. L. 114–90, tit. IV, 129 Stat. 721 (2015 U.S. SPACE Act).
- United States Constitution.
- United States Gen. Mining Act, codified at 30 U.S.C. § 23 (1872).
- United States National and Commercial Space Programs, 51 U.S.C. §§ 10101-71302.
- United States Nuclear Waste Policy Act, codified at 42 U.S.C. § 10101 et seq. (1982).
- United States Space Command, Mission. <https://www.spacecom.mil/Mission/>. Accessed Apr 16, 2021 (Space Command).
- Vancouver Recommendations on Space Mining (Outer Space Inst. Apr 20, 2020) (“Vancouver Recommendations”).
- Wimmer-Schweingruber et al. 2016. *LND for Chang’E 4 Mission*, Fall Gen. Assembly (Am. Geophysical Union Dec 2016).

Correction to: Instrumentation for Planetary Exploration



Emily F. Klonicki-Ference, Michael J. Malaska, Mark P. Panning,
Sarah E. Waller, and Patrick J. Gasda

Correction to:
Chapter 6 in: V. Badescu et al. (eds.), *Handbook of Space Resources*, https://doi.org/10.1007/978-3-030-97913-3_6

The original version of the book was inadvertently published without updating the small instrument parameters in Chapter 6. The correction chapter and the book have been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-3-030-97913-3_6

Index

A

Aamot, 977
Absolute coordinates, 730
Acetaldehyde, 622
Active feedback control, 128
Active sonar, 938
Agglutinates, 402
Airlock, 651
Akatsuki, 351
Alfred Wegener Institute (AWI), 979
Alien life, 1180, 1181, 1184
Aluminum, 625
Analog, 481, 482
Analysis horizon, 1094
Anchoring, 819
Anode, 621, 625, 626
Antarctica, 960
Apollo, 648, 1185
APRILtags, 741
Ariane V, 339
Artemis Accords, 1048, 1112, 1172, 1178, 1179, 1185, 1189–1196
Artemis Program, 1172, 1179, 1189, 1190, 1193, 1196
Artificial gravity, 779
Asteroid, 230, 232, 235, 243, 245, 246, 402, 787, 804, 1181
Asteroid colonies, 764
Asteroid exploration, 841, 842
Asteroid habitats, 768
Asteroid mining, 781, 827
Asteroid resources, 793
Asteroid resource utilization, 832
Astronauts, 1164, 1167, 1176, 1180, 1185, 1190
Athlete, 653
Atmosphere, 351

Attitude dynamics, 47
Attitude stability, 91
Australia, 1165, 1172, 1189, 1196
Autocatalytic modules, 1064
Autonomous Underwater Vehicle (AUV), 987
Autonomy, 967, 972
Available free energy, 1063

B

Backward contamination, 1113
Balloon, 648
Ballooning spider, 44
Bankability, 1117
Basalt, 464
Belgium, 1165, 1189, 1191, 1192
Beneficiation, 417, 418
Bennu, 809
BepiColombo, 339, 341, 345, 348
Biomining, 423
Blood Falls, 960
Blue Horizon, 1040
Bow-shock, 349
Brazil, 1172, 1189, 1196
Brine, 960
Building blocks, 1112, 1178, 1187
BUK, 970
Bundle adjustment, 737

C

Cabin, 656
Callisto, 960
Canada, 1172, 1187, 1190, 1196
Capacitance, 70
Carbonaceous chondrite, 788, 809, 812

Carbon dioxide, 621, 624, 625
 Carbon monoxide, 621, 625
 Cartesian move, 730
 Catalyst, 620, 626, 635
 Catenary equation, 747
 Cathode, 621
 C-complex, 807
 Celestial body, 1167, 1170, 1173, 1174, 1180
 Chang'e, 1190, 1191, 1193
 Chaos terrains, 912
 Charge, 47
 Chariot-type, 655
 China, 1179, 1188, 1190–1194, 1197
 China National Space Administration (CNSA), 1179, 1190, 1194
 Chinese Lunar Exploration Program (CLEP), 1179, 1190, 1191, 1193
 Chromium, *see* Phillips catalyst
 Circumnavigation, 47
 Cis-lunar space, 831
 Cleaning, 967
 Cleanliness, 975
 Clocking angle, 738
 Clockwork Starfish, 820
 Clohessy–Wiltshire, 53
 Closet Approach (CA), 349
 Cockpit, 654
 Cold Regions Research and Engineering Laboratory (CRREL), 977, 978
 Collisionless plasma, 64
 Colorado School of Mines, 1048
 Comets, 788
 Command and Data Handling (C&DH), 729
 Commercial Space Launch Competitiveness Act, 1112
 Committee for Mineral Reserves International Reporting Standards (CRIRSCO), 1038
 Committee for the Peaceful Uses of Outer Space (COPUOS), 1047, 1189, 1191, 1193
 Committee on Space Research (COSPAR), 1181–1184
 Common heritage of mankind, 1116
 Communication, 967
 Communication link, 971
 Concrete, 420
 Conductive heat losses, 964
 Conductive losses, 964
 Conductivity, 63
 Construction, 418

Consumer modules, 1071
 Contamination, 825, 959
 Control, 124
 Convection, 882
 Convective losses, 964, 973
 Conveyance, 400
 Coordinated Universal Time (UTC), 349
 Co-registered, 737
 Cost savings assessment, 1028
 Coupled orbit-attitude dynamics, 96
 Critical refreezing length, 964
 Cryobot, 935–938, 951, 952, 956, 979
 Cryogenic ice, 911
 Curve melting, 983
 Cyclone, 636
 Cymbal flexensional transducer, 951

D

Debye length, 64
 Deep Space Industries, 830
 Deflection, 732
 Degasser, 637
 Degradation, 968
 DEMOKRITOS, 970
 Department of Commerce, 1195
 Department of Defense, 1196
 Desiccant, 631
 Design optimization, 915
 Design Reference Architecture 5.0, 618, 620, 631
 Deutsches Zentrum für Luft- und Raumfahrt (DLR), 982
 Dielectric constant, 63
 Differential charging, 70
 Differential heating, 983
 Digital Elevation Model (DEM), 735
 Dipole model, 125
 Displaced non-Keplerian orbit, 4
 Double dipole, 136
 Downlinked, 737
 Due regard, 1111
 Dust, 47, 49, 401, 408, 647
 Dust layer penetration, 982
 Dynamic braking, 724

E

Earth-Following Orbits (EFOs), 169, 218
 Earth observation, 252
 Eddy currents, 315
 Effective specific heat, 981
 Efficiency models, 962

- Efficiency of utilization of in-situ resources, 1095
- E-Glider, 43
- Electrical thrusters, 269
- Electric field, 64
- Electrodes, 46
- Electrolyzer, 619, 621
- Electromagnetic casting, 326
- Electro-Magnetic Levitator (EML), 313
- Electromagnetic Melting, 318
- Electron, 63
- Electrostatic flight, 47
- Electrostatic force, 73
- Electrostatic hovering, 44
- Electrostatic inflation, 47
- Electrostatic maneuvering, 47
- Electrostatic map, 124
- Electrostatic orbiting, 44
- Electrostatic periodic orbits, 44
- Electrostatic torque, 74
- Electrostriction, 948
- Embodied energy, 1058
- Embodied energy contributions, 1074
- Embodied energy cost of human labor, 1077
- Embodied energy cost of labor, 1100
- Embodied energy expended in a space logistics effort, 1073
- Embodied energy metadata, 1101
- Embodied energy values, 1069
- Enceladus, 957
- Enceladus Explorer (EnEx), 984
- End-effector, 724
- Endothermic, 621
- Energy-based value system, 1104
- Energy budget, 631, 637
- Energy-Dispersive X-ray Fluorescence (EDXRF), 479
- Energy language, 1066
- Energy of space logistics embodied into this logistical mass, 1083
- EnEx-IceMole, 977
- E-sail, 11, 857
- Estimation, 124
- Ethanol, 621
- Ethylene, 617, 621, 628
- Europa, 874, 957, 1181, 1183
- Europa Clipper, 882
- Europa Explorer (EurEx), 987
- European Investment Bank, 1049
- European Space Agency (ESA), 656, 1046
- European Space Resources Innovation Centre (ESRIC), 1049, 1192
- European Union, 1046
- Evolvable Mars Campaign (EMC), 653
- Excavation, 400, 410
- Excavators, 408
- Exchange rate between mass and lifetime embodied energy, 1073
- Executive Order 13,914, 1125
- Exothermic, 621, 626, 635
- Exploration, 955–960, 966, 972, 975, 985, 987
- Extraterrestrial life, 1180, 1181, 1197
- Extraterritorial, 1173, 1178
- Extra-vehicular Activity (EVA), 647
- Extremophilic life, 959
- F**
- Fault protection, 974
- Federal Aviation Administration (FAA), 1195, 1196
- Federal Communications Commission, 1195
- Ferrochrome, 626
- Fiducial markers, 740
- Field of View (FOV), 727
- Fishing, 1164, 1172, 1173, 1186
- Fission reactors, 968
- Fluidized bed, 620, 635
- Flybys, 347
- Footprint, 733
- Forefield exploration, 984
- Formate, 621
- Forward contamination, 1113, 1145
- France, 1173, 1191
- Frozen orbit, 364
- Fuel cell, 652
- Fusion engines, 764
- G**
- Galilean satellites, 874
- Ganymede, 874, 960, 1182
- Gas dryer, 623
- Gaseous Lunar Oxygen from Regolith Electrolysis (GaLORE), 833
- Gateway, 1190, 1194, 1196
- General Mining Act, 1177
- Geodynamics, 936
- Geopolymers, 420
- Geostationary orbit, 1137
- Germany, 1173, 1179, 1190–1192
- Global commons, 1111
- Gold, 815
- Goodness map, 737

Gravity field models, 57
 Greece, 1189, 1191
 Growth dynamics of the space economy, 1103
 Guarded moves, 730

H

Habitability, 652, 883, 936
 Hague International Space Resources Governance Working Group (ISRGGW), 1047, 1112, 1178, 1179, 1187, 1193
 Harmonic drive, 724
 Hayabusa-2, 807
 Heat capacity, 962
 Heat exchanger, 626, 636
 Heat rejection, *see* radiator
 Heat transport, 962
 High-Density Polyethylene (HDPE), *see* polyethylene
 High Efficiency Particulate Air (HEPA), 637
 High-pressure ice, 960
 Hopper, 663
 Hybrid analysis, 1070
 Hybrid probes, 956
 Hydrate, 628
 Hydrated minerals, 817
 Hydrogen, 621, 624, 625
 Hydroxyl, 817

I

Ice caps, 958
 Ice density, 962
 Ice Diver, 977, 985
 Ice melting probe, 955–961, 965–968, 970–973, 976–980, 986, 988, 989
 IceMole, 960, 983
 Ice-ocean interface, 959
 Ice penetration, 966, 976, 980, 981, 988
 Ice screw, 976, 983
 Ice shell, 874, 879
 IceShuttle Teredo, 987
 Icy moons, 955, 957, 960, 961, 965–967, 971, 975, 980, 987, 989
 Imidazole, 629
 Impact Features, 888
 India, 1188, 1191, 1197
 Indirect embodied energy for Crew labor hours, 1076
 Indirect embodiment of energy, 1068
 Inertial Measurement Unit (IMU), 972

Inflatable, 661
 In-ice navigation, 984
 Input/output-based sectoral models, 1070
 In-situ analysis, 976, 987
 In-Situ Resource Utilization (ISRU), 430, 463, 464, 566, 618, 808
 In situ useful mass, 1103
 In-space production function, 1100
 In-space supply chain, 1101
 Institut für Weltraumforschung (IWF), 980
 Instrumentation, 975
 Instrument Context Camera (ICC), 727
 Instrument Deployment Camera (IDC), 727
 Instrument Site Selection Working Group (ISSWG), 738
 Interior environments, 873, 886
 International Institute of Space Law (IISL), 1163, 1174
 International law, 1163, 1164, 1166, 1170, 1174, 1185, 1193, 1194
 International Lunar Research Station, 1179, 1190, 1194
 International Seabed Authority, 1187
 International Space Station (ISS), 1179
 International Space University, 1048
 International waters, 1164, 1169, 1172, 1173, 1175, 1186, 1187
 Inverse kinematics, 732
 Io, 1181
 Ion, 63
 Ionosphere, 351
 Iron, 789
 Iron-nickel meteorites, 811
 Iron-nickels, 810
 Ispace Europe, 1041
 Italy, 1172, 1173, 1192, 1196, 1197
 Itokawa, 808
 IWF I, 977

J

Japan, 1172, 1173, 1192, 1196
 Joint Ore Reserves Committee (JORC), 1038
 Juno, 882
 Jupiter, 1181–1183
 Jupiter Icy Moons Explorer (JUICE), 882
 Jurisdiction, 1164–1170, 1173, 1177, 1178, 1180, 1186

L

Landforms, 887
 Laser system, 986

- Latent heat of melting, 961
- Lateral conduction of heat, 962
- Lateral convection, 981
- Lateral heating, 983
- Lateral refreezing, 964
- Launching state, 1166–1168, 1175
- Lead Zirconate Titanate (pzt), 948, 949
- Levitation, 50
- Liability Convention, 1150, 1165–1167, 1171, 1173, 1175, 1176, 1180
- Life, 873
- Life detection, 936
- Lifetime embodied energy, 1058
- Lifetime embodied energy cost per unit of total lifetime useful mass, 1082
- Lifetime Embodied Energy (LEE), 1069, 1077
- Lifetime energy efficiency, 1099
- Lifetime energy output, 1073
- Lifetime mass, 1084
- Lightness number, 362
- Liquifer Systems Group (LSG), 656
- Localization, 741
- Logistical mass, 1083
- Low-Density Polyethylene (LDPE), *see* polyethylene
- Lunar Orbital Platform Gateway, 271
- Lunar Roving Vehicle (LRV), 648
- Luxembourg, 1023, 1046, 1164, 1172, 1188, 1189, 1191–1193, 1195, 1196
- Luxembourg Institute of Science and Technology, 1048
- Luxembourg Law on the Exploration and Use of Space Resources, 1115
- Luxembourg National Research Fund, 1048
- Luxembourg Space Agency, 1034, 1046

- M**
- Maana Electric, 1042
- Macro-model, 62
- Magnetic detents, 724
- Magnetic field, 338
- Magnetosphere, 339
- Magnetostriction, 948
- Make vs take, 1081
- Mariner 10, 338
- Mars, 230, 232–236, 243–246, 402, 955, 957, 958, 960–962, 965, 966, 968, 970, 980, 981, 1181, 1183
- MarsCruiserOne (MCO), 653
- Mars Design Reference Mission (MDRM), 653
- Mars Surface Transportation System (MSTS), 653
- Mass budget, 631, 637
- Mass emplaced on Mars, 1081
- Matrix, 480
- Maximum power principle, 1064
- Maxwellian distribution, 65
- McMurdo Dry Valleys, 960
- Mechanical hardstops, 724
- Megaregolith, 401
- Melting velocity, 961
- Melt water flow, 962
- Meltwater sampling, 987
- Membrane, 623, 627, 628
- Mercury, 337, 338
- Mercury exploration, 360
- Mercury flybys, 346
- Mercury Magnetospheric Orbiter (MMO, nicknamed Mio), 339, 344
- Mercury Planetary Orbiter (MPO), The, 339, 341
- Mercury steering, 977
- Mercury Transfer Module (MTM), 339
- Mesothermal, 63
- MESSENGER, 338
- Metal-ceramic, 626, 631
- Metal-rich, 788
- Meteorites, 787, 807
- Methane, 617, 621, 628
- Micro-gravity, 251
- Miniaturization, 967, 973
- Mining, 1164, 1169, 1171, 1173, 1177, 1178, 1181, 1182, 1187, 1190, 1193, 1194, 1196, 1197
- Mining Space Summit, 1034
- Mission authorization, 1121
- Mobility, 42, 646
- MOBITAT, 653
- Modularity, 651
- Molecular sieve, *see* dessicant
- Moon, 230, 231, 234, 238, 240, 243–246, 402, 1165–1167, 1170–1172, 1174–1177, 1179–1181, 1183–1194, 1196, 1197
- Moon Agreement, 1165, 1176, 1186, 1187, 1189, 1191, 1193, 1194
- Moon Treaty, 1111
- MTR, Inc., 623, 627
- Multi-Mission RTG (MMRTG), 968

- N**
- NASA, 1166, 1179, 1181, 1185, 1188–1190, 1192, 1193, 1195, 1196

- Navigation, 124, 967, 972
 Near Earth Asteroids (NEAs), 763, 805
 Near earth objects, 1129
 Netherlands, 1165, 1179, 1190, 1193
 Networks of energy flows, 1099
 Noise Level (NL), 941, 943, 946, 952
 Non-Keplerian orbits, 3
 Nuclear electric propulsion, 970
 Nuclear power, 1177
 Nuclear reactor, 970
 Nucleation, 324
 Numerical methods, 891
- O**
- Oceans, 873
 Ocean Worlds, 873, 955
 OffWorld, 1043
 Olivine, 464, 479
 Omnidirectional beam, 951
 Optical fiber, 986
 Optical Surface Reflectors (OSR, i.e. mirrors), 343
 Opus 12, Inc., 621
 Orbital dynamics, 47
 Orbital power plants, 251
 Ordinary chondrites, 790, 808, 811
 Ore, 797
 OSIRIS-REx, 809
 Outer Space Institute, 1187
 Outer Space Treaty, 1110, 1165–1167, 1170–1178, 1180, 1185, 1186, 1189, 1191, 1196, 1197
- P**
- Pacific International Space Center for Exploration Systems (PISCES), 463
 Parent bodies, 804
 Parent rock, 480, 485
 Particle-in-cell, 69
 Passive sonar, 938
 Pendulum steering, 977, 985
 Penetrometer, 522
 Penitentes, 911
 Permeability, 629
 Petroleum Resources Management System (PRMS), 1038
 Phase-change, 962
 Phenocrysts, 480
 Philberth Probe, 977
 Phillips catalyst, 635
 Photoelectron, 64
 Photovoltaic, 652
 Physical value creation, 1063
 Piezoelectricity, 948
 Piezoelectric transducer, 948–950, 952
 Piracy, 1170, 1175
 Plagioclase feldspar, 464, 474
 Planetary defense, 1179, 1183, 1184
 Planetary flyby, 346
 Planetary gears, 724
 Planetary industrial ecosystems and supply chains, 1098
 Planetary protection, 967, 975, 1113, 1176, 1179, 1181–1185
 Planetary protection rules, 957
 Planetary Resources, Inc, 830
 Planetary Science Decadal Survey, 936
 Plastic, *see* polyethylene
 Plate tectonics, 889
 Platinum, 790, 816
 Poisson's equation, 65
 Policy on Planetary Protection, 1180
 Polyethylene, 617, 633
 Polymer electrolyte membrane, *see* proton exchange membrane (PEM)
 Polymerization, 629, 634, 636
 Porosity, 967, 981, 983
 Potential, 67
 Potential Hazardous Asteroids (PHAs), 770
 Power demand, 968
 Power supply, 144
 Power requirements, 321
 Power transmission, 967
 Pressure, 621, 625, 629
 Pressure resistance, 973
 Pressurized rover, 649
 Primary energy source, 1069
 Probe using Radioisotopes for Icy Moons Exploration (PRIME), 936–938, 951, 952
 Process tree models, 1069
 Producer modules, 1071
 Product distribution, *see* yield
 Propellant, 618, 621, 637
 Property, 1172, 1173, 1177, 1186
 Proto-flight, 725
 Proton, 64
 Proton Exchange Membrane (PEM), 621
 Pyroxene, 464, 479
- R**
- Radiation, 252, 647
 Radiation hardness, 974
 Radiator, 342, 624, 637
 Radioactive isotopes, 969

- Radioisotope Thermoelectric Generator (RTG), 937, 939, 945, 946, 950, 968, 979
- RAMA, 653
- Ray casting, 63
- Reachability, 735
- Real property, 1164, 1169, 1172, 1186
- Recoverable Autonomous Sonde (RECAS), 986
- Reduced latent heat of melting, 961
- Reference frame, 50
- Registration Agreement, 1150
- Registration Convention, 1165–1167, 1171, 1173, 1175, 1176, 1180
- Regolith, 70, 400, 430, 439, 521, 624, 813, 911
- Regolith Advanced Surface Systems Operations Robot (RASSOR), 412, 833
- Relationship between economic growth and energy, 1063
- Relative coordinates, 730
- Requirements, 723
- Rescue Agreement, 1150, 1165, 1171, 1185
- Resource exploration, 832
- Resources, 815
- Reverse water-gas shift, 620
- Robot Markup Language (RML), 743
- Robot Sequence Editor (RoSE), 741
- Robot Sequencing and Visualization Program (RSVP), 741
- Rolling, 662
- Roscosmos, 1194
- Rover, 648
- Rubble-piles, 809
- Russia, 1179, 1188, 1190, 1193, 1194
- Ryugu, 807, 809

- S**
- Sabatier, 619
- safe, 730
- Salt evaporites, 912
- Sample and probe recovery, 986
- Sampling system, 925
- Saturn, 1182
- Scoop, 726
- Seafloors, 885
- Security of tenure, 1115
- Sensitivity analysis, 638
- Separation, 416, 623, 627, 628
- SESAME, 969
- S-group, 807
- Shannon–Hartley theorem, 942
- Shielding, 974
- Shuttle Under Ice Shelf (SUSI), 977, 979
- Siderophile elements, 789
- Sidewall heaters, 983
- Sieving, 416
- Silver, 629
- Simulants, 400, 521
- Single dipole, 134
- Single crystal PZN-PT, 949
- Singularity, 733
- Sintering, 419, 431, 464, 465
- Small-body sampling, 47
- Smart dust, 359
- SNAP-10A, 970
- Socio-economic benefits, 1031
- Solar array, 343
- Solar Electric Propulsion, 339
- Solar radiation pressure, 43
- Solar Particle Events (SPE), 647
- Solar radiation flux, 968
- Solar sail, 5, 365, 366, 368, 851
- Solar wind, 49, 64, 338
- Solidification, 324
- Sonar, 936–939, 947, 950, 952
- Sonar equations, 942
- Source of energy, 1081
- South Polar Terrain, 959
- South Pole ice, 941, 943, 946
- Space colonies, 765
- Space debris, 310, 1113
- Space debris remediation, 311
- SpaceDev, 829
- Space economic output, 1102
- Space economy, 312
- Space elevator, 229–235, 239–246
- Space Exploration Vehicle (SEV), 653
- Space logistics emjoules, 1073
- Space resources, 310
- SpaceResources.lu, 1045
- Space station, 250, 254
- Specific embodied energy of energy, 1073
- Specific impulse, 622, 630
- Sphere, 662
- Sphere of influence, 346
- Spherical Environment Exploration Device (SEED), 653
- Stability, 47
- State Department, 1196
- Statistical methods, 895
- Steel, 625, 626
- Stereo baseline error, 735
- Stereo images, 735

Sterilization, 967, 975
 Stony-irons, 810
 Subglacial lakes, 956
 Sublimation, 961
 Subsurface ocean, 958
 Suitlocks, 651
 Suitports, 651
 Supercritical, 633
 Surface geology, 878

T

Taxonomic types, 804
 Taylor Glacier, 960, 984
 Telecommand, 971
 Teledyne Benthos modem, 940
 Telemetry, 971
 Temperature, 621, 625, 640
 Tether, 140, 967, 971
 Thermal characterization, 724
 Thermal drills, 956
 Thermoelastic noise, 748
 3D printing, 421
 3D simulations, 919
 Thrusters, 342
 Thrust vectoring, 205
 Tiepoints, 737
 Titan, 890, 960, 1182
 Titanium, *see* Ziegler-Natta catalyst
 Tool frames, 730
 TOPAZ, 970
 Total lifetime useful mass on Mars, 1081
 Touch-and-Go, 814
 Tracking, 199
 Trailer, 654
 Trajectory model, 962, 965
 Trans Astronautica Corporation, 831
 Transceiver pods, 971
 Transceiver relays, 980
 Transmission Loss (TL), 941, 942
 Triethylaluminum, *see* Ziegler-Natta catalyst
 Trilateration, 972
 2015 U.S. SPACE Act, 1164, 1172, 1173, 1195, 1196
 2017 Luxembourg Space Law, 1164, 1172, 1192
 2019 UAE Space Law, 1164, 1172, 1194

U

Ukraine, 1172, 1194, 1196, 1197
 UN Committee on the Peaceful Use of Outer Space, 1186

Undercooling, 324
 UN General Assembly, 1167, 1187
 United Arab Emirates (UAE), 1164, 1172, 1188, 1192, 1194–1197
 United Kingdom, 1172, 1173, 1195, 1196
 United Nations, 1046, 1047
 United Nations Convention on the Law of the Sea (UNCLOS), 1111, 1173, 1187, 1188
 United States, 1163, 1164, 1167–1169, 1172, 1173, 1175, 1177–1179, 1181, 1183–1186, 1188, 1192, 1193, 1195, 1196
 University of Luxembourg, 1048
 Upstream direct and indirect embodied energy inputs, 1075
 USC-IFEPIIC, 70
 U.S. Deep Seabed Act, 1172, 1173
 U.S. Fishermen's Protective Act, 1172, 1173
 U.S. Geological Survey, 1196
 U.S. Space Command, 1175
 U.S. Space Force, 1176
 Utilization, 1082
 UV radiation, 49

V

VALKYRIE, 977, 986
 Valuation of space infrastructure, 1102
 Value chain, 1026, 1029
 Velocity models, 965
 Venus, 337, 1182, 1184
 Venus flyby, 346, 349
 Vertical Takeoff/Vertical Landing (VTVL), 466
 Vesicles, 483
 Vesiculated, 482, 483
 Vicious circle of mass, 1055
 Volatile-rich, 788
 Volatiles, 808, 823
 Volatiles Investigating Polar Exploration Rover (VIPER), 833
 Volcanism, 885

W

Waste heat, 624
 Water-gas shift, 622, 625
 “Wet” architecture, 974
 Wolf Amendment, 1179, 1190
 Workspace, 724

XX-complex, [807](#)X-ray diffraction analysis, [480](#)X-Ray Diffraction (XRD), [480](#)**Y**YENISEI, [970](#)Yield, [622](#)**Z**Zeolite, [629](#)Ziegler-Natta catalyst, [635](#)