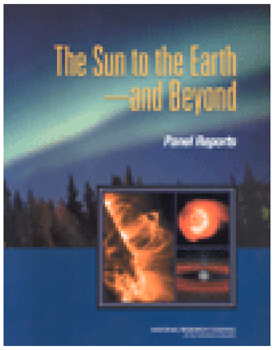
The Sun to the Earth -- and Beyond: Panel Reports



Solar and Space Physics Survey Committee, Committee on Solar and Space Physics, National Research Council

ISBN: 0-309-52938-7, 264 pages, 8 1/2 x 11, (2003)

This free PDF was downloaded from:
http://www.nap.edu/catalog/10860.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

Thank you for downloading this free PDF. If you have comments, questions or want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This book plus thousands more are available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The Sun to the Earth —and Beyond

Panel Reports

Solar and Space Physics Survey Committee
Committee on Solar and Space Physics
Space Studies Board
Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the panels responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by Contracts NASW 96013 and NASW 01001 between the National Academy of Sciences and the National Aeronautics and Space Administration; National Oceanic and Atmospheric Administration Purchase Order No. 40-AA-NR-111308; National Science Foundation Grant No. ATM-0109283; Office of Naval Research Grant No. N00014-01-1-0753; and Air Force Office of Scientific Research Purchase Order No. FQ8671-0101168. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

Cover: The background photo is of the aurora borealis as viewed from the vicinity of Fairbanks, Alaska. The three figures in the inset show the magnetically structured plasma of the Sun's million-degree corona (left); the plasmasphere, a cloud of low-energy plasma that surrounds Earth and co-rotates with it (top right); and an artist's conception of Jupiter's inner magnetosphere, with the Io plasma torus and the magnetic flux tubes that couple the planet's upper atmosphere with the magnetosphere. Ground-based aurora photo courtesy of Jan Curtis; coronal image courtesy of the Stanford-Lockheed Institute for Space Research and NASA; plasmasphere image courtesy of the IMAGE EUV team and NASA; rendering of the jovian magnetosphere courtesy of J.R. Spencer (Lowell Observatory).

International Standard Book Number 0-309-08972-7 (book)
International Standard Book Number 0-309-52593-4 (PDF)

Copies of this report are available free of charge from

Space Studies Board
National Research Council
The Keck Center of the National Academies
500 Fifth Street, N.W.
Washington, DC 20001

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

RECENT REPORTS OF THE SPACE STUDIES BOARD

Satellite Observations of the Earth's Environment: Accelerating the Transition of Research to Operations (2003)

Assessment of Directions in Microgravity and Physical Sciences Research at NASA (prepublication) (2002)
Assessment of the Usefulness and Availability of NASA's Earth and Space Mission Data (2002)
Factors Affecting the Utilization of the International Space Station for Research in the Biological and Physical Sciences (prepublication) (2002)
Life in the Universe: An Assessment of U.S. and International Programs in Astrobiology (prepublication) (2002)
New Frontiers in the Solar System: An Integrated Exploration Strategy (prepublication) (2002)
Review of NASA's Earth Science Enterprise Applications Program Plan (2002)
"Review of the Redesigned Space Interferometry Mission (SIM)" (2002)
Safe on Mars: Precursor Measurements Necessary to Support Human Operations on the Martian Surface (2002)
The Sun to the Earth—and Beyond: A Decadal Research Strategy in Solar and Space Physics (prepublication) (2002)
Toward New Partnerships in Remote Sensing: Government, the Private Sector, and Earth Science Research (2002)
Using Remote Sensing in State and Local Government: Information for Management and Decision Making (2002)

Assessment of Mars Science and Mission Priorities (prepublication) (2001)
The Mission of Microgravity and Physical Sciences Research at NASA (2001)
The Quarantine and Certification of Martian Samples (prepublication) (2001)
Readiness Issues Related to Research in the Biological and Physical Sciences on the International Space Station (2001)
"Scientific Assessment of the Descoped Mission Concept for the Next Generation Space Telescope (NGST)" (2001)
Signs of Life: A Report Based on the April 2000 Workshop on Life Detection Techniques (prepublication) (2001)
Transforming Remote Sensing Data into Information and Applications (2001)
U.S. Astronomy and Astrophysics: Managing an Integrated Program (2001)

Copies of these reports are available free of charge from
Space Studies Board
National Research Council
The Keck Center of the National Academies
500 Fifth Street, N.W., Washington, DC 20001
(202) 334-3477
ssb@nas.edu
www.national-academies.org/ssb/ssb.html

NOTE: Listed according to year of approval for release.

SOLAR AND SPACE PHYSICS SURVEY COMMITTEE

LOUIS J. LANZEROTTI, Lucent Technologies, *Chair*
ROGER L. ARNOLDY, University of New Hampshire
FRAN BAGENAL, University of Colorado at Boulder
DANIEL N. BAKER, University of Colorado at Boulder
JAMES L. BURCH, Southwest Research Institute
JOHN C. FOSTER, Massachusetts Institute of Technology
PHILIP R. GOODE, Big Bear Solar Observatory
RODERICK A. HEELIS, University of Texas, Dallas
MARGARET G. KIVELSON, University of California, Los Angeles
WILLIAM H. MATTHAEUS, University of Delaware
FRANK B. McDONALD, University of Maryland
EUGENE N. PARKER, University of Chicago, Professor Emeritus
GEORGE C. REID, University of Colorado at Boulder
ROBERT W. SCHUNK, Utah State University
ALAN M. TITLE, Lockheed Martin Advanced Technology Center

Staff

ARTHUR A. CHARO, Study Director
WILLIAM S. LEWIS,[1] Consultant
THERESA M. FISHER, Senior Program Assistant

---

[1] On temporary assignment from Southwest Research Institute.

PANEL ON THE SUN AND HELIOSPHERIC PHYSICS

JOHN T. GOSLING, Los Alamos National Laboratory, *Chair*
ALAN M. TITLE, Lockheed Martin Advanced Technology Center, *Vice Chair*
TIMOTHY S. BASTIAN, National Radio Astronomy Observatory
EDWARD W. CLIVER, Air Force Research Laboratory
JUDITH T. KARPEN, Naval Research Laboratory
JEFFREY R. KUHN, University of Hawaii
MARTIN A. LEE, University of New Hampshire
RICHARD A. MEWALDT, California Institute of Technology
VICTOR PIZZO, NOAA Space Environment Center
JURI TOOMRE, University of Colorado at Boulder
THOMAS H. ZURBUCHEN, University of Michigan

PANEL ON SOLAR WIND AND MAGNETOSPHERE INTERACTIONS

CHRISTOPHER T. RUSSELL, University of California, Los Angeles, *Chair*
JOACHIM BIRN, Los Alamos National Laboratory, *Vice Chair*
BRIAN J. ANDERSON, Johns Hopkins University
JAMES L. BURCH, Southwest Research Institute
JOSEPH F. FENNELL, Aerospace Corporation
STEPHEN A. FUSELIER, Lockheed Martin Advanced Technology Center
MICHAEL HESSE, NASA Goddard Space Flight Center
WILLIAM S. KURTH, University of Iowa
JANET G. LUHMANN, University of California, Berkeley
MARK MOLDWIN, University of California, Los Angeles
HARLAN E. SPENCE, Boston University
MICHELLE F. THOMSEN, Los Alamos National Laboratory

PANEL ON ATMOSPHERE-IONOSPHERE-MAGNETOSPHERE INTERACTIONS

MICHAEL C. KELLEY, Cornell University, *Chair*
MARY K. HUDSON, Dartmouth College, *Vice Chair*
DANIEL N. BAKER, University of Colorado at Boulder
THOMAS E. CRAVENS, University of Kansas
TIMOTHY J. FULLER-ROWELL, University of Colorado at Boulder
MAURA E. HAGAN, National Center for Atmospheric Research
UMRAN S. INAN, Stanford University
TIMOTHY L. KILLEEN, National Center for Atmospheric Research
CRAIG KLETZING, University of Iowa
JANET U. KOZYRA, University of Michigan
ROBERT LYSAK, University of Minnesota
GEORGE C. REID, University of Colorado at Boulder
HOWARD J. SINGER, NOAA Space Environment Center
ROGER W. SMITH, University of Alaska

PANEL ON THEORY, MODELING, AND DATA EXLORATION

GARY P. ZANK, University of California, Riverside, *Chair*
DAVID G. SIBECK, NASA Goddard Space Flight Center,[1] *Vice Chair*
SPIRO K. ANTIOCHOS, Naval Research Laboratory
RICHARD S. BOGART, Stanford University
JAMES F. DRAKE, JR., University of Maryland
ROBERT E. ERGUN, University of Colorado at Boulder
JACK R. JOKIPII, University of Arizona
JON A. LINKER, Science Applications International Corporation
WILLIAM LOTKO, Dartmouth College
JOACHIM RAEDER, University of California, Los Angeles
ROBERT W. SCHUNK, Utah State University

PANEL ON EDUCATION AND SOCIETY

RAMON E. LOPEZ, University of Texas, El Paso, *Chair*
MARK ENGEBRETSON, Augsburg College, *Vice Chair*
FRAN BAGENAL, University of Colorado at Boulder
CRAIG DeFOREST, Southwest Research Institute
PRISCILLA FRISCH, University of Chicago
DALE E. GARY, New Jersey Institute of Technology
MAUREEN HARRIGAN, Agilent Technologies
ROBERTA M. JOHNSON, National Center for Atmospheric Research
STEPHEN P. MARAN, NASA Goddard Space Flight Center
TERRANCE ONSAGER, NOAA Space Environment Center

---

[1] Johns Hopkins University Applied Physics Laboratory until summer 2002.

SPACE STUDIES BOARD

JOHN H. McELROY, University of Texas at Arlington (retired), *Chair*
J. ROGER P. ANGEL, University of Arizona
JAMES P. BAGIAN, Veterans Health Administration's National Center for Patient Safety
ANA P. BARROS, Harvard University
RETA F. BEEBE, New Mexico State University
ROGER D. BLANDFORD, California Institute of Technology
JAMES L. BURCH, Southwest Research Institute
RADFORD BYERLY, JR., University of Colorado at Boulder
HOWARD M. EINSPAHR, Bristol-Myers Squibb Pharmaceutical Research Institute (retired)
STEVEN H. FLAJSER, Loral Space and Communications Ltd.
MICHAEL H. FREILICH, Oregon State University
DON P. GIDDENS, Georgia Institute of Technology/Emory University
RALPH H. JACOBSON, The Charles Stark Draper Laboratory (retired)
MARGARET G. KIVELSON, University of California, Los Angeles
BRUCE D. MARCUS, TRW, Inc. (retired)
HARRY Y. McSWEEN, JR., University of Tennessee
GEORGE A. PAULIKAS, The Aerospace Corporation (retired)
ANNA-LOUISE REYSENBACH, Portland State University
ROALD S. SAGDEEV, University of Maryland
CAROLUS J. SCHRIJVER, Lockheed Martin
ROBERT J. SERAFIN, National Center for Atmospheric Research
MITCHELL SOGIN, Marine Biological Laboratory
C. MEGAN URRY, Yale University
PETER W. VOORHEES, Northwestern University
J. CRAIG WHEELER, University of Texas at Austin

JOSEPH K. ALEXANDER, Director

COMMITTEE ON SOLAR AND SPACE PHYSICS

JAMES L. BURCH, Southwest Research Institute, *Chair*
JAMES F. DRAKE, University of Maryland
STEPHEN A. FUSELIER, Lockheed Martin Advanced Technology Center
MARY K. HUDSON, Dartmouth College
MARGARET G. KIVELSON, University of California, Los Angeles
CRAIG KLETZING, University of Iowa
FRANK B. McDONALD, University of Maryland
EUGENE N. PARKER, University of Chicago, Professor Emeritus
ROBERT W. SCHUNK, Utah State University
GARY P. ZANK, University of California, Riverside

ARTHUR A. CHARO, Study Director
THERESA M. FISHER, Senior Program Assistant

Preface

This volume, *The Sun to the Earth—and Beyond: Panel Reports*, is a compilation of the reports from five National Research Council (NRC) panels convened as part of a survey in solar and space physics for the period 2003–2013. The NRC's Space Studies Board and its Committee on Solar and Space Physics organized the study. Overall direction for the survey was provided by the Solar and Space Physics Survey Committee, whose report, *The Sun to the Earth—and Beyond: A Decadal Research Strategy in Solar and Space Physics*, was delivered to the study sponsors in prepublication format in August 2002. The final version of that report was published in June 2003.[1] The survey report and the panel reports are included on the compact disk that is supplied with this volume.

The panel reports provide both a detailed rationale for the survey committee's recommendations and an expansive view of the numerous opportunities that exist for a robust program of exploration in solar and space physics. Although the recommendations of the survey committee are consistent with the priorities expressed by the panels, it was not possible to incorporate all of the panel recommendations into a balanced program that could be carried out over the next decade within a realistic resource envelope.

The preface to the survey committee report, *The Sun to the Earth—and Beyond: A Decadal Research Strategy in Solar and Space Physics*, is reproduced below.

> *The Sun to the Earth—and Beyond: A Decadal Research Strategy in Solar and Space Physics* is the product of an 18-month effort that began in December 2000, when the National Research Council (NRC) approved a study to assess the current status and future directions of U.S. ground- and space-based programs in solar and space physics research. The NRC's Space Studies Board and its Committee on Solar and Space Physics organized the study, which was carried out by five ad hoc study panels and the 15-member Solar and Space Physics Survey Committee, chaired by Louis J. Lanzerotti, Lucent Technologies. The work of the panels and the committee was supported by the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), the National Oceanic and Atmospheric Administration (NOAA), the Office of Naval Research (ONR), and the Air Force Office of Scientific Research (AFOSR).

---

[1] Space Studies Board, National Research Council. 2003. *The Sun to the Earth—and Beyond: A Decadal Research Strategy in Solar and Space Physics,* National Academies Press, Washington, D.C.

The Sun to the Earth—and Beyond is the report of the Solar and Space Physics Survey Committee. It draws on the findings and recommendations of the five study panels, as well as on the committee's own deliberations and on previous relevant NRC reports. The report identifies broad scientific challenges that define the focus and thrust of solar and space physics research for the decade 2003 through 2013, and it presents a prioritized set of missions, facilities, and programs designed to address those challenges.

In preparing this report, the committee has considered the technologies needed to support the research program that it recommends as well as the policy and programmatic issues that influence the conduct of solar and space physics research. The committee has also paid particular attention to the applied aspects of solar and space physics—to the important role that these fields play in a society whose increasing dependence on space-based technologies renders it ever more vulnerable to "space weather." The report discusses each of these important topics—technology needs, applications, and policy—in some detail. *The Sun to the Earth—and Beyond* also discusses the role of solar and space physics research in education and examines the productive cross-fertilization that has occurred between solar and space physics and related fields, in particular astrophysics and laboratory plasma physics.

Each of the five study panels was charged with surveying its assigned subject area and with preparing a report on its findings. The first three panels focused on the important scientific goals within their respective disciplines and on the missions, facilities, programs, technologies, and policies needed to achieve them. In contrast, the Panel on Theory, Modeling, and Data Exploration addressed basic issues that transcend disciplinary boundaries and that are relevant to all of the subdisciplines of solar and space physics. The Panel on Education and Society examined a variety of issues related to both formal and informal education, including the incorporation of solar and space physics content in science instruction at all levels, the training of solar and space physicists at colleges and universities, and public outreach. The reports of the panels will be published in a separate volume titled *The Sun to the Earth—and Beyond: Panel Reports*.

In addition to the input from the five study panels, the committee also received information at a 2-day workshop convened in August 2001 to examine in detail issues relating to the transition from research models to operational models. Participants in the workshop included members of the committee and representatives from the Air Force, the Navy, NOAA, NSF, NASA, the U.S. Space Command, academia, and the private sector.

The committee undertook its work intending to provide a community assessment of the present state and future directions of solar and space physics research. To this end, the committee and the panels engaged in a number of efforts to ensure the broad involvement of all segments of the solar and space physics communities. These efforts included town-meeting-like outreach events held at the May 2001 joint meeting of the American Geophysical Union (AGU) and the American Astronomical Society's (AAS's) Solar Physics Division[2] and at spring and summer 2001 workshops of the following programs: International Solar-Terrestrial Physics (ISTP), Solar, Heliospheric, and Interplanetary Environment (SHINE), Coupling, Energetics, and Dynamics of Atmospheric Regions (CEDAR), and Geospace Environment Modeling (GEM). Each of these outreach events was well attended and provided the committee and panels with valuable guidance, suggestions, and insights into the concerns of the solar and space physics community. Additional community input came from presentations on science themes, missions, and programs at panel meetings, from direct communication with individual panel and committee members by phone and e-mail, and through Web sites and Web-based bulletin boards established by two of the panels. Reports in the electronic newsletters of the AGU's Space Physics and Aeronomy section and of the AAS's Solar Physics Division kept

---

[2] The AGU and the Solar Physics Division of the AAS are the two principal scientific organizations representing the solar and space physics community.

those communities informed of the progress of the study and encouraged their continued involvement in the study process.

Each of the study panels met at least twice during the spring and summer of 2001. The Panel on the Sun and Heliospheric Physics and the Panel on Education and Society met three times. The committee met five times, three times in 2001 and twice in 2002. The panel chairs and vice chairs participated in two of those meetings, during which they presented their panels' recommendations and received comments and suggestions from the committee. The final set of scientific and mission, facility, and program priorities and other recommendations was established by consensus at the committee's last meeting, in May 2002.

The committee's final set of priorities and recommendations does not include all of the recommendations made by the study panels, although it is consistent with them. Each panel worked diligently to identify the compelling scientific questions in its subject area and to set program priorities to address these questions. All of the recommendations offered by the panels merit support; however, the committee took as its charge the provision of a strategy for a strong, balanced national program in solar and space physics for the next decade that could be carried out within what is currently thought to be a realistic resource envelope. Difficult choices were inevitable, but the recommendations presented in this report reflect the committee's best judgment, informed by the work of the panels and discussions with the scientific community, about which programs are most important for developing and sustaining the solar and space physics enterprise.

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's (NRC's) Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

The Sun and Heliospheric Physics
Loren W. Acton, Montana State University,
George Gloeckler, University of Maryland,
Leon Golub, Harvard-Smithsonian Center for Astrophysics, and
Marcia Neugebauer, University of Arizona.
Solar Wind and Magnetosphere Interactions
Stanley W.H. Cowley, Leicester University, United Kingdom,
Barry H. Mauk, Johns Hopkins University,
Ted J. Rosenberg, University of Maryland, and
Jack D. Scudder, University of Iowa.
Atmosphere-Ionosphere-Magnetosphere Interactions
Larry Lyons, University of California, Los Angeles,
Stephen Mende, University of California, Berkeley,
Michael Mendillo, Boston University,
Raymond G. Roble, National Center for Atmospheric Research, and
J. Hunter Waite, Jr., University of Michigan.
Theory, Modeling, and Data Exploration
Joseph B. Gurman, NASA Goddard Space Flight Center,
Lynn M. Kistler, University of New Hampshire,
Dana Longcope, Montana State University,
John D. Reppy, Cornell University, and
Robert J. Strangeway, University of California, Los Angeles.

Education and Society

Susana E. Deustua, American Astronomical Society,
Terry G. Forbes, University of New Hampshire,
Nicola Fox, Johns Hopkins University,
George Nelson, Western Washington University, and
Bruce Partridge, Haverford College.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the survey committee report and the panel reports was overseen by Robert A. Frosch, Harvard University, and Lennard A. Fisk, University of Michigan. Appointed by the National Research Council, they were responsible for making certain that an independent examination of the survey and panel reports was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of each report rests entirely with the authoring committee and the institution.

Contents

APPENDIXES

1

Report of the Panel on the Sun and Heliospheric Physics

SUMMARY

A revolution in solar and heliospheric physics is in progress. A variety of measurements, together with theory and numerical simulations, have created fresh insights into phenomena that occur in the Sun and the heliosphere and have sharpened our basic understanding of the underlying physical processes. Powerful modern computing capabilities now allow us to examine these physical processes and predict their observable signatures in considerable detail. To continue this revolution, the Panel on the Sun and Heliospheric Physics has formed an aggressive plan for solar-heliospheric research in the coming decade. Its plan is built on a systems approach to this broad yet strongly coupled domain.

RESEARCH THEMES

The prime guiding principle behind the major research issues and challenges for the next decade is to understand the processes that link the Sun-heliosphere-Earth system. The panel's recommended new programs are centered on five basic themes that stretch from the solar interior to the outer heliosphere and beyond:

- Exploring the solar interior,
- Understanding the quiet Sun,
- Exploring the inner heliosphere,
- Understanding the active Sun and the heliosphere, and
- Exploring the outer heliosphere and the local interstellar medium.

SCIENCE QUESTIONS FOR NEW RESEARCH INITIATIVES IN SOLAR-HELIOSPHERIC PHYSICS FOR THE COMING DECADE (PRIORITIZED)

Within the foregoing themes the panel has identified and prioritized those science questions requiring new initiatives to continue present progress in solar-heliospheric physics across a broad front:

1. What physical processes are responsible for coronal heating and solar wind acceleration, and what controls the development and evolution of the solar wind in the innermost heliosphere?
2. What determines the magnetic structure of the Sun and its evolution in time, and what physical processes determine how and where magnetic flux emerges from beneath the photosphere?
3. What is the physics of explosive energy release in the solar atmosphere, and how do the resulting heliospheric disturbances evolve in space and time?
4. What is the physical nature of the outer heliosphere, and how does the heliosphere interact with the galaxy?

OPERATIONAL PROGRAMS AND MISSIONS

If the current pace of progress in solar and heliospheric physics is to continue over the next decade, it is essential that key capabilities of the current space program in solar-heliospheric physics continue at least until they are replaced by missions in development, by approved missions awaiting development, or by the new initiatives recommended in this report. These capabilities are needed for a variety of high-priority science objectives and for the routine monitoring of the Sun and heliosphere that is critical for accurate specification and prediction of short-term space weather and longer-term space climate. In particular, it is essential that NASA maintain capabilities to image the corona at x-ray and extreme ultraviolet (EUV) wavelengths, to image coronal mass ejections in white light, to do helioseismology, and to measure the solar wind plasma, magnetic field, and energetic particle variations in near-Earth interplanetary space. The panel supports the continued active tracking of the Wind mission for targeted research topics and as a backup to the Advanced Composition Explorer (ACE) as a 1-AU monitor of heliospheric conditions. **The panel also specifically recommends continuation of two missions that are uniquely sampling difficult-to-reach heliospheric regions: Ulysses, as long as it is technically possible to do so; and Voyagers 1 and 2, as long as they are capable of providing measurements necessary to characterize the location and nature of the termination shock and heliopause.**

The panel also recognizes that extended, continuous, well-calibrated observations from space are critically important for detecting and measuring the now indisputable variability of the Sun's irradiance **and strongly recommends that continuous irradiance measurements from both the ground and space be continued indefinitely.**

Ground-based solar observatories carry out a variety of research programs and also provide valuable long-term synoptic observations. Each of these observatories contributes to one or more of the panel's research priorities, as do those of the ground-based neutron monitor network. The panel has not attempted to prioritize the ongoing programs of these institutions.

PROGRAMS IN DEVELOPMENT

The panel's recommendations for new initiatives presume that the missions and programs presently under active development and listed below will become operational within the coming decade to address the high-priority science objectives in solar-heliospheric physics for which they are designed.

- *Solar Terrestrial Relations Observatory (STEREO).* A two-spacecraft mission with identical in situ and remote sensing instrumentation on both spacecraft. STEREO is designed to study the origin and heliospheric propagation of disturbances driven by coronal mass ejections and their products in the ecliptic plane out to 1 AU.
- *Solar-B.* A joint Japanese-U.S.-U.K. mission that provides coordinated optical, EUV, and x-ray measurements to determine the relationship between changes in the photospheric magnetic field and changes in the structure of the chromosphere and corona.
- *Synoptic Optical Long-term Investigation of the Sun (SOLIS).* A suite of three National Solar Observatory (NSO) instruments at Kitt Peak, Arizona, designed to make sustained and well-calibrated observations relating to long-term solar variability.
- *Global Oscillations Network Group (GONG++).* Includes identical Michelson Doppler imaging instruments at six sites around the world to allow nearly uninterrupted full-disk observations of solar oscillations and magnetic fields. It is anticipated that the GONG experiment, which is in the process of being upgraded, will be operated for at least a solar cycle in order to study how solar interior dynamics evolves over the solar cycle at a wide range of depths.

APPROVED PROGRAMS

The following approved programs, which are not yet under full development, are prerequisites to the panel's recommended new programs.

- *Solar Dynamics Observatory (SDO).* A NASA Living With a Star (LWS) mission to study the Sun from the subsurface layers of the convection zone to the outer corona. It will carry an array of telescopes to image the inner solar atmosphere over a wide temperature range, an advanced Doppler package to image subsurface structures and detect sunspots developing on the far side of the Sun, an EUV irradiance monitor to study both short- and long-term variations in the solar irradiance that arise in response to changes in the solar magnetic field, and one or more coronagraphs to image the solar corona out to ~15 R_S. Instrumentation proposals for this mission have been submitted and are awaiting selection. [See note, p. 45.]
- *Advanced Technology Solar Telescope (ATST).* A ground-based National Science Foundation (NSF) program to provide precise, sensitive, high-resolution (0.1 arcsec) measurements of the solar magnetic and velocity fields with a broad set of diagnostics over a wavelength range from 0.3 to 35 micrometers. The telescope will have a very large aperture (4 m) and employ adaptive optics to attain these measurement goals and will be used to study solar magnetic fields from the density scale length of the photosphere up through the 5,000,000 K coronal plasma. This program is in the definition phase and is a major technical challenge.

RECOMMENDATIONS FOR MAJOR NEW INITIATIVES (PRIORITIZED)

1. ***A solar probe mission to the near-Sun region (4–60 R_S) to determine the origin and evolution of the solar wind in the innermost heliosphere via in situ sampling.*** The region inward of 0.3 AU is one of the last unexplored frontiers in our solar system, the birthplace of the heliosphere itself. Remote sensing observations and in situ sampling of the solar wind far from the Sun have provided tantalizing glimpses of the physical nature of this region. However, to understand how the solar wind originates and evolves in the inner heliosphere requires direct in situ sampling of the plasma, energetic particles, magnetic field, and waves, as close to the solar surface as possible—the panel's top science priority for the coming decade. Such measurements will determine how energy flows from the interior of the Sun through the surface and into the solar atmosphere, heating the corona and accelerating the wind, and will also reveal how the wind evolves with distance in the inner heliosphere. These measurements will revolutionize our basic understanding of the expanding solar atmosphere. The panel therefore strongly recommends a solar probe to the near-Sun region that emphasizes in situ measurements of the innermost heliosphere. The generic solar probe recommended by the panel is not necessarily identical to the Solar Probe mission for which NASA released an Announcement of Opportunity in September 1999 that placed equal emphasis on both in situ and remote sensing observations. For a first solar probe, the panel strongly believes that the in situ measurements are of the highest priority and should not be compromised. In general, the panel did not find various rationales given for including remote sensing instrumentation on a first mission to the near-Sun region to be compelling. Nevertheless, the panel appreciates that a solar probe mission provides perhaps the first opportunity to measure the photospheric magnetic field in the polar regions of the Sun via remote sensing. Such measurements will address the panel's second science priority and should be the secondary objective of a solar probe mission.

2. *The Frequency Agile Solar Radiotelescope (FASR) to image the Sun with high spatial and spectral resolution over a broad frequency range (0.1 to 30 GHz).* Radio imaging and radio spectroscopy provide unique insights into the solar chromosphere and corona. Combining radio imaging with radio spectroscopy provides a revolutionary new tool to study energy release in flares and coronal mass ejections (CMEs) and the thermal structure of the solar atmosphere in three dimensions. Moreover, radio imaging spectroscopy provides a range of powerful techniques for measuring magnetic fields in the corona. For example, measurements of gyroresonance emission can be used to determine the magnetic field strength in active regions at the base of the corona, observations of gyrosynchrotron radiation from mildly relativistic electrons can be used to probe the coronal magnetic field in solar flares, and multiband Stokes-V observations of solar free-free emission can be utilized to provide a measure of the longitudinal field to strengths as low as a few gauss. In addition, observations of radio depolarization and Faraday rotation can be used to measure even smaller magnetic fields in particular source regions (e.g., CMEs) or along particular lines of sight within the outer corona. FASR will probe both the quiet and the active solar atmosphere and is uniquely suited to measure coronal magnetic fields, nonthermal emissions from flares and CMEs, and the three-dimensional thermal structure of the solar atmosphere. Thus it directly addresses aspects of the panel's top three science priorities.

3. *Virtual Sun, a focused, interagency theory/modeling/simulation program to provide physical understanding across the Sun-heliosphere-Earth system.* Understanding the physical connections between the Sun, the heliosphere, and Earth is the prime guiding principle behind the major research issues and challenges for the next decade. The system is strongly coupled and highly nonlinear, linking spatial scales from current sheets to the size of the heliosphere and varying on time scales from fractions of a second to millennia. Its complexity has long been an obstacle to a full understanding of key mechanisms and processes, let alone to the construction of global models of the entire system. However, during the last decade we have broadened considerably our theoretical understanding of the Sun-heliosphere-Earth system, have collected a rich observational base to study it, and have witnessed a rapid development of supercomputing architectures. Together, these developments suggest that the time is ripe to complement the U.S. observational program in solar and space physics with a bold theory and modeling initiative that cuts across disciplinary boundaries. The Virtual Sun program will incorporate a systems-oriented approach to theory, modeling, and simulation and ultimately will provide continuous models from the solar interior to the outer heliosphere. The panel envisions that the Virtual Sun will be developed in a modular fashion via focused attacks on various physical components of the Sun-heliosphere-Earth system and on crosscutting physical processes. Two problems that appear ready for such a concentrated attack are the problem of the solar dynamo and that of three-dimensional magnetic reconnection in the solar atmosphere and heliosphere. Approximately 10 years will be required to achieve the goals of this mission. The panel envisions that the program will require both continuity and community oversight to meet such ambitious goals. In particular, individual components should be competitively selected and reviewed periodically to assess quantitative progress toward completion of a working Virtual Sun model.

4. *U.S. participation in the European Space Agency's (ESA's) Solar Orbiter mission for a combined in situ and remote sensing study of the Sun and heliosphere 45 R_S from the Sun.* Solar Orbiter is a natural successor to SOHO to explore the Sun and its interaction with the heliosphere. Selected by ESA for launch in the 2008–2015 time frame, Solar Orbiter will use a unique orbital design to bring a comprehensive payload of imaging and in situ particle and field experiments into an elliptical orbit with a perihelion of 45 R_S. At this distance Solar Orbiter will approximately co-rotate with the Sun. An overall goal of the mission is to reveal the magnetic structure and evolution of the solar atmosphere and the effects of this evolution on the plasma, energetic particles, and fields in the inner heliosphere. The orbital plane will increasingly become tilted with respect to the ecliptic plane, so that near the end of the mission the spacecraft will attain a solar latitude of 38 degrees. Thus, over the course of the mission Solar Orbiter will provide data on the magnetic field and convective flows at high latitude that are essential for understanding the solar dynamo. The panel finds that U.S. participation in this mission would be a cost-effective way for U.S. scientists to address various aspects of the top three science priorities. The mission will be particularly attractive if it includes instrumentation to investigate particle acceleration close to the Sun.

5. ***A multispacecraft heliospheric mission to probe in situ the three-dimensional structure of propagating heliospheric disturbances.*** Solar wind disturbances driven by CMEs are inherently complex three-dimensional structures. Our understanding of the evolution and global extent of these disturbances has largely been built on single-point in situ measurements obtained at and beyond 1 AU, although some multispacecraft observations of heliospheric disturbances have been obtained and STEREO will provide stereoscopic imaging and two-point in situ measurements of CME-driven disturbances. The panel believes that a multispacecraft heliospheric mission consisting of four or more spacecraft less than 1 AU from the Sun, separated in both radius (inside 1 AU) and longitude and emphasizing in situ measurements, promises a significant leap forward in our understanding of global aspects of the evolution of solar wind disturbances. A mission of this kind will illuminate the connections between solar activity, heliospheric disturbances, and geomagnetic activity and will directly addresses the third science priority; it is an essential element of NASA's Living With a Star program.

6. ***A reconnection and microscale (RAM) probe to examine the solar corona remotely with unprecedented spatial (~10 km) and temporal (millisecond to second) resolution.*** Observations and theory have long indicated that magnetic reconnection plays a key role in rapid energy release on the Sun. Although magnetic reconnection and its repercussions have long been studied intensively in Earth's magnetosphere via both observations and theory, many questions remain about its operation in the solar corona, where physical conditions are considerably different from conditions in the magnetosphere. Moreover, in situ sampling deep in the Sun's atmosphere is clearly out of reach. High-resolution spatial and temporal observations of the solar atmosphere are required to make further progress for understanding how magnetic reconnection operates in the solar atmosphere, in particular for understanding its role in the magnetic restructuring and rapid energy release characteristic of solar disturbances. This mission thus addresses aspects of the second and third science priorities for the coming decade.The panel finds that a RAM mission will also provide data extremely useful for understanding how wave transport and dissipation occur in the solar atmosphere, an aspect of the first science priority.

7. ***An interstellar sampler mission for the remote exploration of the interaction between the heliosphere and the local interstellar medium.*** The boundary between the solar wind and the local interstellar medium (LISM) is one of the last unexplored regions of the heliosphere. Very little is currently known about the shape and extent of this region or the nature of the LISM. The physical nature of these regions will be studied by an interstellar sampler mission using a combination of remote sensing and in situ sampling techniques at heliocentric distances between about 1 and 4 AU. The panel finds that such a pioneering mission would reveal new properties of the interstellar gas and the transport of pickup ions in the heliosphere and would thus directly address the fourth science priority. This mission is a natural precursor to a more ambitious probe to penetrate the interstellar medium directly.

PROGRAMS REQUIRING TECHNOLOGY DEVELOPMENT (NOT PRIORITIZED)

Several missions have been identified that address the panel's high-priority science questions, but as presently conceived, these missions require further technology development. The panel recommends that in the coming decade NASA develop the necessary technologies (for example, propulsion, power, communications, and instrumentation) to prepare for the following solar-heliospheric missions: (1) an interstellar probe, to pass through the boundaries of the heliosphere and penetrate directly into the interstellar medium with state-of-the-art instrumentation; (2) a multispacecraft mission to obtain a global view of the Sun, to reveal the Sun's polar magnetic field and internal flows, to provide three-dimensional views of coronal mass ejections, and to observe internal flows, surface magnetic fields, and the birth of active regions everywhere; and (3) a particle acceleration solar orbiter to investigate particle acceleration in the innermost heliosphere and in solar flares at an observation point 0.2 AU from the Sun.

NEW RESEARCH OPPORTUNITIES (NOT PRIORITIZED)

The panel recognizes several opportunities for new solar and heliospheric measurements that could provide breakthroughs in understanding, and recommends specifically that the following measurements and/or developments be pursued with vigor:

- Instrumentation to observe the chromosphere-corona transition region in the 300–1,000 Å band;

- Solid-state detectors for solar UV observations;
- Low-frequency helioseismology measurements to search for g-mode oscillations;
- Radar studies of the quiet and active solar corona;
- Instrumentation and techniques for imaging and mapping the global heliosphere;
- Spectral-spatial photon counting detectors for x-ray and EUV wavelengths to study reconnection on the Sun; and
- Minaturized, high-sensitivity instrumentation for in situ measurements.

POLICY ISSUES (NOT PRIORITIZED)

The panel makes several policy recommendations, some of which parallel those in the 2001 NRC report *U.S. Astronomy and Astrophysics: Managing an Integrated Program:*

- **The panel strongly encourages NASA, NSF, and other agencies that fund solar and heliospheric physics to continue interagency planning and coordination activities to optimize the science return of ground- and space-based assets. It encourages a similar high level of planning and coordination between NSF's Astronomical Sciences (AST) and Atmospheric Sciences (ATM) Divisions.**
- **The panel recommends that NSF plan for and provide comprehensive support for scientific users of its facilities. This includes support for data analysis, related theory efforts, and travel.**
- **The panel recommends that NASA support instrumentation programs, research programs, and software efforts at national and university ground-based facilities where such programs are essential to the scientific aims of specific NASA missions and/or the strategic goal of training future personnel for NASA's mission.**
- **The panel recommends that NSF and NASA study ways in which they could more effectively support education and training activities at national and university-based facilities. This support is particularly needed for training scientists with expertise in developing experiments and new instruments. The national laboratories have capabilities that could be better exploited by the universities. The panel recommends that both NSF and NASA study the idea of forming Centers of Excellence with strong university connections and tied to national facilities as a means of sustaining university-based research efforts and of educating and training the scientists, technicians, and instrument builders of the next generation. These centers should have lifetimes of 10 to 15 years and should be reviewed every 2 to 3 years to ensure they remain on track.**

1.1 INTRODUCTION

The Sun is a magnetic star, while the solar wind is both the prototype stellar wind and the only stellar wind we can hope to sample directly with in situ measurements. Solar, heliospheric, geomagnetic, and ionospheric activity are all linked via the solar wind to the variability of magnetic fields that pervade the solar atmosphere. Solar activity and resulting heliospheric disturbances can have profound impacts on our technological society, while long-term variations in the Sun's total radiative output are thought to affect Earth's climate. The Sun's magnetic field is generated by the magnetic dynamo processes occurring within the turbulent convection zone that occupies the outer 30 percent (by radius) of the Sun. These fields emerge from beneath the photosphere on a wide range of scales, from small fibril concentrations in the intergranular lanes to large active regions. The Sun exhibits a 22-year cycle of global magnetic activity, involving sunspot eruptions with well-defined rules for field polarity and emergence latitudes during the cycle. A major challenge is to understand the physical processes that produce the Sun's magnetic field, heat the corona, and accelerate the solar wind, and to understand the mechanisms that connect the solar interior to the solar atmosphere, to the heliosphere, and to Earth's magnetosphere. It is also a challenge to measure changes in the solar irradiance, to relate irradiance changes to the evolving solar magnetic field, and to understand how changes in solar irradiance might affect Earth's climate. The totality of the interactions of the Sun with the heliosphere and Earth is the focus of NASA's Sun-Earth Connections Theme and its Living With a Star program, as well as of the National Space Weather Program, led by NSF. These interactions are also of considerable practical importance to agencies such as NOAA, DOD, and DOE.

Throughout its study the Panel on the Sun and Heliospheric Physics has considered the Sun and the heliosphere as a strongly coupled system. In Section 1.2 of this report the panel highlights some of the significant accomplishments in solar-heliospheric physics from the last decade, while in Section 1.3 it identifies five basic research themes that encompass most of the major unsolved problems and unexplored frontiers in solar-heliospheric physics for the coming decade. These themes relate to perceived opportunities for the future and extend from the solar interior to the outer heliosphere:

- Exploring the solar interior,
- Understanding the quiet Sun,
- Exploring the inner heliosphere,
- Understanding the active Sun and heliosphere, and
- Exploring the outer heliosphere and the local interstellar medium.

Within these themes the panel has identified the outstanding science questions that currently are at the cutting edge of solar-heliospheric physics, and at the end of Section 1.3 it prioritizes those questions requiring new initiatives to continue present progress across a broad front.

Section 1.4 summarizes existing, in development, and approved programs that are continuing the revolution in solar and heliospheric physics but that alone will not resolve the panel's prioritized set of questions. The panel's recommended new initiatives are all linked to that set of questions and are described in detail in Section 1.5.

In Section 1.6, the panel identifies several opportunities for new measurements that could provide breakthroughs in our understanding of solar and heliospheric processes. Section 1.7 discusses the links between solar-heliospheric physics and other physics disciplines. The panel's recommendations on policy and education are provided in Section 1.8, which also provides a final recommendation on program support.

1.2 SIGNIFICANT ACCOMPLISHMENTS IN THE LAST DECADE

Recent years have witnessed an extraordinary and ongoing revolution in solar and heliospheric physics, as is evident in the following short, and necessarily limited, summary of some of the research highlights from the last decade. These highlights are spread across the wide domain of solar-heliospheric research and form the basis and framework for our recommended new initiatives for the coming decade.

INTERIOR

- Differential rotation as a function of depth has been measured in the convection zone of the Sun, and surprisingly strong velocity shears were discovered at the base and top of the zone. These shear layers are likely birthplaces of the Sun's large- and small-scale magnetic field, respectively.
- Large-scale meridional flows from 10 to 30 m/s were discovered within the convection layer.
- Signatures of newly emerging active regions were detected beneath the solar surface, yielding a potential new tool for predicting future sites of solar activity.
- Improved pressure and temperature models of the solar interior provided by helioseismic measurements still appear to rule out an astrophysical solution to the neutrino problem; neutrino oscillations were discovered.
- The solar irradiance has been shown to vary in a complicated way with the advance of the solar activity cycle.
- The helium mass fraction abundance in the Sun has been sensitively measured to be 0.2468, which disagrees with estimates for the cosmic helium abundance.

QUIET SUN

- Even the quiet Sun was found to be ceaselessly dynamic.
- More than 95 percent of the magnetic flux in the quiet Sun was discovered to emerge from beneath the surface in less than a day (Figure 1.1).
- Many coronal loops are heated within ~10,000 km of their footpoints rather than uniformly or at the loop tops.
- Coronal plasmas were discovered to be highly inhomogeneous in the cross-field direction on scales of a few hundred kilometers or less, implying a corresponding degree of inhomogeneity in the coronal heating mechanism (Figure 1.2).
- Coronal ions were discovered to have large thermal anisotropies, suggesting that ion cyclotron waves may be a dominant source of ion heating in the corona.
- Microflares were revealed to be common in and near the chromospheric network.
- Coronal magnetography was pioneered in radio and infrared measurements.

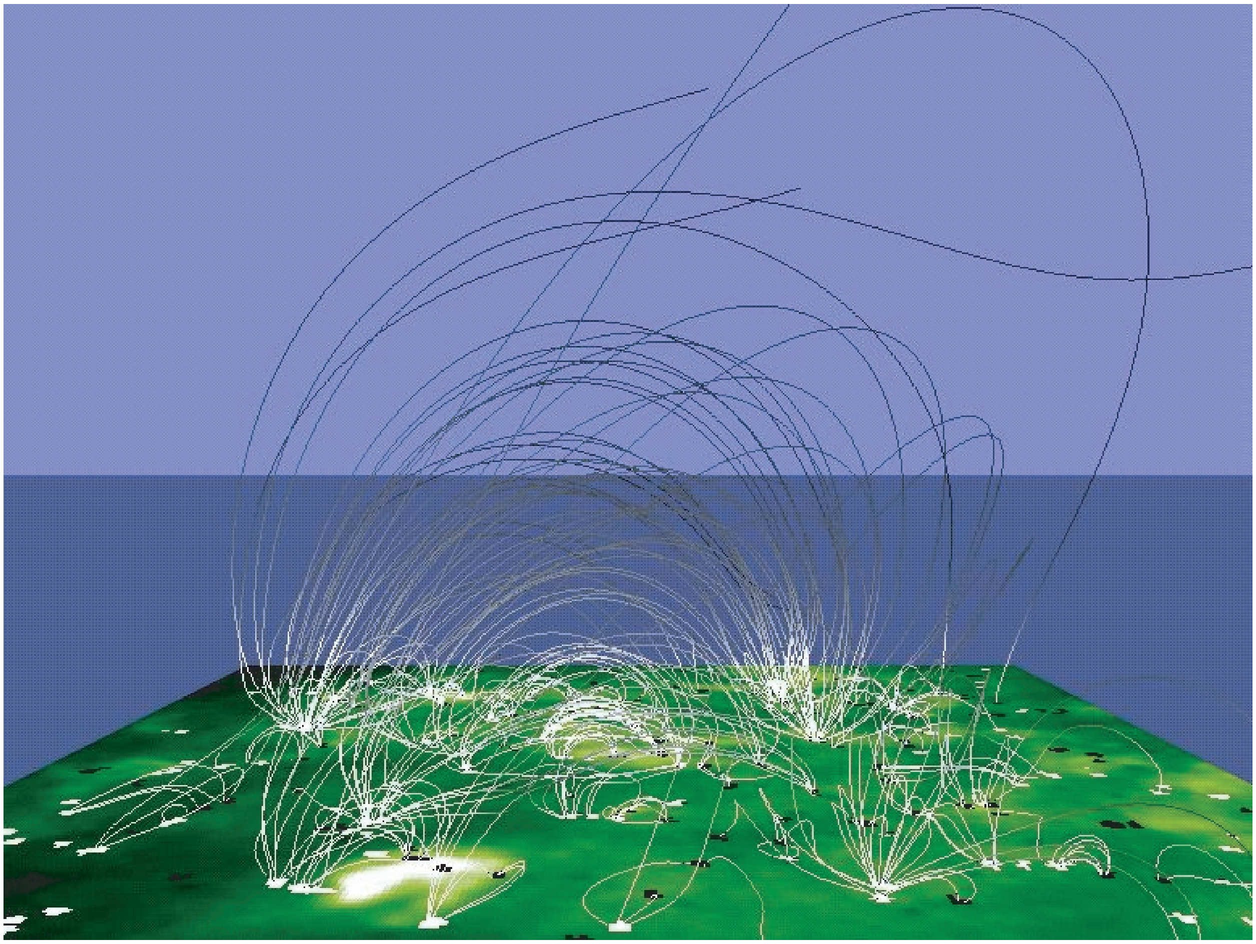


FIGURE 1.1 The magnetic "carpet"—fine-scale concentrations of magnetic flux with opposite polarities (white and black spots, respectively) that cover the solar surface and that emerge and disappear on a time scale of ~40 hours. The carpet is shown here superposed on an EUV image of the lower solar corona from the SOHO/EIT experiment. Field lines extending above the carpet as large loops are derived from models using SOHO/MDI measurements of the magnetic field in the photosphere. Courtesy of Stanford-Lockheed Institute for Space Research.

QUIET HELIOSPHERE

- Fast solar wind from high latitudes was found to dominate the three-dimensional heliosphere at solar minimum (Figure 1.3); slow and variable wind from all latitudes was found to dominate the heliosphere in the years prior to and at the solar maximum.
- A significant portion of the slow solar wind acceleration was found to occur well away from the Sun, out to at least 30 R_S.
- Slow wind and fast wind were discovered to have consistently different ionic compositions and elemental abundances, providing keys to understanding their different origins at the Sun.
- The global structure of corotating interaction regions was determined; these interaction regions have opposed north-south tilts in the opposite solar hemispheres.
- Co-rotating energetic particle events were discovered at very high solar latitudes near the solar activity minimum, suggesting a new model for the heliospheric magnetic field.
- A new source of pickup ions, thought to be solar wind deposited on and re-emitted from interplanetary dust grains, was discovered in the inner heliosphere.
- The open magnetic flux density was found to be nearly constant with latitude.

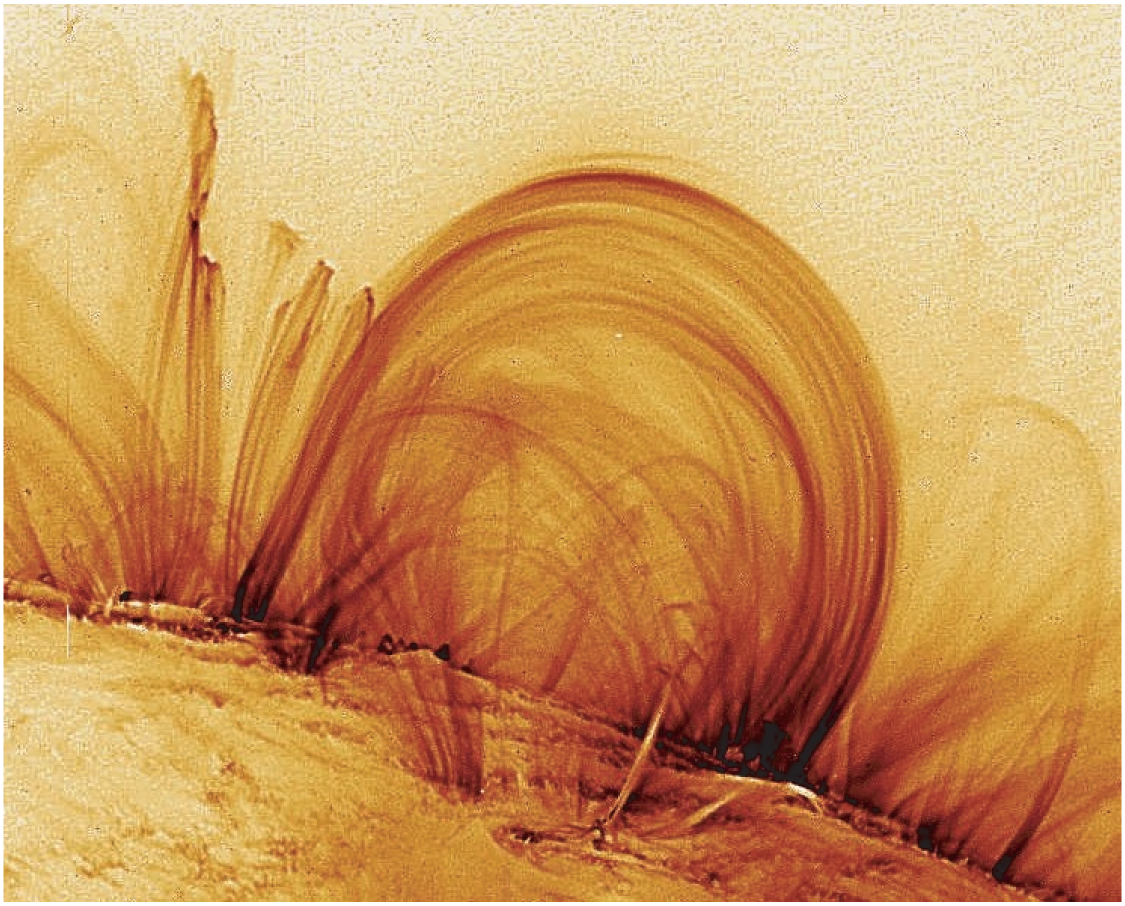


FIGURE 1.2 Coronal loops typically exhibit filamentary structure with cross-field scales down to the resolution limit (several hundred kilometers) of the TRACE telescope, as shown in this image of extreme ultraviolet FeIX emissions from coronal plasma at a temperature of ~1 MK. Courtesy of the TRACE team.

ACTIVE SUN

- The vital role of magnetic reconnection was recognized in most forms of solar activity.
- The notion that solar flares are the cause of CMEs and major space disturbances was seriously challenged; a new magnetic field- and CME-centered paradigm emerged.
- Blastlike global coronal waves were discovered propagating across the Sun in association with some large flares and CMEs.
- Two classes of solar energetic particle events were recognized in the heliosphere: impulsive events accelerated during flaring activity and the much larger gradual events accelerated in the solar wind by CME-driven shocks.
- Trans-iron nuclei with $36 \leq Z \leq 83$ were found to be overabundant in some impulsive solar energetic particle events by a factor of ~1,000, an important clue for understanding acceleration processes at the Sun.

ACTIVE HELIOSPHERE

- CMEs were firmly established as the cause of transient shock wave disturbances in the solar wind, nonrecurrent geomagnetic disturbances, gradual solar energetic particle events, and Forbush decreases of cosmic rays.
- The solar wind was discovered to be highly structured at all latitudes during the approach to and at solar maximum.

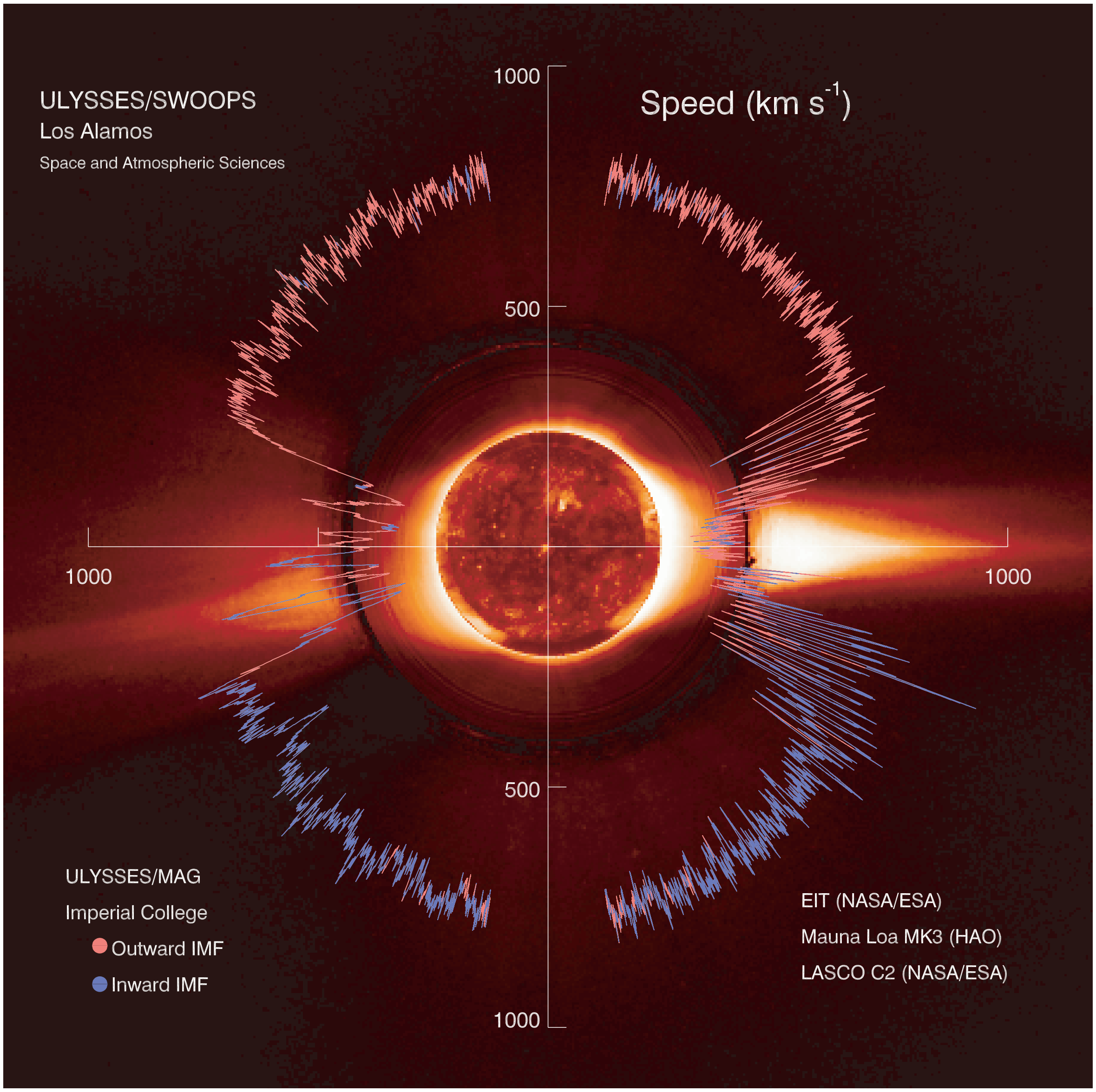


FIGURE 1.3 Solar wind speed and magnetic field polarity measured by Ulysses as a function of heliolatitude during its first polar orbit of the Sun on the declining phase of solar activity and near the solar activity minimum, overlaid with three concentric images of the corona obtained from the SOHO/EIT, the Mauna Loa coronagraph, and the C2 coronagraph on SOHO. Color coding of the speed profile indicates magnetic field polarity: red for outward pointing and blue for inward. Courtesy of the Ulysses solar wind plasma physics team.

- Mixtures of closed, open, and disconnected magnetic topologies were discovered within CMEs in the solar wind.
- The discovery that suprathermal ions are always present in the slow solar wind provided direct evidence for persistent ion acceleration in the inner heliosphere.
- The discovery of rapid intensity variations in small solar energetic particle events provided direct evidence for the random walk of open field lines on the solar surface.
- A variety of seed populations for solar energetic and corotating particle events were discovered; the relative importance of these seed populations varies from event to event.

DISTANT HELIOSPHERE

- The radio signature of the interaction of globally merged interaction regions with interstellar plasma just beyond the heliopause provided the first direct measure of the size of the heliosphere.
- Interstellar pickup ^{1}H, ^{3}He, N, O, and Ne were discovered and the cosmologically significant ^{3}He/^{4}He ratio was measured. The composition of the local interstellar material was deduced from pickup ion and anomalous cosmic ray measurements.
- A gradual deceleration of the solar wind associated with mass loading by pickup ions was detected in the distant heliosphere.
- The termination shock was confirmed to be a powerful accelerator and found to energize some particles up to GeV energies.
- Cosmic rays were discovered to have limited access to the heliosphere over the poles of the Sun, contrary to expectations, while the modulation of cosmic rays was observed to extend to distances beyond 100 AU.
- A hydrogen "wall" was predicted and detected at the nose of the heliosphere in the direction of the solar system's motion through the local interstellar medium.

1.3 SCIENCE THEMES FOR THE COMING DECADE

The panel's recommendations for new initiatives in solar and heliospheric physics in the coming decade are based on science questions arising in five basic research themes; these research themes and the underlying science questions are discussed below in some detail. At the end of this chapter, the panel prioritizes the science questions requiring new research initiatives in the coming decade.

EXPLORING THE SOLAR INTERIOR

The complex turbulence within the solar convection zone exhibits remarkable properties that have largely defied theoretical explanation. One fundamental question concerns how convection redistributes angular momentum to produce a differential rotation that varies with radius and latitude. A second question concerns local and global magnetic dynamo processes occurring within the convection zone. The panel believes local chaotic processes cause small-scale magnetic fields, while global processes create the cyclic magnetic active regions that have very well-defined rules for field polarity and emergence latitudes. These two issues are intimately linked, for the global dynamo action yielding the 22-year cycles probably is very sensitive to the Sun's internal angular velocity profiles and meridional circulations. Little is known in detail about the operation of the global dynamo, partly because it is a difficult region to observe and partly because of problems in simulating the complex and highly nonlinear dynamics in the convection zone. A major challenge for the coming decade is to understand the complex operation of this deep convective shell that is responsible for solar magnetism.

Helioseismology, the study of the acoustic p-mode oscillations of the solar interior, has provided a remarkable new window for studying dynamical processes deep within the Sun. Nearly continuous helioseismic observations from SOHO and GONG have revealed that the deeper radiative interior rotates as a solid body, possibly owing to the existence of primordial magnetic fields, while the convection zone exhibits prominent differential rotation. These two regions are joined at a complex shear layer, the tachocline (Figure 1.4). Near the surface a thin but pronounced shear boundary layer, in which the angular velocity increases with depth at intermediate and low latitudes, has been discovered. Local domain helioseismic studies have begun to probe that near-surface shear layer, revealing large-scale flow patterns that are modulated by surface activity, evolving meridional circulations, propagating banded zonal flows, and complex, evolving flow patterns that may be associated with the largest scales of deep convection (Figure 1.5). It is anticipated that the upgraded GONG experiment will operate for at least a solar cycle in order to study the evolving solar interior dynamics over a wide range of depths for an extended period. In addition, the helioseismology experiment on the Solar Dynamics Observatory (see "Operational Program and Missions," in Section 1.4) will provide an unprecedented examination of the near-surface convection zone region.

What Is the Origin of the Solar Cycle?

Observations and simple models suggest that strong, organized, toroidal magnetic fields, generated from existing poloidal fields in the tachocline, emerge through the photosphere to form active regions. The weaker poloidal field is then thought to be regenerated either by turbulence throughout the convection zone or by the breakup of twisted active regions near the surface. Re-

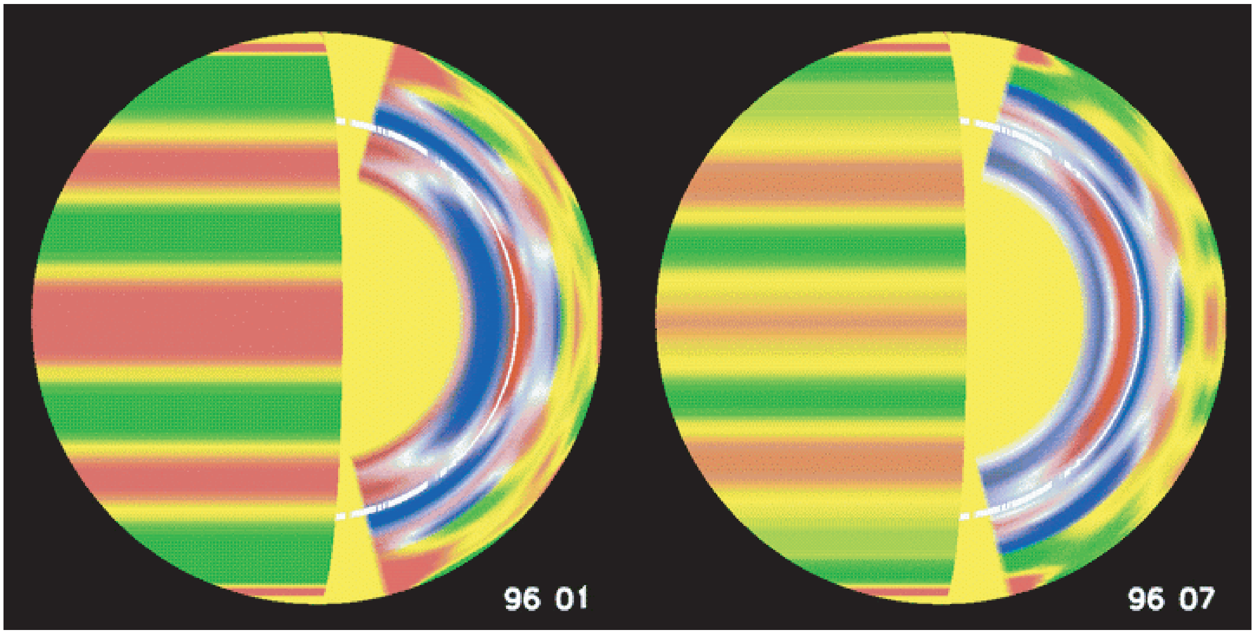


FIGURE 1.4 Cutaway images of solar rotation in January (left) and July (right) 1996 illustrating changes in internal rotation with time. Near the surface, faster rotation is indicated by red, slower by green, and intermediate by yellow; below 0.85 R, faster rotation is indicated by red and slower by blue. The left-hand side of each sphere shows the surface view. The white line indicates 0.71 R, the base of the convection zone that appears to coincide with the position of the tachocline. Courtesy of the National Solar Observatory.

cent mean-field dynamo models have shown that a dynamo that generates toroidal and poloidal fields in separate regions can produce field strengths similar to those inferred from observations. Hence, differential rotation, cyclonic turbulence, and magnetic buoyancy all must play important roles in the generation of the observed large-scale magnetic fields associated with the 22-year activity cycle. Rapid advances in massively parallel computer architectures are now enabling detailed studies of the processes thought to be crucial to the global dynamo. Although fully self-consistent MHD simulations of the dynamo will be computationally infeasible into the near future, many elements of solar dynamo processes can and should be tackled in the next decade within a focused theory initiative (see “Focused Theory/Modeling/Simulation Mission: A Virtual Sun,” in Section 1.5).

What Is the Structure Beneath the Convection Zone?

The radiative zone spans the inner 70 percent by radius of the solar interior, with the innermost 25 percent occupied by the nuclear burning core. The stratification and rotation of much of this deep interior have been probed with helioseismology using p-mode data, and the results have greatly improved stellar structure models. The helioseismic inversions also imply that the nuclear core probably has not experienced recent dynamic overturning events that would have mixed the core composition. The now well-determined models yield estimates of neutrino generation that are about threefold greater than what has been detected experimentally. The recent discovery of neutrino oscillations may reconcile neutrino observations with structural models. In addition, there are hints of interesting 1.3-year variations in rotation rates that are out of phase above and below the tachocline, suggesting dynamic magnetic links between the convection zone and the radiative interior (Figure 1.4). Our knowledge of the innermost Sun remains very incomplete, largely because few p-modes can penetrate into these central regions. The critical issues summarized above call for renewed efforts to seek to detect internal gravity (g-modes) associated with the deep interior, with astrometry possibly providing new routes to be explored (see “AlGaN Solid-State Detectors for Solar Ultraviolet Observations,” in Section 1.6).

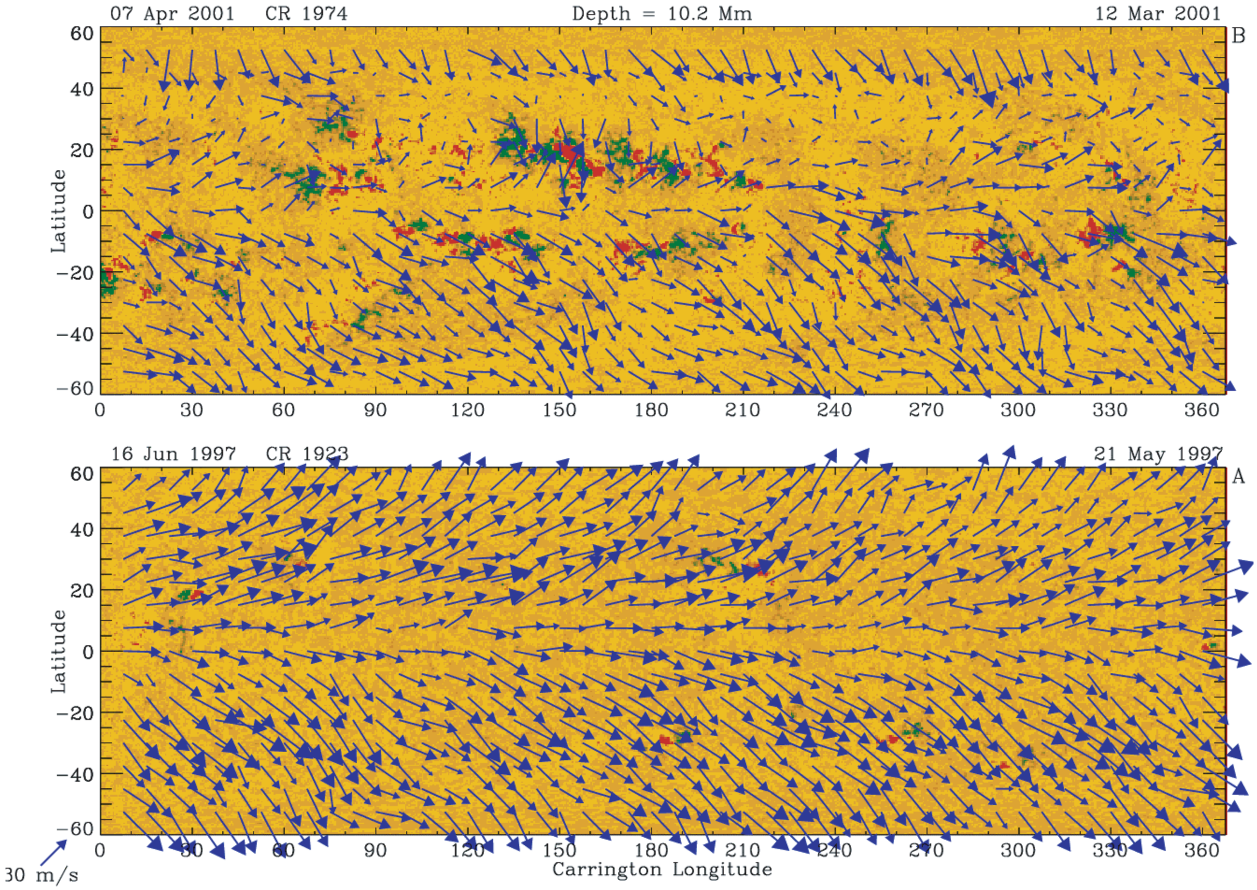


FIGURE 1.5 Local helioseismic probing of subsurface flows with ring-diagram methods reveals solar subsurface weather (SSW). Shown are synoptic maps of horizontal flows at a depth of 10.2 Mm for Carrington rotations 1923, year 1997 (bottom), and 1974, year 2001 (top), underlaid with surface magnetic field patterns (red and green indicate field polarity). In 1997, when the Sun is quiet magnetically, SSW involves meandering jets and wavy flow structures; the low-order undulations in longitude are reminiscent of jet stream flow in Earth's atmosphere. In 2001, when the Sun is active, the large-scale patterns of SSW are strikingly different, with the evolving flows exhibiting major deflections in the vicinity of active complexes. The Northern Hemisphere in 2001 shows the reversed circulation of a second meridional cell directed equatorward at midlatitudes, whereas in the Southern Hemisphere the meridional flows are consistently poleward. SOURCE: D.B. Haber, B.W. Hindman, J. Toomre, R.S. Bogart, and R.M. Larsen, 2002, Evolving submerged meridional circulation cells within the upper convection zone revealed by ring-diagram analysis, *Astrophys. J.* 570: 855–864.

UNDERSTANDING THE QUIET SUN

The "quiet Sun" is a misnomer, as a decade of intense scrutiny has revealed. The magnetic field inside and outside active regions emerges through the photosphere, spreads, and cancels constantly, accompanied by ceaseless dynamic and radiative manifestations of ongoing energy release on a wide range of spatial and temporal scales. Even long-lasting structures like prominences are far from static. Some fraction of the energy emerging through the photosphere heats the solar atmosphere to temperatures far above the 6000 K photosphere. We know that heating the corona requires only a small fraction of the energy flux heating the chromosphere, yet both regions still present major theoretical and observational challenges to understanding the heating processes. A coherent program of observations and model development for the evolving plasma and magnetic field properties in the chromosphere and corona is needed to address the following unsolved problems during the next decade.

How Does Magnetic Structure Emerge and Evolve?

Solar magnetic flux emerges on virtually every scale of convection. In active regions, where the flux distribution as a function of size is now well characterized, the flux turnover time appears to be days. In ephemeral regions, which collectively contain over 95 percent of the photospheric flux, the turnover time is much shorter, on the order of several hours. Strong magnetic fields of mixed polarity, which have spatial scales at or below the resolution limit of current instrumentation, collect in the chromospheric network, which as a result is poorly characterized. An active topic of investigation is to determine how convection distorts and twists these flux concentrations as they rise through the convective zone and penetrate the photosphere. Existing models are highly simplified and generally treat one or two discrete flux tubes instead of the complex, continuous distribution present both above and beneath the Sun's surface.

The large-scale distribution of magnetic flux on the Sun can store large amounts of energy, which may be released gradually or explosively, and ultimately determines the coupling from the Sun to the heliosphere. The small-scale, mixed-polarity fields, on the other hand, may provide most of the energy that heats the "quiet" corona and powers the solar wind. To understand the evolving distribution of fields over the Sun and into the heliosphere throughout the solar cycle, we must identify and understand the sources of magnetic flux, the flux emergence process, mechanisms for spreading the flux, and the manner in which flux disappears from the surface. Progress has been made on all fronts over the last decade, due to new observations, ideas, and models as well as to advances in computer capabilities, but more work in this area is needed. The panel recommends two pertinent observational quests for the coming decade: measurements of the time-varying magnetic field from photospheric to coronal heights and detailed measurements of the spatial and temporal scales on which magnetic flux emerges and dissipates through the solar cycle (see subsections "Frequency Agile Solar Radiotelescope," "Focused Theory/Modeling/Simulation Mission: A Virtual Sun," and "A Reconnection and Microscale Probe," in Section 1.5, and "Instrumentation to Observe the Solar Atmosphere at 300 to 1,000 Angstroms," in Section 1.6). These observations will not only open a window onto the governing processes beneath the surface but will also establish the fundamental properties that any successful model of the Sun's magnetic field must reproduce. The panel also recommends that a substantial effort be made to introduce greater realism into models of the subphotospheric rise and resulting emergence of magnetic flux with a wide range of spatial scales, as a logical extension of theoretical advances in understanding the solar dynamo (see subsection "Focused Theory/Modeling/Simulation Mission: A Virtual Sun," in Section 1.5).

What Heats the Solar Atmosphere?

Many fundamental aspects of the chromosphere are unknown: for example, the heating mechanism(s), the source of the highly inhomogeneous flows, and the relationship between the thermal and magnetic topologies. Substantial improvements in observational techniques and modeling are needed to determine which of several promising mechanisms transports subsurface energy into the chromosphere and dissipates it there. The chromosphere poses a special challenge to simulations, because the presence of neutral atoms and molecules, the comparable energy content of plasma and magnetic field, and the effects of radiative transport cannot be ignored there. The detailed thermal and dynamical structuring also remains puzzling: for example, static, one-dimensional semiempirical models do not predict the observed amount of cool gas or the thick vertical structure and do not incorporate important dynamical effects. Sorely lacking are temporally and spatially resolved observations of the thermal state of the inhomogeneous and dynamic chromosphere (see subsections "Frequency Agile Solar Radiotelescope" and "U.S. Participation in ESA's Solar Orbiter Mission," in Section 1.5), as well as physics-based models that include all relevant thermodynamic processes, radiative transport, and realistic three-dimensional topologies.

The corona is thought to be heated either by wave resonance processes or by a distribution of discrete events ("nanoflares"), although alternative models exist. Wave heating models specifically predict where and on what time scales energy deposition occurs in coronal holes and loops. To discriminate among various models, we need to derive the rate of energy deposition in coronal structures as a function of position and time from observed histories of the plasma temperature, density, and magnetic field. In addition, the observed wave spectrum must be extended to higher frequencies to determine whether the waves enter the corona from below or are generated in the corona by particle beams or reconnection (see "Primary Recommendations (Prioritized)," in Section 1.5). Numerical studies of wave heating are only beginning to include appropriate three-dimensional magnetic geometries and enough of the important physics to enable comparing calculations and observations.

The contribution of nanoflares to coronal heating and dynamics depends sensitively on several factors, including their rate of occurrence and the energy output per event. At present, the corresponding properties of observed x-ray bursts, which are orders of magnitude more energetic, have been extrapolated to small energies in order to determine the occurrence rate and energy content of nanoflares. It is thus critically important to obtain more sensitive, high-cadence observations of the smallest energy release events on the Sun. Moreover, we need to understand the mechanism responsible for energy release in nanoflares, currently believed to be reconnection and its by-products. Direct observations of the signatures of reconnection as it operates in the corona (subsection "A Reconnection and Microscale Probe," in Section 1.5), coupled with greater theoretical insight into the physics of reconnection (subsection "Focused Theory/Modeling/Simulation Mission: A Virtual Sun," in Section 1.5), would enable substantial progress to be made on the long-standing problem of coronal heating.

EXPLORING THE INNER HELIOSPHERE

The most important unexplored region of the heliosphere is within ~0.3 AU of the Sun, where the solar wind originates. The processes operating in this near-Sun region create the heliosphere and, apart from solar irradiance variations, determine the influence of the Sun on Earth and its magnetosphere. It is in the near-Sun region that the corona is heated and the plasma accelerated to supersonic speeds in both a quasi-steady form and in the episodic coronal mass ejection events. This is also the region where solar wind turbulence arises and where coronal structures, such as plumes, streamers, and holes, evolve into solar wind structures of varying speed, density, kinetic temperature, composition, and magnetic field strength. To understand the processes that heat and accelerate the solar wind and that determine the evolution of coronal structure into solar wind structure, this region must be explored directly as close to the Sun as a spacecraft can survive.

To date the solar corona has only been observed remotely, through imaging, radio sounding, spectroscopy, and downstream in situ sampling. These observations have improved dramatically over the past decade and provide tantalizing glimpses of the physical nature of the innermost heliosphere. To discriminate among competing theories and to make a major breakthrough in our understanding of the coronal origins of the solar wind, however, we need to know physical parameters that are difficult or impossible to measure remotely, including the structure, strength, and fluctuation spectrum of the ambient magnetic field, the hydromagnetic waves/turbulence down to scales comparable with the thermal ion gyroradii, phase-space distribution functions of both thermal and suprathermal particle populations, and the manner in which these quantities evolve with distance from the Sun in the innermost heliosphere. Thus, there is a compelling need for direct in situ measurements in the inner heliosphere close to the Sun (see subsection "A Solar Probe Mission," in Section 1.5). Our exploration of this heretofore unsampled environment almost certainly will also yield unanticipated results that will modify our understanding of the Sun and the heliosphere.

What Is the Origin of the Solar Wind?

Since the initial prediction of the solar wind by Parker in 1958, numerous theories have been developed for mechanism(s) by which the corona is heated and the solar wind is accelerated. Because in situ measurements at 1 AU and remote observations of UV emission from the corona by SOHO show indirect evidence for ion heating by resonance with ion-cyclotron waves, most recent theories assume the introduction of hydromagnetic waves at the base of the corona. Various sources have been suggested for this wave energy. The wave excitation may be stationary or episodic in nature. Measurement of the intensity spectrum of magnetic fluctuations in the corona and its variation is required to discriminate between possible wave origins. Measurement of ion distribution functions in or near the solar wind acceleration region will establish whether the ions are indeed heated and accelerated by ion-cyclotron waves, or perhaps by an entirely different mechanism, such as plasma jets from nanoflares.

Solar wind acceleration is closely linked to coronal heating. Although some early models suggested that the energy flux responsible for solar wind acceleration is supplied by electron heat conduction and exospheric models of the solar wind expansion are still being explored, it is now generally thought that the electrons do not carry sufficient energy to drive the coronal expansion. Instead, the solar wind acceleration is thought to be associated with the ion thermal pressure gradient and with the outward pressure gradient of Alfvén waves. Model-independent in situ measurements are required to settle this issue and to determine if the acceleration process is fundamentally the same or different in high- and low-speed wind, respectively. Direct measurements of ion and electron distribution functions and the wave fields present in the innermost heliosphere (see "A Solar Probe Mission," in Section 1.5) should answer the fundamental questions about coronal heating, solar wind acceleration, and inner heliosphere evolution that have dominated much of solar and heliospheric physics since the discovery of the corona.

Exploring the Fundamental Physics of Collisionless Plasmas

The solar wind is a superb laboratory for direct investigation of processes in collisionless plasmas that have general application to other space plasmas and to astrophysical domains where the plasma can not be directly sampled. These processes include those that govern the basic evolution of plasma particle distribution functions; the development and evolution of waves and turbulence; the generation and structure of collisionless shocks and other discontinuities; particle acceleration to relativistic energies by shocks, turbulence, and magnetic reconnection; and plasma instabilities associated with various forms of phase space "free energy." These processes occur throughout the solar wind, and much of our present understanding of these processes and phenomena is derived from past heliospheric investigations. However, the near-Sun region in the innermost heliosphere is unique. Here the ions are subject to scattering on ion-cyclotron waves, solar gravity, and magnetic focusing, and their electron distribution function is expected to change dramatically with increasing radial distance. The distribution function should reveal the development of the field-aligned beam component that carries the electron heat flux and its modification by plasma microinstabilities and wave-particle scattering. The wave power spectrum will reveal clearly for the first time the onset of hydromagnetic turbulence. Collisionless shocks driven by CMEs are expected to peak in strength in this region, with high intensities of proton-excited waves and effective acceleration of ions to GeV energies. The inner heliosphere is also the region where inner source pickup ions are created from neutrals emitted by interplanetary dust. The pickup process and modifications of pickup ion distributions are best studied in these regions. Solar energetic particle fluxes should be intense in the very inner heliosphere, and their acceleration mechanisms should not be obscured by interplanetary transport. The electromagnetic emissions due to solar energetic particle interactions with the solar atmosphere are also more intense in the inner heliosphere, and secondary neutrons can be observed at low energies before they decay. These secondary photons and particles provide high-resolution information on solar energetic particles trapped in solar active regions that are not otherwise directly observable.

UNDERSTANDING THE ACTIVE SUN AND THE HELIOSPHERE

Stressed magnetic fields are thought to provide the free energy for a wide range of transient energetic phenomena on the Sun, ranging from the smallest microjets and microflares to the largest flares and CME/filament eruptions. The largest of these energetic bursts drive solar wind disturbances, accelerate particles to high energies, and contribute to variations in the solar irradiance, all phenomena directly affecting the near-Earth environment. A primary goal of Living With a Star and the National Space Weather Program is to understand the underlying physics of these events sufficiently well to predict geoeffective solar disturbances. In part, this requires that we establish the precise relationship between flares and CMEs as well as the relationship of flares and CMEs to newly emerging magnetic flux, active prominences, transient coronal holes, and the global waves that propagate away from some flare and CME initiation sites. Some of the fundamental questions that must be answered to understand the physics of eruptive events are summarized below.

How Is Magnetic Energy Stored and Explosively Released?

Solar magnetic fields can accumulate and store excess energy through twisting, either before or after emergence through the photosphere. These stresses are transmitted to the coronal field, which, according to most models, then serves as the reservoir to be tapped by explosive energy release. Consequently, measuring and understanding the degree to which the coronal field is nonpotential is one prerequisite to understanding solar eruptive activity. To date, the energy content of nonpotential coronal magnetic fields has been estimated through extrapolations of photospheric vector magnetograms and through field morphologies inferred from EUV and soft x-ray images and/or white-light coronagraph data. Although such estimates have demonstrated that ample magnetic free energy is available to drive observed energetic phenomena, we need direct, quantitative information about the coronal magnetic field strength and topology to form a complete picture of the physical origins of these events. Prospects for the coming decade are encouraging in this regard: Both microwave imaging spectroscopy and coronal magnetography in the optical/infrared (IR) regime may provide reliable measurements of the coronal magnetic field for the first time (in Section 1.4, see "Ground-Based Programs," and in Section 1.5, see "Frequency Agile Solar Radiotelescope").

Theory and observations by Yohkoh, SOHO, and TRACE indicate that magnetic reconnection plays a key role in rapid energy release on the Sun. On macroscopic scales, results from the first generation of three-dimensional MHD models of reconnection demonstrate that two-dimensional models are of limited applicability to the real solar atmosphere. Efforts are under way to construct physics-based models for many forms of solar activity. For example, new models of CMEs have been developed in which magnetic reconnection operates in different ways: as a pre-eruption trigger, as a "tether-cutting" mechanism, or as a posteruption path to relaxation. The scenario actually preferred by the Sun in the initiation and evolution of CME eruptions remains controversial, requiring more realistic models and definitive observational tests. Because competing CME models assume distinctly different initial conditions, a pressing observational issue for the coming decade is to determine the pre-eruption magnetic field and plasma conditions. We now know from in situ observations that at least 30 percent of CMEs in the solar wind have a magnetic flux rope structure, yet it remains uncertain how and where these flux ropes are generated. In addition, we do not yet understand the origin of observed mixtures of closed, open, and (occasionally) disconnected field lines within CMEs in the solar wind, how those mixtures relate to the problem of magnetic flux buildup in the heliosphere, or how the observed anomalous ionic compositions within CMEs arise. Finally, models capable of exploring the macroscopic consequences of reconnection cannot address the microscopic means by which field lines are first severed and then reconnected, nor can they predict the kinetically determined by-products of reconnection, such as particle acceleration or plasma wave generation. To understand the reconnection process and its links to solar and heliospheric activity, we need a combined and well-focused theoretical/simulation, observational, and experimental effort (see "Frequency Agile Solar Radiotelescope," "Focused Theory/Modeling/Simulation Mission: A Virtual Sun," and "A Reconnection and Microscale Probe," in Section 1.5).

When magnetic energy is released, particles can be accelerated to high energies in several ways: as a direct consequence of reconnection-associated electric fields, by shocks, or by stochastic processes such as wave-particle interactions. Models of stochastic acceleration in flares successfully account for certain properties of energetic electrons and ions derived from both remote and in situ data, yet full closure between theory and observation remains elusive. Precisely when and where particle acceleration occurs in a flaring volume, as well as the properties of the waves on which they resonantly scatter, have not been determined. There are also questions about the relative roles of acceleration and transport in producing the charge state and composition characteristics of solar particle events that have been observed by ACE, Wind, and SOHO during the current solar maximum. Further progress in this area will require observations of ion charge state and composition at high energies; observations of low-energy neutrons and particles close to the Sun, where transport effects are minimized; and sensitive, high-resolution, hard x-ray, gamma-ray, and radio observations (see "Primary Recommendations (Prioritized)," in Section 1.5).

How Do Heliospheric Disturbances Evolve?

The recognition that CMEs are the primary drivers of heliospheric disturbances with the greatest impact on the near-Earth space environment was one of the most significant highlights of the past decade. Their intrinsic variety and evolving three-dimensional structures have been made acutely apparent by observational advances (e.g., SOHO) in detecting these events both close to and relatively far (30 R_S) from the Sun (Figure 1.6). In addition, in situ heliospheric observations and simulations have demonstrated that the effect of a given CME on the heliosphere depends critically not only on its inherent mass and momentum but also on the nature of the flow regime into which it is launched. Indeed, a particular CME can produce radically different effects at different heliocentric distances, latitudes, and longitudes (Figure 1.7). The physical processes involved in CME-disturbance propagation in the heliosphere are reasonably well understood in principle, but the application of these principles to a specific event is constrained by our lack of knowledge of initial conditions within the CME and the state of the ambient plasma and magnetic field into which the CME propagates. In order to test and validate models of disturbance propagation in the heliosphere, we need accurate measurements of CME characteristics close to the Sun and in situ measurements of helio-

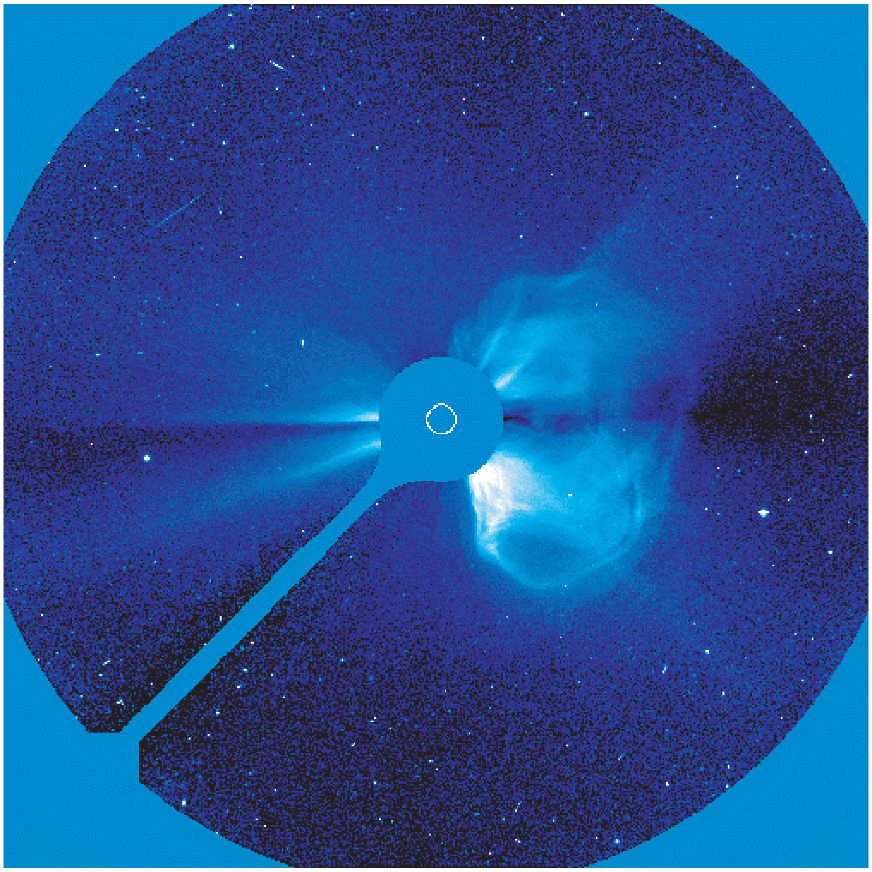


FIGURE 1.6 A coronal mass ejection observed in white light with the wide field of view coronagraph on SOHO. Courtesy of the Large Angle and Spectrometric Coronagraph (LASCO) science team.

spheric conditions before and during the disturbances as functions of heliocentric distance, latitude, and longitude (see subsection "A Multispacecraft Heliospheric Mission," in Section 1.5).

To date, CMEs have been detected near the Sun through imaging (e.g., SOHO), while heliospheric disturbances have been identified solely through their plasma, magnetic field, and energetic particle signatures, currently being provided by ACE and Ulysses. Despite key advances in analysis techniques over the past decade, unambiguous identification of CME material in the solar wind far from the Sun remains difficult. The disconnect between solar and heliospheric observations has frustrated many attempts to relate coronal and solar wind disturbances, particularly at times of frequent solar eruptions. Missions such as STEREO should begin to bridge this gap and will help provide a comprehensive picture of the entire disturbance process from birth at the Sun to heliospheric response. However, full understanding of disturbance propagation in the solar wind, an essential component of both the National Space Weather and the Living With a Star programs, will require multipoint observations at different heliocentric distances and longitudes (see subsection "A Multispacecraft Heliospheric Mission," in Section 1.5).

Acceleration at CME-driven shocks produces the most dramatic solar energetic particle events observed within the heliosphere. The physics of particle acceleration at shocks is understood in principle, but a number of crucial questions remain unanswered. For example, we do not know what seed particles are accelerated in these events, whether shocks acting alone are capable of accelerating particles to the GeV energies sometimes observed, and whether scattering near the shocks is primarily produced by ambient or self-generated waves. There is a pressing need for self-consistent numerical

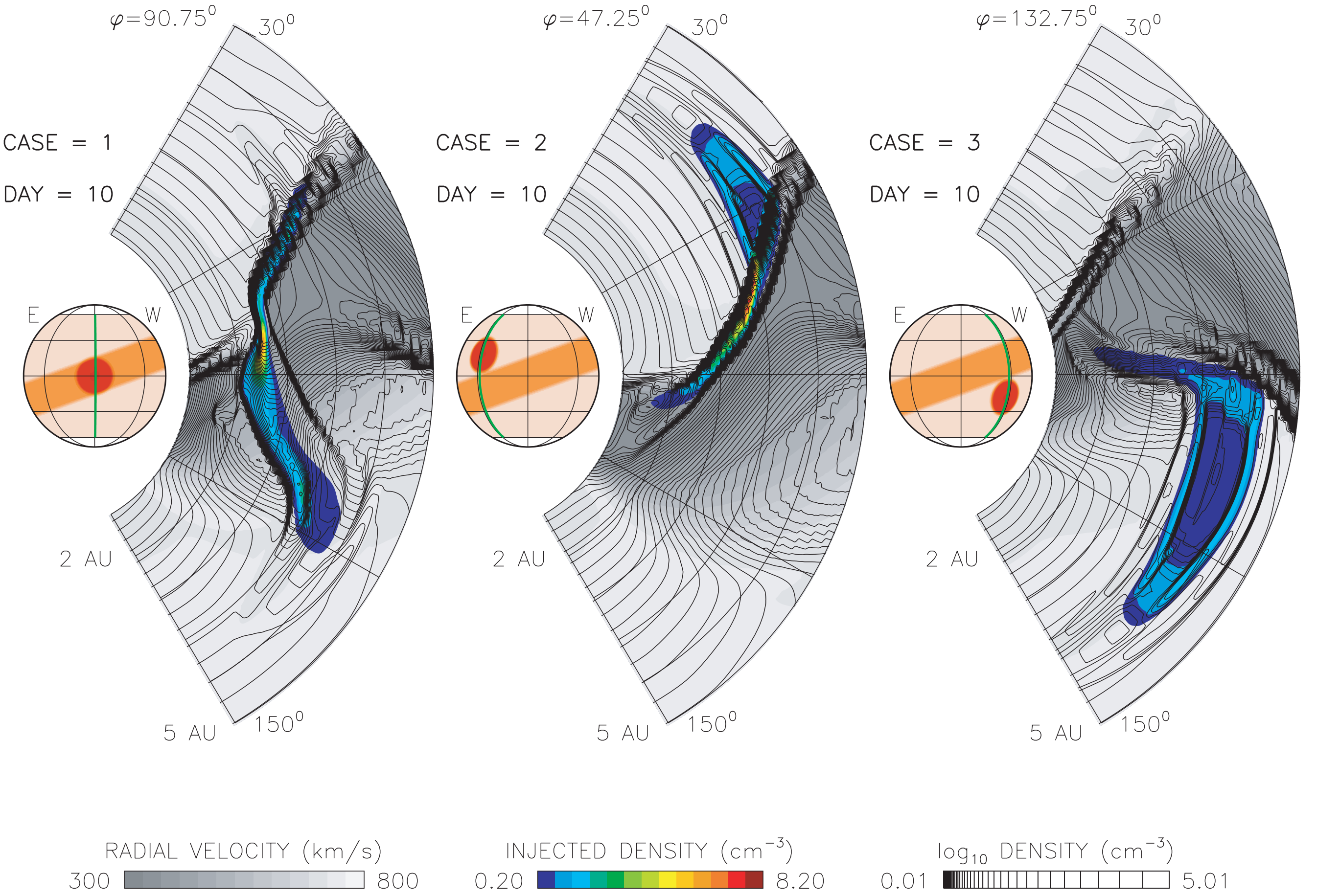


FIGURE 1.7 Simulated, color-coded, central-meridian (green line) plots of radial velocity and density for three different heliospheric disturbances initiated 10 days earlier at 0.14 AU. The solid red circle in each panel indicates the location of the initial disturbance relative to a tilted low-latitude band (orange) of slow solar wind, which is surrounded by faster wind extending up to the polar regions of the Sun. Plasma within the initial disturbance pulse, which lasted for 14 hours, had the same speed as the high-latitude wind and an internal pressure eight times greater than that within the ambient wind at all latitudes. Note the considerable latitudinal distortion of the injected material as it propagates out into the heliosphere in all three cases. Courtesy of D. Odstrcil, Astronomical Institute, Ondrejov, Czech Republic, and V. Pizzo, SEC/NOAA, Boulder, Colorado.

models of the complex interplay between particles and waves in the vicinity of the shock front and for incorporation of fluid and kinetic effects into a coherent picture of CME-driven particle acceleration. Observations designed to probe acceleration sites with both in situ and remote sensing instrumentation (see, in Section 1.5, subsections "U.S. Participation in ESA's Solar Orbiter Mission," "A Multispacecraft Heliospheric Mission," and "A Particle Acceleration Solar Orbiter") will play a critical role in deciphering the physics of shock acceleration.

EXPLORING THE OUTER HELIOSPHERE AND THE LOCAL INTERSTELLAR MEDIUM

As the solar wind expands through the solar system it eventually reaches a point where it can no longer hold off the pressure of the interstellar medium, and it undergoes a shock transition, forming the solar-wind termination shock (Figure 1.8). We have only limited knowledge of the termination shock, the interface between the solar wind and the interstellar plasma (the heliopause),

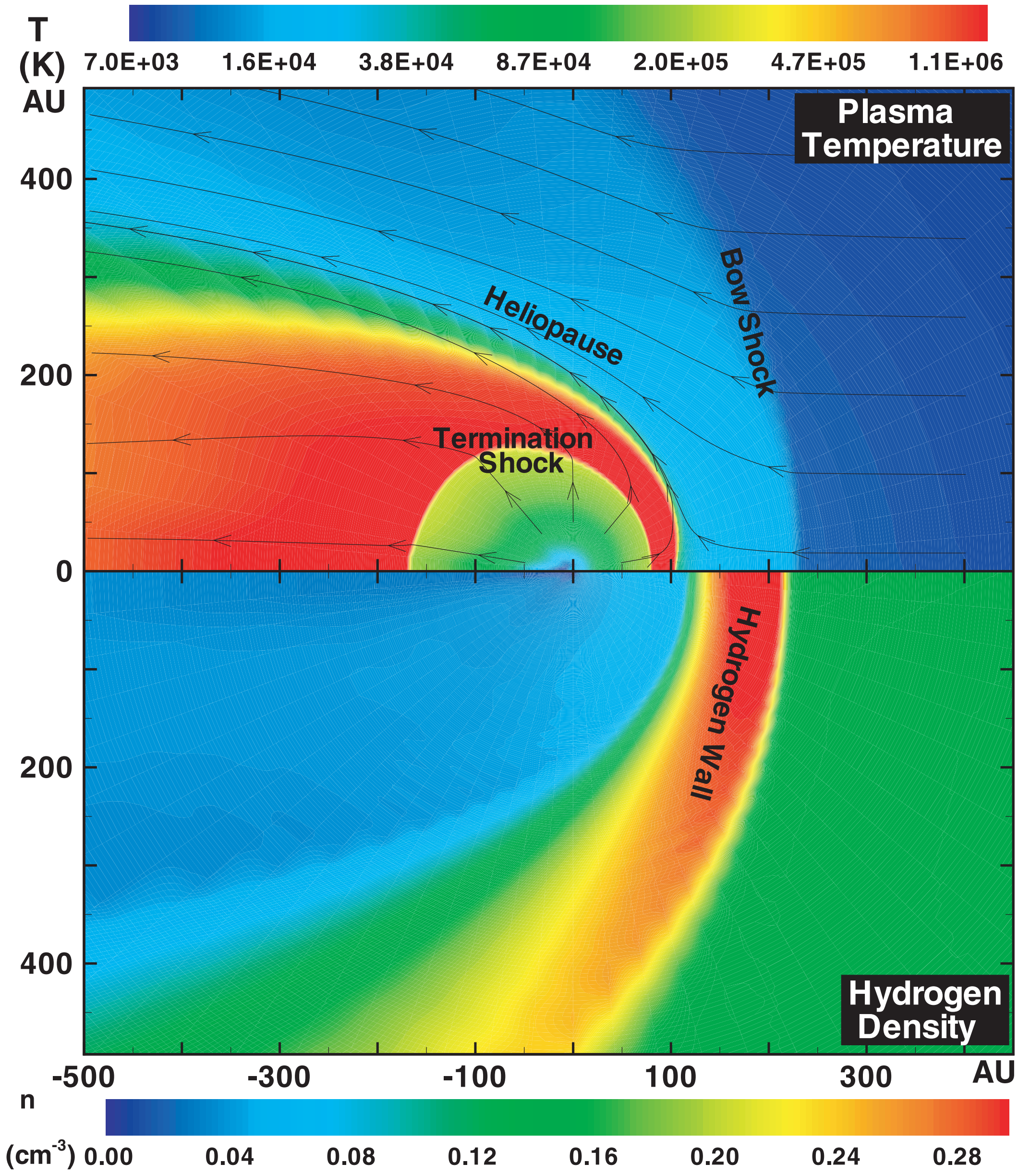


FIGURE 1.8 Model calculations simulating the interaction of the heliosphere with the interstellar medium. Prominent features include the termination shock, the heliopause, the hydrogen wall immediately upstream from the heliopause, and the upstream bow shock, which exists only if the motion of the heliosphere through the interstellar medium is supersonic. Courtesy of G. Zank, University of California, Riverside.

and the local interstellar medium beyond. Although high-energy cosmic rays, neutral interstellar gas, and large interstellar dust grains are able to cross the heliopause, the solar system is effectively shielded from the interstellar plasma, magnetic fields, low-energy cosmic rays, and small dust grains. Thus the outer heliosphere and its dynamics affect the space environment of Earth. One of the great frontiers for space science during the 21st century will be to explore the boundaries of the heliosphere and the interstellar medium beyond.

How Do the Sun and Heliosphere Interact with the Galaxy?

The Sun is moving through the LISM with a relative velocity of ~26 km/s. As interstellar plasma encounters the blunt nose of the heliosphere its flow is diverted, much like the solar wind is diverted around Earth's magnetosphere. The heliosphere and its boundaries provide a unique laboratory for studying plasma processes and the interaction of a star with its environment. The termination shock and heliopause are expected to move by ~10 to 20 AU over the solar cycle in response to changes in the dynamic pressure of the solar wind. Indeed, those boundaries may never be truly stationary. Some of these changes are caused by large interplanetary shocks created during violent outbursts of solar activity. The Voyager spacecraft have discovered intense and long-lasting radio emission from the direction of the nose of the heliosphere, occurring approximately a year after the largest episodes of solar activity. This radio emission is apparently excited as strong interplanetary shocks interact with interstellar plasma in front of the heliopause.

In the course of the Sun's journey around our galaxy it traverses a wide range of interstellar conditions. For the past few thousand years the Sun has apparently been immersed in a low-density (~0.2 cm^{-3}) cloud with a temperature of ~7000 K. The portion of the interstellar gas that is neutral (including a large fraction of H, He, N, O, Ne, and Ar) is able to penetrate the heliosphere, where some of the material is ionized by solar UV or charge exchange with the solar wind and become pickup ions. These pickup ions are convected into the outer heliosphere, where some fraction are apparently accelerated to cosmic ray energies at the termination shock and become the anomalous cosmic rays. Some of the accelerated anomalous cosmic rays charge-exchange with neutral hydrogen to produce energetic neutral atoms (ENAs). Charge exchange processes involving shock-heated solar wind beyond the termination shock produce lower-energy ENAs that may provide a means of imaging the global heliosphere. A combination of in situ and remote sensing measurements is needed to explore the many processes occurring in the dynamic boundary region.

The location of the termination shock and heliopause and the size of the heliosphere are somewhat uncertain and undoubtedly variable. As of September 1, 2001, Voyager-1 was at ~82 AU and moving outward at 3.6 AU/year, followed by Voyager-2 at ~65 AU. Although the termination shock has not yet been detected directly, evidence from several approaches indicates that the shock is currently between 85 and 100 AU from the Sun in the upstream direction, suggesting that Voyager-1 will probably encounter the shock one or more times within the next few years. These encounters will establish the scale-size of the heliosphere, provide a direct measure of the pressure of the LISM, and initiate a new era of studies of the interaction of a star with its environment. Since it will be two decades or more before a more capable spacecraft can reach this boundary region, it is critically important to keep the Voyagers operating as long as they are returning useful scientific data on the heliosphere's interaction with the LISM.

What Is the Physical Nature of the Local Interstellar Medium?

The nearby interstellar medium includes species that are predominantly ionized (e.g., C, S, and Si), those that are mostly neutral (H, He, N, O, Ne, and Ar), and others that are mainly locked up in grains (e.g., Al, Ca, and Fe). Although pickup ions and anomalous cosmic rays provide information on species that are predominantly neutral in the LISM, we presently have no information on the elemental and isotopic composition of elements that are mostly ionized, nor do we know to what extent refractory elements have condensed into grains. In addition, we have almost no knowledge of the direction and strength of the interstellar magnetic field, or of the intensity and composition of low-energy cosmic rays outside the heliosphere. Direct sampling of the composition of interstellar matter would provide a benchmark for comparison with solar-system abundances and provide constraints on galactic chemical evolution. Abundance measurements of isotopes that include ^{2}H, ^{3}He, ^{13}C, ^{18}O, ^{22}Ne, ^{26}Mg, and ^{30}Si would constrain cosmological and nucleosynthesis models and provide a more accurate picture of the evolution of the solar system, our galaxy, and the universe.

To advance studies of the outer heliosphere and LISM over the coming years we need a program that includes (1) continuation of the Voyagers, (2) a mission to several AU that would measure the neutral LISM that penetrates the inner heliosphere and provide ENA and EUV images of the heliospheric boundaries, and (3) a technology development program leading to an interstellar probe carrying advanced instrumentation to explore the boundaries of the heliosphere and our local interstellar environment. The panel believes that some of the intermediate goals of such a program can be achieved by an Explorer-class mission to several AU;

however, the panel has not prioritized Explorer-class missions in this report.

A PRIORITIZED SET OF SCIENCE QUESTIONS FOR THE COMING DECADE REQUIRING MAJOR NEW RESEARCH INITIATIVES

Some of the science questions discussed in this chapter will be addressed by existing missions and programs or those in the active queue (see Section 1.4). To continue progress across a broad front it is necessary to prioritize the outstanding science questions that will not be resolved by existing, in development, or approved missions and programs. With the above in mind, the panel has identified and prioritized the following outstanding science questions that are ripe for attack in the coming decade:

1. What physical processes are responsible for coronal heating and solar wind acceleration, and what controls the development and evolution of the solar wind in the innermost heliosphere?
2. What determines the magnetic structure of the Sun and its evolution in time, and what physical processes determine how and where magnetic flux emerges from beneath the photosphere?
3. What is the physics of explosive energy release in the solar atmosphere, and how do the resulting heliospheric disturbances evolve in space and time?
4. What is the physical nature of the outer heliosphere, and how does the heliosphere interact with the galaxy?

1.4 EXISTING AND ANTICIPATED PROGRAMS

OPERATIONAL PROGRAMS AND MISSIONS

Space-Based Missions

The current fleet of operational space missions includes a very capable array of remote-sensing and in situ instrumentation that has been responsible for many significant scientific advances over the past decade (see Section 1.2 for some highlights). Missions that are scheduled to obtain solar and heliospheric observations beyond FY 2002 are listed (alphabetically) in Table 1.1, along with the science themes that they address (the five themes are introduced in Section 1.3). If the current pace of progress is to continue over the next decade, **it is essential that key capabilities of the current program continue at least until they are replaced by missions in development (Table 1.2) or by the new initiatives (Section 1.5).** These capabilities include the ability to obtain continuous and redundant solar irradiance measurements, to image the corona in x rays and the EUV, to image CMEs in white light, to do helioseismology, and to measure the solar wind plasma, magnetic field, and energetic particle variations in near-Earth interplanetary space. Such capabilities are needed both for high-priority science objectives and to achieve the goals of the Living With a Star program. They are also essential for the routine monitoring of the Sun and heliosphere, which is critical for accurate specification and prediction of short-term space weather and longer-term space climate by NOAA and DOD and for detecting and measuring the now indisputable variability of the Sun's irradiance. With regard to the latter, the panel notes that it is only because of a sequence of overlapping, well-calibrated experiments that we are beginning to understand why solar luminosity varies and how it could change sufficiently on human time scales to affect terrestrial climate.

The panel also specifically recommends continuation of two missions that are uniquely sampling difficult-to-reach heliospheric regions: Ulysses and Voyagers 1 and 2. Ulysses should be continued as long as it is technically possible to do so, and Voyagers 1 and 2 should be continued as long as they are capable of providing measurements necessary to characterize the locations and natures of the termination shock and heliopause. The panel strongly supports NASA's plan to keep Wind operating as a backup 1-AU monitor of heliospheric conditions and to enable observational campaigns for targeted research topics. In developing its recommendations for the coming decade, the panel assumes that adequate resources will continue to be available for mission operations and data analysis for RHESSI.

Ground-Based Programs

The ground-based solar observatories are leading laboratories for the development of new solar experiments and carry out a variety of research programs. They have also provided the long-term synoptic observations crucial to understanding the underlying mechanisms of solar variability. Without the interlocking efforts of many ground observatories, we would not have some of the essential framework on which our leading models of the

TABLE 1.1 Operational Space Missions

Mission	Description	Themes
Advanced Composition Explorer (ACE)	A mission to study the elemental and isotopic composition of solar wind, solar energetic particles, and cosmic rays that also provides real-time solar wind data from L1.	2, 4, 5
Genesis	Discovery-class mission that will collect solar wind samples from L1 orbit and return them to Earth for laboratory analysis. Also provides solar wind data from L1.	2, 4
Geostationary Operational Environmental Satellites (GOES)	NOAA meteorological satellites that also provide real-time solar x-ray, solar particle, and magnetic field data. GOES-12 carries a new Solar X-ray Imager (SXI).	4
Ramaty High Energy Solar Spectrographic Imager (RHESSI)	Explorer mission to study explosive energy release in solar flares using gamma-ray and x-ray imaging.	4
Solar, Anomalous, and Magnetospheric Particle Explorer (SAMPEX)	Polar-orbiting mission that measures the composition of solar energetic particles, anomalous cosmic rays, and trapped radiation.	4, 5
Solar and Heliospheric Observatory (SOHO)	ESA-led mission to study the Sun's outer atmosphere, the solar wind, and solar internal structure using helioseismology.	1, 2, 4
Transition Region and Coronal Explorer (TRACE)	Explorer mission with an EUV/UV telescope to study dynamic phenomena on the Sun with arc-second resolution.	2, 4
Ulysses	ESA/NASA mission that will soon complete its second orbit over the poles of the Sun carrying a payload composed mainly of particles and fields instruments.	2, 4, 5
Voyager Interstellar Mission	Voyager 1 and 2 are now speeding towards the solar wind termination shock with hopes of eventually reaching the nearby interstellar medium.	5
Wind	ISTP mission that provides an extensive set of particle and field measurements of the solar wind in the vicinity of and largely upstream of Earth.	2, 4, 5

NOTE: The research themes are (1) exploring the solar interior, (2) understanding the quiet Sun, (3) exploring the inner heliosphere, (4) understanding the active Sun and heliosphere, and (5) exploring the outer heliosphere and the local interstellar medium.

TABLE 1.2 Programs in Development or Awaiting Launch

Mission	Description	Themes
National Polar-orbiting Operational Environmental Satellite System (NPOESS)	A multiagency program for polar-orbiting satellites that will provide a variety of meteorological and climate data including the total solar irradiance.	2
Solar-B	An ISAS-led mission that will measure the full solar vector magnetic field on small scales along with coordinated optical, EUV, and x-ray measurements.	2, 4
Solar Mass Ejection Imager (SMEI)	A multiagency mission led by the USAF with an all-sky camera to image CMEs as they propagate through the solar wind.	3, 4
Solar Terrestrial Relations Observatory (STEREO)	Twin spacecraft that provide stereo imaging and in situ measurements of coronal mass ejections as they propagate from the Sun to 1 AU.	3, 4
Triana	Earth-observing mission that will orbit L1; it includes solar wind and magnetic field instruments.	2, 4

NOTE: The research themes are (1) exploring the solar interior, (2) understanding the quiet Sun, (3) exploring the inner heliosphere, (4) understanding the active Sun and heliosphere, and (5) exploring the outer heliosphere and the local interstellar medium.

TABLE 1.3 Primary Ground-Based Solar Observatory Contributions

	Arecibo	BBSO	NSO/ GONG[a]	NSO/ SOLIS	NSO/ Other[b]	HAO	SOON	MSO	MWSO	NRAO	OVRO	PSPT[a]	RSTN[a]	SFSO	WSO
Longitudinal B fields		S	S	S	S			S	I						S
Vector B fields		I		S	I			S							
Ca II K-line photometry		I		S	S							S		S	
H-alpha photometry		S		S			S								
He photometry				S	S	S									
Continuum photometry				S								S		S	
Coronal emission					S			I							
Coronal continuum						S									
Helioseismic data			S						I						
10-cm radio flux	I									I	I		S		
Radio emission data	I									I	I				
Primary research themes[c]	3,4,5	2,4	1	2,3,4	2,3,4	1–4	4	1–4	1	3,4,5	3,4,5	1,2,4	3,4	2,4	2,3,4

NOTE: S, synoptic observing programs; I, intermittent observing programs.

[a] Network operated by several institutions.

[b] "Other" is NSO's Dunn Solar Telescope, the McMath-Pierce Solar Telescope, and the Evans Solar Facility.

[c] The research themes are (1) exploring the solar interior, (2) understanding the quiet Sun, (3) exploring the inner heliosphere, (4) understanding the active Sun and heliosphere, and (5) exploring the outer heliosphere and the local interstellar medium.

Key: BBSO, Big Bear Solar Observatory; NSO/GONG, National Solar Observatory/Global Oscillation Network Group; NSO/SOLIS, NSO/Synoptic Optical Long-Term Investigation of the Sun; HAO, High Altitude Observatory; SOON, Solar Optical Observing Network; MSO, Mees Solar Observatory; MWSO, Mt. Wilson Solar Observatory; NRAO, National Radio Astronomy Observatory; OVRO, Owens Valley Radio Observatory; PSPT, Precision Solar Photometric Telescope (network); RSTN, Radio Solar Telescope Network; SFSO, San Fernando Solar Observatory; WSO, Wilcox Solar Observatory.

magnetic Sun and heliosphere have been constructed. The support of these observatories for present space missions and their role in training future solar physicists are vital to the health of this field. Table 1.3, although not exhaustive, describes many of the principal components of the ground-based network for observing the Sun and also indicates roughly to which of the five research themes each observatory will contribute. In addition to the solar ground observatories, a network of neutron monitors spread around the world measures the intensity of and anisotropies in the cosmic ray flux arriving at Earth. Such measurements probe the magnetic structure of the heliosphere and contribute to research theme 4. The panel has not attempted to prioritize the ongoing programs of the ground-based institutions and facilities.

NASA Supporting Research and Technology Program

A large number of ideas, models, and instruments that are in common use in solar and heliospheric physics can trace their origins to NASA's Supporting Research and Technology (SR&T) program. This wide-ranging program supports theory and modeling, data analysis and interpretation, instrument development, and suborbital flight opportunities, as well as ground-based and laboratory studies in support of NASA's flight program. Many scientists in solar and heliospheric research had their graduate work supported in part by this program, and SR&T funding has contributed to many of the exciting discoveries and developments in this field over the past decade. The SR&T program has also contributed substantially to the development of future instruments, programs, and mission concepts considered in this report. Despite these successes, funding for the SR&T program remained relatively flat during the past decade. The panel is encouraged by reports that NASA is currently seeking to increase funding for this important component of its overall program.

During the past year NASA's SR&T program has been reorganized into 11 research "clusters." In the 2001 SR&T Senior Review, chartered by NASA to review the effectiveness of the SR&T program and recommend possible changes, the overall program was found to be very productive, and the Solar and Heliospheric Cluster was one of those judged to be deserving of increased funding. **The panel supports the Senior Review recommendation for increased SR&T funding for the Solar and Heliospheric Cluster.**

Suborbital Programs

The Suborbital program has long been an essential component of the solar/heliospheric research program. For example, rocket experiments have obtained the highest resolution x-ray images of the Sun to date, while balloon experiments have detected gamma rays from the Sun and have investigated the solar modulation of galactic cosmic-ray electrons and positrons. The capability of flying long-duration balloons in Antarctica promises to open new opportunities in solar and heliospheric physics. In addition, suborbital programs provide relatively inexpensive opportunities to test new detector technologies and to train students in detector design, development, and construction, and in campaign management and data analysis. Although the NASA, NSF, and DOD suborbital programs emphasize small projects, which are not explicitly considered in this report, **the panel strongly recommends continuation of this highly cost-effective program.** However, in spite of the importance of the suborbital program for all of solar and space physics, the current NASA management of the sounding rocket component through a commercial contractor has resulted in higher costs, less technical support, inefficient management, and a lower flight rate. The current budget and launch rate of a few per year cannot sustain a viable sounding rocket community. **The panel recommends that NASA evaluate the management structure of the sounding rocket program and increase both the level of support and the flight rate.**

PROGRAMS IN DEVELOPMENT

Space Missions

The space missions listed (alphabetically) in Table 1.2 are either awaiting launch or scheduled for launch within the next few years. Major programs in development include the following:

- *Solar-B*. A joint Japanese-U.S.-U.K. mission that will obtain coordinated optical, EUV, and x-ray measurements to determine the relationship between changes in the photospheric magnetic field and changes in the structure of the chromosphere and corona. Solar-B will measure the full vector magnetic field in the photosphere down to 0.2 arcsec to isolate elemental flux tubes with a sensitivity to the transverse component of the field vector of about 100 G.
- *Solar Terrestrial Relations Observatory (STEREO)*. A two-spacecraft mission with identical remote sensing and in situ particle and field instrumentation on both spacecraft. It will study the origin and heliospheric propagation of coronal mass ejections in the ecliptic plane. The angular separation between the two spacecraft in the ecliptic plane will gradually increase with time, with one spacecraft leading and one trailing Earth in its orbit about the Sun. Eventually one spacecraft will be able to observe in situ CME-driven disturbances imaged by the other spacecraft.

Ground-Based Programs

- *Synoptic Optical Long-term Investigations of the Sun (SOLIS)*. The program is becoming operational at NSO on Kitt Peak. It consists of a suite of three new instruments to make optical measurements of processes bearing on solar variability whose study requires well-calibrated and sustained observations over a long period of time. The instruments are a vector spectromagnetograph (high-sensitivity, full-disk measurements of magnetic fields), a full-disk imager (for spectral images of solar disk activity), and a solar spectrometer (for accurate spectral line profiles of the Sun as a star). The expected 2.3 TB of daily raw data will be processed in a manner that allows selected products to be retrieved promptly via Web interfaces.
- *Global Oscillations Network Group*. The group operates identical Michelson Doppler imaging instruments at six sites around the world to allow nearly uninterrupted full-disk observation of solar oscillations and magnetic fields. This helioseismology experiment involves very active international scientific participation both in operating the sites and in the intricate analysis of the data to make inferences about structure and dynamics within the solar interior. The GONG instruments have just been upgraded to new CCD detectors with 1024 x 1024 pixel arrays (as GONG+) and now await the implementation of high-performance computing systems to allow full primary analysis of the 32-fold greater data stream (as GONG++). It is anticipated that the GONG experiment will be operated for at least a solar cycle in order to study for an extended period how solar interior dynamics evolves over a wide range of depths.

APPROVED PROGRAMS

The following approved programs, which are not yet under full development, are prerequisites for the panel's recommended new initiatives.

Solar Dynamics Observatory

The Solar Dynamics Observatory [see note, p. 45] is a Living With a Star mission that is designed to provide essential data for understanding (1) the near-surface region of the convection zone, (2) the emergence of magnetic fields from the convection zone, (3) the resulting restructuring of the chromosphere and transition region, (4) the processes that reconfigure magnetic fields in the solar atmosphere while releasing energy gradually or explosively, (5) the magnitude and cause of both short- and long-term variations in the full-disk solar irradiance spectrum, and (6) coronal events driven by variations in the solar magnetic field.

The baseline SDO consists of the following instrumentation:

- A Helioseismic and Magnetic Imager (HMI) to study the origins of solar variability using solar oscillations and the longitudinal photospheric magnetic field;
- An Atmospheric Imaging Assembly (AIA) to study coronal energy storage and release in rapidly evolving coronal structures over a broad temperature range;
- A spectrometer for irradiance in the EUV to study both short- and long-term variations in the full-disk solar irradiance spectrum in the EUV; and
- A white-light coronagraphic imager to study transient and steady-state coronal plasma emissions.

SDO will obtain whole-Sun images and will fly in a geosynchronous orbit that allows very high data rates (~200 Mbps) almost continuously, obtaining:

- Maps of the flows, temperatures, and magnetic fields in the solar interior;
- Maps of the surface velocity pattern;
- Images of the inner corona at temperatures from 50,000 to 5,000,000 K;
- Irradiance maps of the solar surface; and
- Images of the white light corona out to ~15 R_S.

If a vector magnetograph is selected (currently a second-order priority), SDO will also obtain the magnetic field vector over the solar surface to deduce the magnetic stresses and current systems in the photosphere associated with impulsive events and evolving magnetic structures.

The SDO instrumentation represents a major advance over instruments flown on previous missions such as SOHO and TRACE. For example, the field of view of AIA will be larger than the corresponding instrument on TRACE by a factor of 26 and will have a resolution that is 25 times better than the EIT instrument on SOHO and a cadence that is 100 times faster. Moreover, AIA will obtain images in all wavelength bands simultaneously to separate temporal, thermal, and spatial evolution in the solar atmosphere. HMI will do time-distance helioseismology over the entire surface rather than in a box 10 arc minutes on a side as did SOHO. In addition, the continuous and very high SDO data rate (200 Mbps as compared with a maximum rate on SOHO of 160 kbps, available only 8 hours per day) allows transmission of high-resolution images throughout the day.

The mission plan is to obtain the following:

- Continuous high-resolution Dopplergrams for acoustic imaging of subsurface regions,
- Full-disk solar and coronal imaging in multiple wavelengths,
- Three-dimensional acoustic images that show the full history of active-region development in space and time,
- Spatial spectroscopy to follow the connection of structures from the photosphere to the corona and to measure temperature and densities,
- Irradiance monitoring to measure solar luminosity variations in the EUV, and
- High-temporal-resolution images of the solar corona in white light to follow disturbance evolution in the inner and outer corona.

The SDO measurements will complement those being made by other solar missions operating at the same time. For example, SDO will act as STEREO's third eye, placing a third coronagraph, a magnetograph, and an x-ray imager between STEREO's twin spacecraft. STEREO in turn will provide the directivity, velocity, and topology of CMEs observed as "halo" events by SDO. In addition, SDO will contribute to the objectives of Solar-B by adding continuous observations of the upper corona and higher-cadence, continuous observations of the lower corona.

Instrument proposals have been selected for SDO, and the mission is scheduled for launch in 2007. Estimated cost (panel's best guess based on information provided by NASA) is $350 million. **The panel strongly recommends that the Solar Dynamics Observatory program proceed to the full development stage.**

Advanced Technology Solar Telescope

Upcoming solar space experiments will provide intermittent observations of photospheric magnetic fields and EUV measurements of coronal plasma conditions

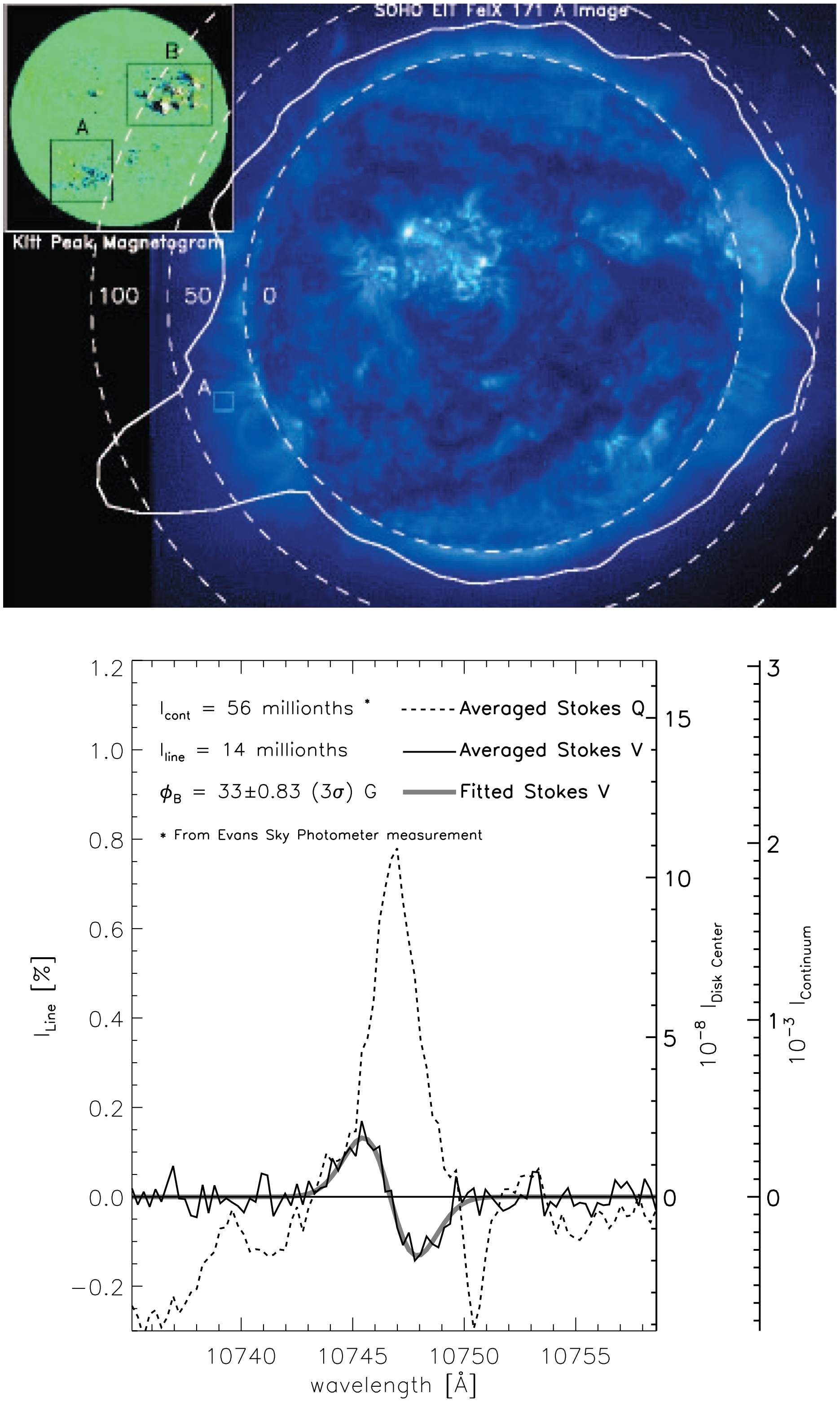


FIGURE 1.9 This SOHO/EIT Fe IX 17.1-nm image shows the structure of the solar region observed by the NSO IR spectropolarimeter, and the figure below shows Fe XIII Stokes V and I line profiles, which reveal evidence of a 30-G magnetic field. These data illustrate that coronal magnetic fields can be directly measured, although a large-aperture telescope such as the ATST is required to achieve the spatial and temporal resolution required for testing coronal models. Courtesy of H. Lin, University of Hawaii.

with angular resolution characterized by their 1-m-aperture-class telescopes. However, a larger-aperture telescope that operates in the infrared is needed to measure coronal magnetic fields above the solar limb. At optical and IR wavelengths, Zeeman splitting measurements are a model-independent remote-sensing technique for measuring magnetic fields. With current developments in infrared array technology, coronal Zeeman observations with sensitivities of a few tens of gauss have been achieved while observing regions tens of thousands of kilometers above the limb. Figure 1.9 illustrates an example of such measurements and demonstrates that coronal magnetic fields can be directly measured with presently available techniques. However, a telescope larger than any currently practical space solar telescope is required for such measurements to become a routine tool for understanding the solar atmosphere and to achieve high-angular-resolution measurements of the photospheric magnetic field.

The Advanced Technology Solar Telescope is a solar-community effort led by the National Solar Observatory and funded by NSF (presently only for the design phase). The goal is to design and build a 4-m facility that employs adaptive optics and that is aimed at determining the solar magnetic field from the photosphere up through the corona. In the lower atmosphere the telescope will achieve flux density sensitivity of a few gauss or less. ATST will provide observations of the solar atmosphere at a high temporal cadence and with better than 0.1-arcsecond resolution, which is sufficient to resolve the pressure scale height and the photon mean free path in the solar atmosphere. Thus, it will enable critical observational tests of models of solar plasma processes.

The telescope will be located at a site that offers superb seeing and sustained periods of clear weather and will replace existing NSO facilities at Kitt Peak and Sacramento Peak. It is anticipated that ATST will begin operations in the 2009–2010 time frame. The estimated cost is $70 million for construction (based on a preliminary NSO-sponsored design study) and $14.0 million per year for 5 years of operations and science (operational cost estimated to be 20 percent per year of the total construction cost). **The panel believes that the Advanced Technology Solar Telescope is an extremely promising avenue for extending our understanding of the Sun's magnetic field that takes full advantage of the research and technical capabilities of the NSO. It strongly recommends that the program proceed to the full development stage.**

1.5 RECOMMENDED NEW INITIATIVES

The panel's recommendations for new initiatives in this chapter presume that missions and programs already under development or approved for development will become operational within the coming decade to address the high-priority science objectives in solar-heliospheric physics for which they are designed.

PRIMARY RECOMMENDATIONS (PRIORITIZED)

A Solar Probe Mission

The region inward of 0.3 AU is one of the last unexplored frontiers in our solar system, the birthplace of the heliosphere itself. Remote sensing observations and in situ sampling of the solar wind far from the Sun have provided tantalizing glimpses of the physical nature of this region. However, to understand how the solar wind originates and evolves in the inner heliosphere, the panel's top science priority for the coming decade, we need direct in situ sampling of the plasma, energetic particles, magnetic field, and waves as close to the solar surface as possible. Such measurements will determine how energy flows upward in the solar atmosphere, heating the corona and accelerating the wind, and will also reveal how the wind evolves with distance in the inner heliosphere. They will revolutionize our basic understanding of the expanding solar atmosphere. The panel therefore strongly recommends a solar probe to the near-Sun region that emphasizes in situ measurements of the innermost heliosphere. The generic solar probe the panel recommends is not necessarily identical to the Solar Probe mission for which NASA released an Announcement of Opportunity in September 1999 and that placed equal emphasis on in situ and remote sensing observations. For a first solar probe, the panel strongly believes that the in situ measurements are of the highest priority and should not be compromised. In general, the panel did not find the arguments given for remote sensing measurements to be sufficiently compelling to include such measurements as a top priority on a first mission to the near-Sun region. Nevertheless, the panel appreciates that a solar probe mission provides perhaps the first opportunity to measure the photospheric magnetic field in the polar regions of the Sun via remote sensing. Such measurements will help address our second science priority and could be a secondary objective of a solar probe mission.

Solar Probe is currently listed as a part of NASA's Living With a Star Program but is not a funded mission. The panel recommends that the generic solar probe discussed here replace Solar Probe in that program; however, the overall orbital characteristics of a solar probe will be essentially the same as before. Over a 10-day period the probe will sample the corona and solar wind inside 60 R_S at both high and low heliographic latitudes. The primary goals of this mission are these:

- Locate the sources and trace the flow of energy that heats the corona.
- Determine the acceleration processes and find the source regions of the fast and slow solar wind.
- Identify the acceleration mechanisms and locate the source regions of solar energetic particles, and use these particles as remote probes of physical conditions closer to the Sun.
- Determine how the plasma, energetic particles, magnetic field, and waves evolve within the innermost heliosphere.

A secondary objective is to determine the magnetic structure of the Sun's polar regions.

To meet these science objectives, a solar probe must approach as close to the solar surface as is possible and carry a combination of in situ plasma, energetic particle, magnetic field, and plasma wave experiments that will accomplish the following:

- Measure with high spatial and temporal resolution fundamental plasma parameters (e.g., velocity distribution functions, density, temperature, velocity, composition) and magnetic field parameters (field strength and direction, power spectra) and their evolution with distance from the Sun.
- Characterize the nonthermal properties of ion and electron distribution functions.
- Identify wave modes and turbulence present in the inner heliosphere, their evolution with heliocentric distance, and their effect on particle distribution functions.
- Measure the composition, flux, and anisotropy of energetic ions and electrons with high spatial and temporal resolution.

The panel appreciates that for the high-speed wind most of the plasma heating and acceleration probably occurs at altitudes below that accessible to a solar probe. However, strong signatures of the processes that heat and accelerate the wind should still be evident in the data obtained at closest approach. Moreover, present observations indicate that the acceleration of the slow wind continues out to at least 30 R_S.

If it does not compromise the in situ measurements or add unduly to the overall mission cost, it is also desirable that a solar probe include a remote sensing experiment to measure the photospheric magnetic field in the Sun's polar regions.

The solar corona and inner heliosphere achieve their simplest form near solar activity minimum, and the data returned from a first encounter with the near-Sun region during the years surrounding solar minimum would have the most straightforward scientific interpretation. However, a mission flown at any phase of the solar activity cycle will return unsurpassed and almost certainly unexpected information on the origins of the plasmas, fields, and energetic particles that fill the heliosphere. The panel therefore believes that the launch of a solar probe should not be seriously constrained by timing relative to the solar activity cycle.

Since the payload the panel recommends for a solar probe differs somewhat from that most recently studied by NASA, **the panel urges NASA to begin immediately to study the technological issues and costs associated with a solar probe and with implementing this extremely vital mission in the coming decade.** Previous estimates for a solar probe mission, including one pass by the Sun at an altitude of 3 R_S, were ~\$600 million (estimate provided by Jet Propulsion Laboratory). This first mission to the near-Sun region will be one of the great explorations of this new century.

Frequency Agile Solar Radiotelescope

Radio imaging and radio spectroscopy provide unique insights into the quiet and active Sun. Combining radio imaging with radio spectroscopy provides a revolutionary new tool to study energy release in flares and coronal mass ejections (CMEs) and the thermal structure of the solar atmosphere in three dimensions. Moreover, radio imaging spectroscopy provides a variety of powerful techniques for measuring magnetic fields in the corona. For example, measurements of gyroresonance emission can be used to determine the magnetic field strength in active regions at the base of the corona, observations of gyrosynchrotron radiation from mildly relativistic electrons can be used to probe the coronal magnetic field in solar flares, and multiband Stokes-V observations of solar free-free emission can be utilized to provide a measure of the longitudinal field to strengths as low as a few gauss. In addition, observations of radio depolarization and Faraday rotation can be used to measure even smaller magnetic fields in particular source regions (e.g., CMEs) or along particular lines of sight within the outer corona.

The Frequency Agile Solar Radiotelescope (FASR) is a proposed (largely to NSF) ground-based, solar-dedicated, interferometric array designed to perform time-resolved, broadband imaging spectroscopy at radio frequencies from ~0.1 to ~30 GHz, thereby probing the Sun's atmosphere from the middle chromosphere to the outer corona. It is optimized to exploit a wide variety of techniques to measure coronal magnetic fields by performing time-resolved, radio imaging spectroscopy across nearly three decades in frequency. It was endorsed by the NRC's Astronomy and Astrophysics Survey Committee as a mid-size, ground-based project. Table 1.4 provides the specifications for FASR.

FASR will be a versatile and powerful instrument, providing unique data to solar and space physicists for studying basic physical processes operating in the solar atmosphere such as reconnection, plasma heating and acceleration, and electron transport. FASR will address aspects of the panel's top three science priorities by providing the following:

- Measurements of coronal magnetic fields and currents;
- Observations of nonthermal particle emission associated with flares, CMEs, and coronal shocks;
- Observations of the three-dimensional structure of the solar atmosphere from chromospheric to coronal heights;
- Observations of sites and mechanisms of coronal heating and solar wind acceleration; and
- Observations of coherent emission mechanisms and plasma wave-particle interactions.

TABLE 1.4 FASR Characteristics

Property	Specification
Frequency range	~0.1–30 GHz
Frequency resolution	~3% (2–30 GHz) <1% (0.1–2 GHz)
Time resolution	<1 sec (2–30 GHz) ~0.1 sec (0.1–2 GHz)
Antenna size	D = 3–5 m
Number of antennas	~100
Number of baselines	~5,000
Polarization	full Stokes
Independent data channels	4–8 pairs
Angular resolution	20/f arcsec (f in GHz)
Field of view	1125/(fD) arcmin (f in GHz; D in m)

FASR will contribute significantly to the Living With a Star and the National Space Weather Program by providing a number of real-time or near-real-time data products. It will also make important contributions to synoptic studies of the Sun and will yield valuable proxies for solar activity and total solar irradiance.

The estimated construction cost for FASR is $30 million (based on the estimates of individuals advocating the facility). Operations and science support will cost $6 million per year for 5 years (operational cost estimated to be 20 percent per year of the total construction cost), leading to a total cost of $60 million. This versatile and powerful instrument represents a major advance over any existing solar radio telescope (see Table 1.5), and is expected to remain the world's premier solar radio instrument for two decades or more after completion.

Focused Theory/Modeling/Simulation Mission: A Virtual Sun

Understanding the physical connections between the Sun-heliosphere and the Earth is the prime guiding principle behind the major research issues and challenges for the next decade. The Sun-heliosphere-Earth system is strongly coupled and highly nonlinear, linking spatial scales from current sheets to the size of the heliosphere and varying on time scales from fractions of a second to millennia. Its complexity has long been an obstacle to attaining a full understanding of key mechanisms and processes as well interconnections within the system, let alone construction of global models of the entire system. However, during the last decade we have broadened considerably our theoretical understanding of the Sun-heliosphere-Earth system, have collected a rich observational base from which to study it, and have witnessed a rapid development of supercomputing architectures. Together, these developments suggest that the time is ripe to complement the U.S. observational program in solar and space physics with a bold theory and modeling initiative that cuts across disciplinary boundaries.

The panel recommends a multiagency, 10-year program to initiate development of a Virtual Sun—a global numerical model of the Sun-heliosphere system consisting of interconnected modules that calculate the behavior of important structural components subject to the pertinent controlling physical processes. The Virtual Sun will be flexible enough to accommodate new modules, to allow improvements in current modules, and to run at different levels of sophistication and resolution. An essential goal is to continually integrate our ever-improv-

TABLE 1.5 Existing Solar Radio Capabilities

Observatory	Country	Angular Resolution	Frequencies	Type
Gauribidanur	India	5'	40–150 MHz	2D mapping
Nançay	France	arcmin	150–450 MHz	2D mapping
RATAN-600	Russia	240''–15''	1–20 GHz	fan beam
OVRO	United States	90''–5''	1–18 GHz	2D mapping
Siberian SRT	Russia	20''	6 GHz	2D mapping
Nobeyama	Japan	15'', 8''	17, 34 GHz	2D mapping
Itapetinga	Brazil	2'	48 GHz	Multibeam
SST	Argentina	3'–1'	212, 410 GHz	Multibeam
Metsahovi	Finland	4'–1'	22, 37, 90 GHz	Single beam
Bruny Island	Australia		3–20 MHz	Spectrograph
Izmiran	Russia		25–260 MHz	Spectrographs
Ondrejov	Czech Republic		0.8–4.5 GHz	Spectrographs
Tremsdorf	Germany		40–800 MHz	Spectrographs
Zurich	Switzerland		0.1–8 GHz	Spectrographs
Espiunica	Portugal		150–650 MHz	Spectrographs
Nançay	France		10–40 MHz	Spectrograph
Culgoora	Australia		18–1800 MHz	Spectrographs
Hiraiso	Japan		25–2500 MHz	Spectrographs
ARTEMIS	Greece		100–469 MHz	Spectrograph
Beijing/Yunnan	China		0.7–7.3 GHz	Spectrometers
DRAO	Canada		2800 MHz	Fixed frequency
Cracow	Poland		410–1450 MHz	6 fixed frequencies
SRBL	United States		0.4–15 GHz	Fixed frequencies
Nobeyama	Japan		1.0–86 GHz	7 fixed frequencies
Hiraiso	Japan		200, 500, 2800 MHz	3 fixed frequencies
Trieste	Italy		237–2695 MHz	6 fixed frequencies

ing understanding of the basic physics into the different modules and their connecting links. The physics of each parameter employed in the individual modules should be clearly understood and precisely related by basic equations to the processes that it represents. The development of an overarching framework that joins the modules consistently, and permits straightforward incorporation of observational inputs/tests and scientific visualization, will be a particularly challenging but essential aspect of this mission. The general value of end-to-end modeling has been recognized in other programs, most notably the DOD MURI and CHSSI programs and the NSF Science and Technology Centers, and by the Living With a Star science architecture team. However, the Virtual Sun is envisioned to have a far more extensive scope than those efforts.

Examples of Virtual Sun components include models of the emergence of magnetic loops in the photosphere, the evolution of MHD turbulence in the solar wind, the evolution of CME-driven disturbances, including the acceleration of energetic particles, and the interaction of the solar wind with the LISM. Two modules that the panel particularly recommends for immediate development are those pertaining to the solar dynamo and magnetic reconnection, the source of solar magnetism and its prime mechanism for releasing stored energy, respectively. Note that these research thrusts have been chosen because of their immediate readiness for substantial progress; other modules needed to build a Virtual Sun will be tackled when they are ready for a focused theoretical and numerical attack.

Major advances in supercomputing are now enabling high-resolution, three-dimensional simulations of those dynamical processes that are viewed as critical elements in the operation of the global dynamo (see “Exploring the Solar Interior,” in Section 1.3). Given the vast range of dynamical scales involved in the solar convection zone, it is unlikely that fully self-consistent MHD simulations of the global dynamo can be achieved in the very near future. However, a hierarchy of cutting-edge simulations that solve the nonlinear MHD equations without recourse to major simplifications in the physics are now tractable, thus permitting study of the primary building blocks in full dynamical detail.

A focused, coordinated initiative should also take advantage of existing and planned resources to make major breakthroughs in understanding the basic physics of three-dimensional magnetic reconnection from micro- to macroscales and in identifying and quantifying how reconnection operates under varying circumstances within the solar system. We can now model reconnection with kinetic codes that simulate particle behavior within the dissipation regions at the heart of the reconnection process, simulate the macroscopic evolution of three-dimensional reconnection with steadily improving resolution (Figure 1.10), and utilize hybrid codes that combine both kinetic and MHD approaches. An ongoing, concerted analytical and numerical effort to understand reconnection in the magnetosphere (the Geospace Environment Modeling [GEM] challenge) serves as a successful prototype for a much broader attack on the reconnection problem in the more diverse arena of the Sun and heliosphere.

A major effort by several critical-mass groups is required to develop the theoretical underpinnings and the suite of complex numerical tools necessary to address the important physics and boundary conditions; adapt these tools to the specific environments of interest; refine the models to reflect and reproduce the latest observations obtained by the facilities and missions discussed in Sections 1.4, 1.5, and 1.6; and link these models at common boundaries to form a working analogue of the Sun-heliosphere system. The panel notes also that a simplified version of the Virtual Sun will be invaluable for education and public outreach.

A coherent theory mission of this scope will require approximately 10 years to achieve its goals. The panel envisions that the program will require both continuity and community oversight to meet such ambitious goals. In particular, individual components should be competitively selected and reviewed periodically to assess quantitative progress toward completion of a working Virtual Sun model. The program will require ample access to national supercomputing facilities and funding for researchers and should therefore be a natural venue for interagency support by NASA, NSF, DOE, DOD, and NOAA. The cost will be on the order of $5 million per year (this is simply an educated guess by the panel). Supercomputing resources would be provided separately through high-performance computing programs at participating agencies. The panel contends that such an effort is crucial to achieving the goals of the substantial U.S. and international investments in ground- and space-based hardware recommended for the coming decade.

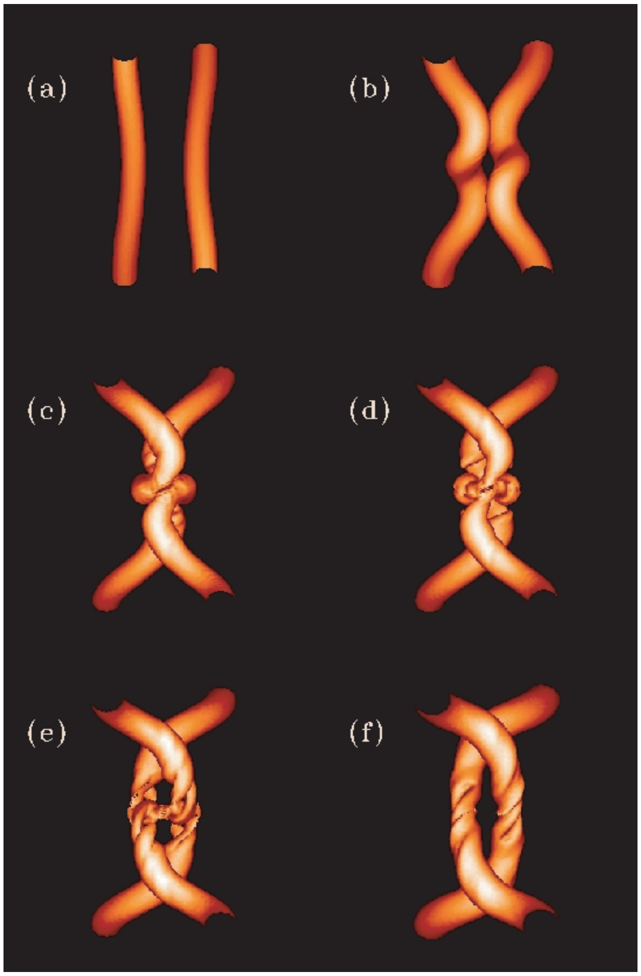


FIGURE 1.10 Three-dimensional MHD simulation (at 256^3 resolution) of the reconnection of a pair of twisted magnetic flux tubes. The images shown are isosurfaces where the magnetic field strength $|\mathbf{B}|$ equals $|\mathbf{B}|$max/2. The tubes are initially oriented perpendicular to each other and are pushed toward each other at a small fraction of the Alfvén speed (panels (a) and (b)). The flux tubes collide at two points (panel (b)) and then reconnect at these two points (panel (c)). This double reconnection allows the tubes to exchange the sections between the two contact points (panels (d) and (e)) and, as a result, by the end of the simulation the two tubes have effectively "tunneled" through each other. This topological change releases energy by converting a portion of the twist in the tubes into linking between the two tubes. The time frame shown here covers about 70 tube Alfvén crossing times. Courtesy of Mark Linton, Naval Research Laboratory.

U.S. Participation in ESA's Solar Orbiter Mission

Solar Orbiter is a natural successor to SOHO to explore the Sun and its interaction with the heliosphere.

Selected by the European Space Agency for launch in the 2008–2015 time frame, Solar Orbiter will use a unique orbital design to bring a comprehensive payload of imaging and particle and field experiments into an elliptical orbit with a perihelion of 45 R_S and an eventual maximum heliospheric latitude of 38 degrees. This orbit will, at times, provide a unique co-rotating vantage point for a combined in situ and remote sensing investigation of the Sun and heliosphere.

The primary goals of this mission are to provide data essential for understanding the solar dynamo, the magnetic structure and evolution of the solar atmosphere from the equator to the poles, and the effects of this evolution on the inner heliosphere. It thus directly addresses certain aspects of our top three science priorities. To meet these science objectives, Solar Orbiter will carry a suite of instruments that will do the following:

- Probe via helioseismology the flow patterns in and beneath the photosphere.
- Capture high-resolution EUV images of the Sun's time-varying atmosphere.
- Measure x-ray, gamma-ray, and neutron fluxes from particle acceleration processes in the solar atmosphere.
- Measure the in situ properties of the solar wind (solar wind distribution functions, velocity, density, temperature, composition, and magnetic field) and solar energetic particles.

The goals for Solar Orbiter and the types of instrumentation to be included on it are complementary to those of a solar probe. Solar Orbiter does not approach nearly as close to the Sun as does a solar probe and thus will not capture in situ the earliest phases of the solar wind expansion and evolution. On the other hand, it will remain inside 0.3 AU for a far longer period and will even co-rotate with the Sun for weeks at a time. Unlike a solar probe, it will also include a reasonably full complement of remote-sensing experiments to obtain a close-up view of the Sun, including measurements of the Sun's polar regions late in the mission.

The panel highly recommends a strong NASA involvement in ESA's Solar Orbiter mission, similar to its engagement in other successful joint ventures with ESA, such as SOHO and Ulysses. This would be a very cost-effective way for the U.S. solar and heliospheric community to address aspects of the panel's top three science priorities. The panel finds this mission would be particularly attractive if it includes both in situ and remote sensing instrumentation to investigate particle acceleration close to the Sun. The expected cost to NASA depends on the level of U.S. engagement in this high-leverage mission. The panel's very rough estimate is that a minimum of ~$100 million will be required for meaningful U.S. participation in the mission.

A Multispacecraft Heliospheric Mission

Solar wind disturbances driven by CMEs are inherently complex three-dimensional structures. Our understanding of the evolution and global extent of these disturbances has largely been built on single-point in situ measurements obtained at and beyond 1 AU, although some multispacecraft observations of heliospheric disturbances have been obtained and STEREO will provide stereoscopic imaging and two-point in situ measurements of CME-driven disturbances. The panel believes that a multispacecraft heliospheric mission, consisting of four or more spacecraft separated in both radius (inside 1 AU) and longitude and emphasizing in situ measurements, promises to dramatically advance our understanding of the global aspects of the evolution of these events. A mission of this kind will illuminate the connections between solar activity, heliospheric disturbances, and geomagnetic activity, will directly addresses the panel's third science priority, and will be an essential element of NASA's Living With a Star program.

The main scientific issues that should be addressed by a multispacecraft heliospheric mission include determination of the following:

- How CME-driven disturbances evolve as functions of heliocentric distance and longitude in a structured ambient solar wind;
- The internal structure and magnetic topologies of CMEs as well as effects of external field draping;
- How energetic particle populations accelerated by CME-driven shocks vary as functions of heliocentric distance and longitude; and
- How the large-scale structure of the solar wind evolves in longitude and with distance in the inner heliosphere in the ecliptic plane.

The optimum multispacecraft heliospheric mission will comprise a small constellation of at least four spacecraft separated in solar longitude and radius with at least one orbital perihelion at or within ~0.5 AU. The spacecraft orbits will lie in and near the ecliptic plane, while their relative positions will drift with time. Each spacecraft will carry the same complement of in situ magnetic field, plasma, energetic particle detectors, and radio wave receivers. The mission will thus provide a highly detailed two-dimensional observational slice through the ambient solar wind and through CME-driven heliospheric disturbances propagating near the ecliptic plane. The scientific return from a multispacecraft heliospheric mission will be considerably enhanced if it is timed to be concurrent with coronagraph observations of CME disturbances departing from the solar atmosphere.

A multispacecraft mission of this nature is of paramount importance in understanding and predicting the form and intensity of heliospheric disturbances impinging on Earth's magnetosphere and hence should be an essential part of NASA's Living With a Star program. The costs associated with this mission have not been extensively studied. The panel believes the mission will be equivalent to the Solar Terrestrial Probe mission and estimates it will cost of the order of $350 million.

A Reconnection and Microscale Probe

Observations and theory have long indicated that magnetic reconnection plays a key role in rapid energy release on the Sun. Although magnetic reconnection and its repercussions have long been studied intensively in Earth's magnetosphere via both observations and theory, many questions remain about its operation in the solar corona, where physical conditions differ considerably from those in the magnetosphere. Moreover, in situ sampling deep in the Sun's atmosphere is clearly out of reach.

Understanding reconnection in the solar context requires imaging (Figure 1.11) and spectroscopy of the fine-scale plasmas affected by reconnection—a challenging but technologically feasible task for the next decade. The Reconnection and Microscale (RAM) probe

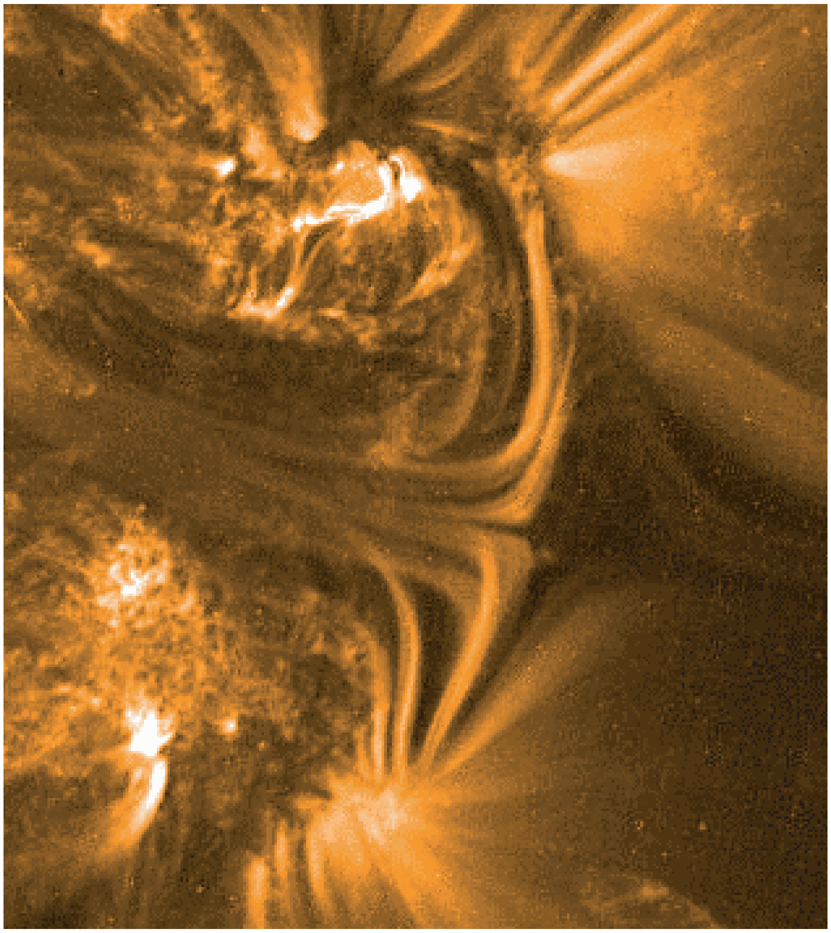


FIGURE 1.11 Close-up of a coronal null point outlined by 195-Å emission (T ~ 1.5 MK), as seen by the TRACE instrument. Courtesy of A. Title, Stanford-Lockheed Institute for Space Research, Palo Alto, California.

will be a focused, remote-sensing mission designed to meet these goals. Situated at the L1 point or in a polar Sun-synchronous orbit with a nominal lifetime of 3 to 5 years, RAM will provide multiwavelength EUV and soft X-ray observations of the Sun with the following characteristics:

- Spatial resolution comparable to estimated coronal current-sheet, null-point, and reconnection jet widths (≥10 km, or 0.01 arcsec when viewed from 1 AU);
- Temporal resolution comparable to the characteristic fast-mode wave crossing time of such structures and to typical wave periods (milliseconds to seconds);
- Spectral resolution and field of view capable of measuring flow speeds into and out of reconnection sites (~10 to 1,000 km/s);
- Simultaneous imaging of plasmas from transition-region to flare temperatures;
- Sufficient imaging sensitivity to detect emission from wave and shock compressions; and
- Full-Sun context imaging of the surrounding magnetic field and plasma conditions.

By deciphering the evolving dynamics and energetics of fine-scale coronal plasmas, a RAM probe will make major breakthroughs on several of the outstanding problems in solar physics, including coronal heating, CME initiation, and solar wind acceleration, thus addressing different aspects of the panel's top three science priorities. Most of the technologies required to build the above instruments are either direct extensions of well-tested methods or are being adapted now from existing commercial/military applications. The estimated life-cycle cost of this Solar Terrestrial Probe (STP)-class mission is $275 million (based on estimates of individuals advocating the mission).

An Interstellar Sampler Mission

The boundary between the solar wind and the LISM is one of the last unexplored regions of the heliosphere. Very little is currently known about the shape and extent of this region or the nature of the LISM. Certain aspects of these regions can be studied by a combination of remote sensing and in situ sampling techniques, thus addressing the panel's fourth science priority for the coming decade. An interstellar sampler mission (ISM), which will set the stage for a direct probe of the outer heliosphere boundaries and the LISM, will undertake the following:

- Measure directly the distribution functions of neutral interstellar H, He, N, O, Ne, and Ar and the corresponding pickup ions to establish the physical state of the interstellar gas and the heliospheric transport of pickup ions.
- Measure simultaneously the solar EUV emission and solar wind plasma properties to determine the ionization rates of the neutral gas and their time variations.
- Measure precisely the elemental and isotopic composition of the interstellar material (specifically, ^{2}H, ^{3}He, ^{22}Ne, and ^{18}O) to extract important information on the evolution of the early universe, galaxy, and Sun.
- Image the outer heliosphere to reveal its structure and dynamics, using both solar EUV emission backscattered from interstellar O^{+} beyond the heliosphere and energetic neutral hydrogen created in the heliosphere boundary region by charge-exchange interactions between neutral H and protons accelerated at the termination shock.
- Measure energetic co-rotating ion events, the contribution of "inner source" dust to heliospheric pickup ions, and the anomalous cosmic ray component that comes from pickup ions.

The ISM orbit will be elliptical with aphelion and perihelion near 4 AU and 1 AU, respectively. The spacecraft will be highly autonomous and solar-powered using conventional propulsion. The scientific objectives can be accomplished for $315 million (based on estimates of individuals advocating the mission).

FUTURE MISSIONS (UNRANKED) REQUIRING TECHNOLOGY DEVELOPMENT

Several new missions are needed to achieve high-priority science objectives, but they will not be possible unless new technology (or a new approach) is developed. In particular, some missions await advanced propulsion technology, which would also enable other exploratory missions within the solar system. There are also cases where existing technology is available, but new technology could provide enhanced capabilities or greater access to unexplored regions or new viewing perspectives.

An Interstellar Probe

Within the next decade the Voyagers will establish the size of the heliosphere and make fundamental discoveries about the termination shock and the region beyond, but their 25-year-old instruments will be un-

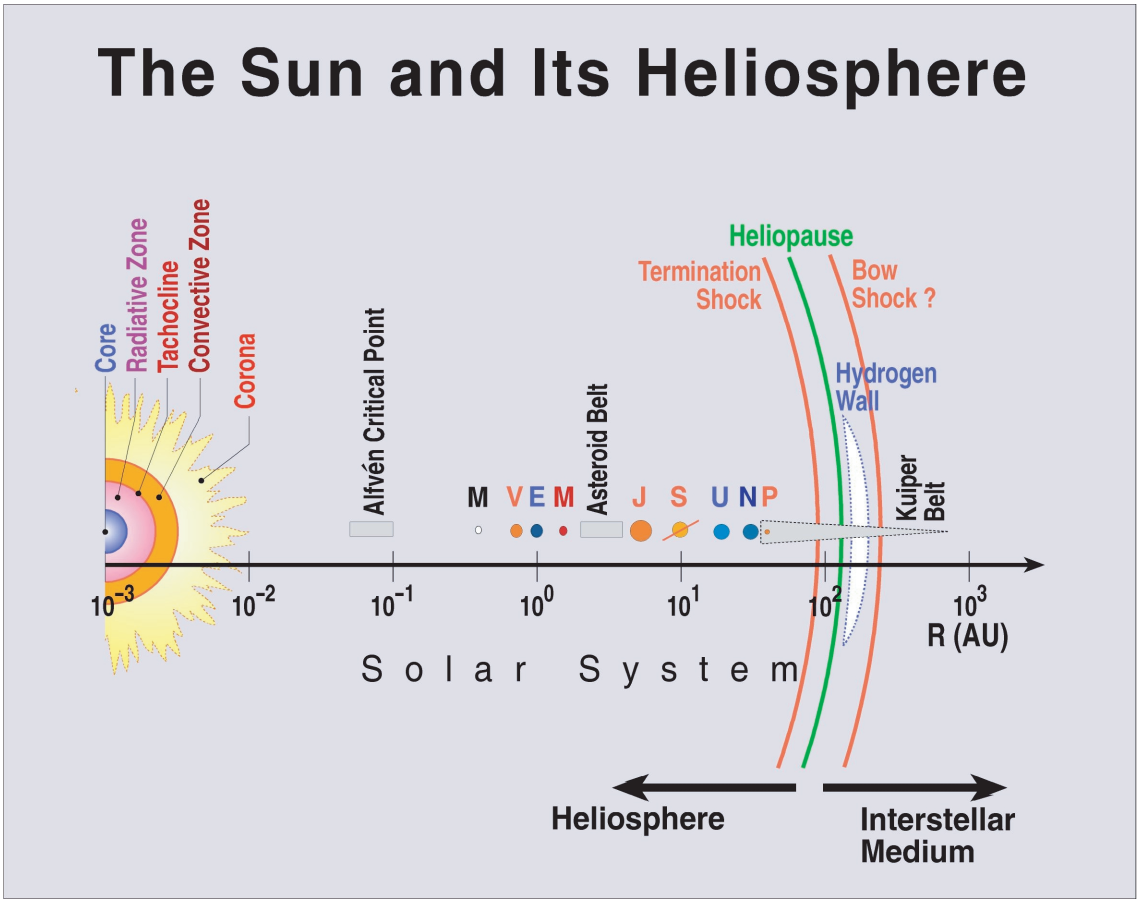


FIGURE 1.12 The heliosphere from the Sun's center to its outer boundaries, shown on a logarithmic scale stretching from 10^{-3} to 10^{3} AU. Courtesy of R. Mewaldt, California Institute of Technology.

able to measure many key properties or answer many basic questions. A new mission, an interstellar probe, is needed to carry instruments specifically designed for comprehensive study of the heliospheric boundaries and exploration of our local galactic environment (Figure 1.12). Once beyond the heliopause, an interstellar probe will discover the properties of interstellar gas, dust, the interstellar magnetic field, and low-energy cosmic rays unaffected by the heliosphere. Direct measurements could be made of the composition of interstellar dust and of the elemental and isotopic abundances of ionized and neutral gas and low-energy particles, including key species such as ^{2}H, ^{3}He, ^{13}C, and ^{26}Mg. While this mission primarily addresses the panel's fourth science priority, objectives beyond solar and heliospheric physics can also be addressed, including outstanding questions in planetary physics, astrophysics, and cosmology. Although the scientific importance of an interstellar probe has been recognized by previous National Academy of Sciences and National Research Council studies and by the NASA Strategic Plan, this mission is unlikely to happen without advanced propulsion technology.

The principal scientific goals of an interstellar probe are as follows:

- Explore the outer heliosphere and the nature of its boundaries.
- Explore the outer solar system in search of clues to its origin.
- Explore the interaction of our solar system with the interstellar medium.
- Explore the nature of the nearby interstellar medium.

To achieve these broad, interdisciplinary objectives requires advanced instruments designed for comprehensive, in situ studies of plasma, energetic particles, fields, and dust in the outer heliosphere and LISM, as well as neutral atom and UV imaging instruments to map the large-scale structure and dynamics of the outer solar system and the global heliosphere.

The core instruments for this mission could be built today, but new concepts should also be considered, such as a molecular analyzer to identify organic compounds, a small infrared spectrometer to map dust distributions and the cosmological infrared radiation background, and a small CCD camera to survey Kuiper-Belt objects >1 km in size. This mission would require a radioisotope power system, and it would benefit from advanced communications and lightweight spacecraft and instrument technologies.

To penetrate significantly into the LISM, an interstellar probe should reach at least 200 AU. To reach this far in ~15 years requires a velocity of ~14 AU/year (about four times the velocity of Voyager 1), for which advanced propulsion is required. Solar sails and nuclear-electric propulsion appear promising, but neither has been tested in space. The estimated mission cost, including 15 years of mission operations and data analysis, is ~$500 million. (This is a very rough guess made by the panel; it assumes a solar sail mission and is based on a previous JPL study.[1] When last updated by JPL, the cost estimate for a solar sail interstellar probe was $483 million, roughly consistent with the cost quoted here. The panel believes the estimate does not include the cost for developing the solar sail technology, which presumably will be used by other future missions as well.)

Sending a spacecraft beyond the heliopause to explore our local galactic neighborhood will be one of the grand scientific enterprises of this century. Developing technology to enable our first venture into the space between the stars should have very high priority.

A Global Solar Mission

Studies of the Sun's corona, activity, and interior are greatly hindered by the fact that almost all observations to date have been made from near Earth. The Sun's rotation hides much of the surface from our view for two weeks at a time, during which substantial changes can occur. Furthermore, the Sun's poles are not completely visible from anywhere in the ecliptic, so our knowledge of the magnetic field, thermal structure, and dynamics of the polar regions is incomplete at best. A global solar mission, combining spacecraft on the farside of the Sun (or distributed in longitude about the Sun) with observations from Earth and spacecraft viewing from over the solar poles, would enable a complete instantaneous picture of the Sun and its activity. Among the accomplishments expected from such a comprehensive mission are these:

- Measuring for the first time the Sun's evolving polar magnetic field and subsurface polar motions;
- Probing three-dimensional structures deep beneath the surface using two-position helioseismology techniques, and predicting when and where active regions will emerge over the entire Sun;
- Probing coronal magnetic fields with x-band and Ka-band Faraday rotation measurements;
- Examining three-dimensional thermal and magnetic structures from a polar perspective;
- Tracking the complete life cycle of active regions and coronal holes;
- Linking variations in the high-latitude heliosphere to surface conditions; and
- Measuring the global effects of dynamic events with complementary stereoscopic imaging and in situ observations.

These observations would address the panel's second and third science priorities by exploring the role of polar convection in solar magnetic field evolution, understanding the mechanisms by which magnetic field reversal occurs, exploring the azimuthal and latitudinal structure of the corona and streamer belt, and understanding the three-dimensional structure of CMEs and polar plumes.

An orbital encounter with Venus would place a spacecraft on the far side of the Sun. The more difficult task, however, will be putting a spacecraft into a polar orbit. Preliminary studies of a polar mission, the solar polar orbiter, have evaluated placing imaging and in situ instruments into a circular polar orbit at 0.5 AU, where it would circle the Sun from pole to pole 3 times a year, in 3:1 resonance with Earth. This approach would achieve the desired orbit within 3–4 years via solar sail propulsion. Solar sails have not yet been tested in space, however, so other advanced propulsion methods should also be considered. A polar perspective could also be reached with conventional propulsion using a Ulysses-type trajectory and, while the orbital period is necessarily greater, the key objectives could be met by ~150 to 300 days of observations at latitudes from 60 to 90

---

[1] Gavit, S.A. 1999. *Interstellar Probe Mission Architecture and Technology Report,* Internal document JPL-D-18410. Jet Propulsion Laboratory, Pasadena, Calif., October.

degrees. Motivated by the unique promise of these new perspectives for addressing outstanding solar-heliospheric problems, the panel recommends that a plan be formulated to achieve our first coordinated polar and farside views of the Sun.

The panel believes that each of the missions envisioned for a global view of the Sun will be in the STP class and estimates their cost to be on the order of $350 million each.

A Particle Acceleration Solar Orbiter

The past decade has witnessed remarkable progress in our understanding of particle acceleration at the Sun and in the heliosphere. While it is now recognized that solar energetic particles (SEPs) are accelerated both in solar flares and at CME-driven shock waves, the details of these particle acceleration processes remain elusive. The 1-AU separation between the primary SEP acceleration region in the corona (for both flare and shock SEPs) and the near-Earth satellites that observe SEPs makes it difficult to disentangle acceleration and propagation effects. Thus a mission must travel close to the Sun to observe SEPs in their infancy, before contamination by propagation effects. Such a mission should occur near the peak of the solar activity cycle to maximize the number of events observed.

A particle acceleration solar orbiter (PASO) with a 0.2-AU perihelion passage, hover capability, and a suite of high-energy (UV, x-ray, gamma-ray) imagers and particle (electron, neutron, ion composition) detectors, would enable us to relate acceleration signatures at the Sun directly to particles observed in space, one aspect of the panel's third science priority. The mission will undertake the following:

- Delineate the spatial and temporal evolution of SEP sources.
- Elucidate basic SEP acceleration and transport processes.
- Determine the relative importance of electrons and ions for the flare energy budget.
- Extend the size distribution of flares to lower intensities and lower energies to gain insight on coronal heating.
- Detect low-energy solar neutrons for the first time.

The estimated cost of the mission is $400 million (based on estimates of individuals advocating the mission). Solar sail technology will be required to reach 0.2 AU, although many of the key science objectives could be obtained at lower cost but with reduced sensitivity, with conventional propulsion, and a 0.3- to 0.5-AU orbit (e.g., as a component of Solar Orbiter). A PASO mission is the next step required to advance our understanding of solar energetic particles and ultimately will lead to improved predictions of this key space weather hazard.

1.6 NEW RESEARCH OPPORTUNITIES (NOT PRIORITIZED)

The panel recognizes several opportunities for new solar and heliospheric measurements that could provide breakthroughs in understanding, and it recommends specifically that the following measurements and/or developments be pursued with vigor.

INSTRUMENTATION TO OBSERVE THE SOLAR ATMOSPHERE AT 300 TO 1,000 ANGSTROMS

Full characterization of the transition between the upper chromosphere and corona is critical to understanding how nonthermal heating occurs in the solar atmosphere. These regions are difficult to model because the plasma beta drops rapidly with increasing height from greater than unity to less than 10^{-2}. Observations in these regions are also difficult, for two fundamental reasons: first, the key physical processes occur very rapidly, thus requiring optical systems with high efficiency; second, the materials commonly used for solar imaging have low reflectivity, between 300 and 3,000 Å, where most of the strong spectral lines in the upper chromosphere and transition region, as well as many coronal lines of interest, are formed. With suitable development, improvements upon two existing technologies offer potential avenues for spectroscopic imaging of the Sun in this challenging, yet crucial, wavelength range.

Most metallic films do not have high reflectivity below ~1,000 Å. The traditional materials used with broadband reflectance down to 500 Å are silicon carbide, platinum, and osmium. To make good imagers in the sub-1,000-Å region, the optical surfaces must be more uniform by a factor of 6 to 10 than for the visible spectrum. Based on recent developments in fabrication of large silicon carbide optics and chemical vapor deposition techniques for making polishable high-reflectance coatings, it should be possible to develop silicon carbide optics with surfaces sufficiently good to make high-quality, broadband imagers down to ~500 Å. However, maintaining the shape and surface quality in optics constructed of these materials remains a problem.

On the other hand, normal-incidence EUV coatings, which allow narrow-band spectral imaging, necessarily have high reflectivity over narrow spectral ranges and do not work well above 300 Å. Recent research has shown that materials exist that allow fabrication of multilayer coatings at wavelengths as long as 400 Å, and there is promise of coating systems for even longer wavelengths. Relatively modest research efforts on normal incidence coatings should thus allow the production of spectral imagers at wavelengths well above 300 Å.

AlGaN SOLID-STATE DETECTORS FOR SOLAR ULTRAVIOLET OBSERVATIONS

Solid-state detectors made from aluminum-gallium nitride ($Al_xGa_{1-x}N$) materials have outstanding potential to become the detectors of choice for most UV applications in the wavelength range 1,000 to 3,000 Å. AlGaN detectors are lightweight and compact and can record photons at very high rates. AlGaN detectors can readily make use of CMOS technologies so that individual pixels can be addressed randomly for highly versatile and rapid readouts. AlGaN is a wide-bandgap material, making the detector inherently solar blind at visible wavelengths, operable at room temperatures without thermal backgrounds, and radiation hard. In addition, AlGaN devices offer very high UV detective quantum efficiencies (DQE > 85 percent) that are very stable. AlGaN detectors that are solar blind and that have DQEs in excess of 60 percent have already been demonstrated in the laboratory.

While AlGaN devices offer great promise as superb UV detectors, the technology is still in its infancy and will require substantial and prolonged development. The primary issue for UV solar applications is a very large nonthermal background produced by material defects. Fortunately, there is considerable (multibillion dollar) commercial and military interest in this material, including the desire for a solar-blind UV image sensor. Tremendous strides have been made in recent years to reduce the defects and dislocations responsible for these unwanted backgrounds. The time is ripe to take advantage of the enormous worldwide investment being made in AlGaN materials research to develop UV image sensors that meet the unique set of requirements for solar and heliospheric applications.

LOW-FREQUENCY HELIOSEISMOLOGY

During the past decade, acoustic helioseismology was highly productive as a probe of the solar interior. Nevertheless, our last unexplored region of the solar interior remains the innermost core—the region ultimately responsible for defining the solar luminosity and the solar neutrino flux. Unfortunately, the inner 10 to 20 percent of the Sun's radius is nearly impenetrable to surface acoustic waves.

A potential exists to sense the kinematic and thermodynamic structure of the core remotely by measuring lower-frequency waves in the visible photosphere—modes with periods of tens of minutes to many hours, in contrast to 5-min acoustic modes. While buoyancy waves (solar "g-modes") may still bear fruit for this problem, the absence of progress on this front, despite two decades of effort, suggests that new techniques should be sought.

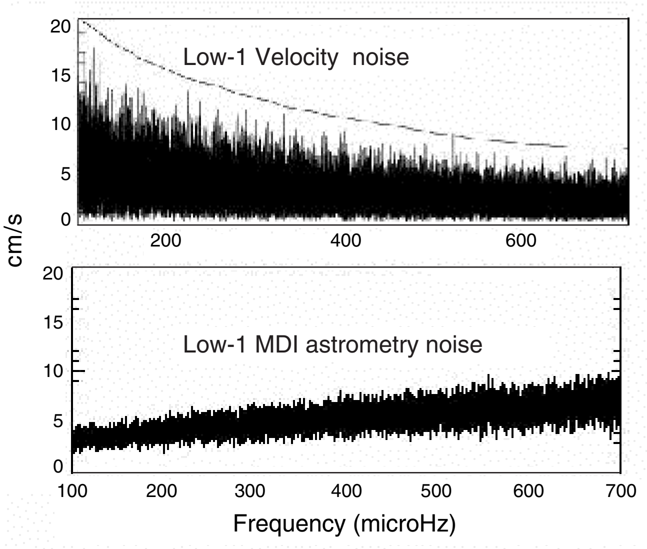


FIGURE 1.13 The solar background noise at low frequencies, a regime where g-modes or r-modes could be detected, is illustrated here. Root-mean-square astrometric noise levels, expressed in cm/s, decrease at lower frequencies (as computed from SOHO/MDI limb astrometry, lower panel), while the solar velocity noise increases with lower frequency (upper panel). Courtesy of T. Appourchaux, Solar System Division, ESA-European Space Research and Technology Center.

A promising avenue may be the use of solar astrometric techniques, which are especially sensitive to low-frequency g-mode and inertial (r-mode) oscillations. Figure 1.13 shows how the solar noise background declines, and thus the modal sensitivity increases, with decreasing signal frequency using astrometric techniques. In contrast, current Doppler methods suffer from increasing noise with lower frequency.

RADAR STUDIES OF THE QUIET AND ACTIVE SOLAR CORONA

Although passive radio-wave observations of the Sun and interplanetary medium are well established as important observational tools, active radio ranging of the near-Sun environment has had little development since the pioneering work of J.C. James in the 1960s, which demonstrated that radar echoes could be obtained under many conditions. The data contained a variety of puzzling features, including anomalously high radar cross sections and Doppler features. A key element of the interpretational difficulty was our primitive understanding of coronal structure and dynamic activity at that time: for example, neither CMEs nor coronal holes had yet been discovered. Moreover, synoptic solar magnetic field observations are not available for that time period so it is impossible, even retrospectively, to relate the observed echoes with likely coronal structures.

Now, however, detailed contextual observations of the corona are routinely available and radar ranging techniques are mature. Given these advances, any reasonable opportunity to explore the possibilities of solar radar as an observing tool should be exploited. Active sensing of the solar corona offers the possibility of mapping coronal structures, of detecting CMEs and measuring their radial velocity components, and of exploring wave-wave interactions in the solar corona.

INSTRUMENTATION AND TECHNIQUES FOR IMAGING AND MAPPING THE GLOBAL HELIOSPHERE

Most in situ heliospheric observations are local in nature, and it is often difficult to understand the global context of such measurements. New technologies for detecting energetic neutral atoms (ENA) and certain UV lines show considerable promise for performing measurements that can directly determine key aspects of the three-dimensional structure of the heliosphere and, possibly, its time dependence. Similarly, data analysis approaches using tomographic techniques have demonstrated new capabilities for imaging the large-scale structure of the inner heliosphere. Continued progress in these areas will depend critically on integrating new observations with advanced models in an iterative manner. The panel recommends further development of the following new technologies for imaging the global heliosphere.

- *ENA imaging*. Newly developed ENA-imaging instruments and techniques have been demonstrated for magnetospheric applications by the IMAGE mission. The same imaging principles should apply to the remote sensing of energetic particle acceleration processes on the Sun, in interplanetary space, and at the outer boundaries of the heliosphere. Although current techniques have been limited by very small geometric factors, more sensitive designs using new detection methods and background-suppression techniques appear promising.
- *UV imaging*. It has recently been proposed that the global shape of the heliosphere can be mapped by measuring solar EUV line radiation that is back-scattered by galactic plasma beyond the heliopause. Mapping of the heliopause can be carried out using the oxygen (O^+) resonance line (83.4 nm) provided that sufficient improvement in detection sensitivity can be achieved. This technique may also provide an independent measure of the ionization state of the local interstellar medium.
- *Heliospheric tomography*. Within the inner heliosphere, tomographic techniques have been applied to white-light images of the corona and radio observations of interplanetary scintillation in efforts to image large-scale solar wind structures. Considerable progress has been made in reconstructing images of quasi-steady corotating structures, and these techniques are now being applied to transient disturbances such as those driven by CMEs. Given the importance of understanding such structures on a heliospheric scale, the panel recommends the continued development of tomographic techniques to image structures in the inner heliosphere.

SPECTRAL-SPATIAL PHOTON-COUNTING DETECTORS FOR THE X-RAY AND EUV REGIONS

Localized plasma heating and jetlike Alfvénic flows are key signatures of magnetic reconnection that are difficult to detect with present-day instrumentation. To probe and understand reconnection in solar activity, a new type of instrument must be developed that meets the following stringent requirements: (1) the ability to resolve a wide range of plasma velocities from ~10 to 1,000 km/s both along and transverse to the line of sight;

(2) enough sensitivity to capture the onset of a CME, flare, or other reconnection-driven event; and (3) spatial resolution commensurate with the scale of the reconnection site and the width of the jets, coupled with a sufficiently large field of view to capture the dynamic and topological effects of reconnection.

A new class of detectors—microcalorimeters, superconducting tunnel junctions, and transition edge sensors, which simultaneously provide spatial, spectral, and temporal information—shows promise as near-ideal sensors for studying magnetic reconnection. While these detectors differ in their modes of operation, all are photon-counting detectors that have high (2 to 10 eV) energy resolution in the soft x-ray range from 1 to 10 keV, temporal resolution of hundreds of microseconds to milliseconds, and the capability of being fabricated as pixel arrays. Astronomical experiments employing such detectors have already flown in space, and detectors in this class are the baseline for several contemplated future astronomical missions such as Constellation X.

Recommended missions such as a reconnection and microscale probe (see, in Section 1.5, subsection "A Reconnection and Microscale Probe") would benefit greatly from full development of these imaging spectroscopy technologies. For solar physics applications these detectors must be adapted to handle higher count rates than are experienced in astronomical applications. Suitably modified, they will be able to detect and measure the highly time-dependent and directional flows that are the hallmarks of magnetic reconnection in the solar corona, both along and transverse to the line of sight and to record the associated thermalization of the ambient plasma. Experiments employing these detectors can provide the first fully three-dimensional picture of reconnection in energetic solar phenomena, from coronal heating to coronal mass ejections, thus answering some of the most fundamental questions in solar and heliospheric physics (themes 2 and 4).

MINIATURIZED, HIGH-SENSITIVITY INSTRUMENTATION FOR IN SITU MEASUREMENTS

The scientific payloads for exploratory missions such as an interstellar probe will be allotted only limited mass and power resources. Although most of the core instruments envisioned for this mission have considerable flight heritage and could be built today, there is a need for miniaturized, low-power versions of existing designs if the full range of science objectives for this and other exploratory missions is to be achieved. In addition, new technology is needed to develop instruments such as a small neutral gas spectrometer, a small molecular analyzer, and a small infrared spectrometer.

Missions within the heliosphere, such as ISM, that will remotely sample the interstellar medium need a high-sensitivity neutral gas spectrometer to measure a variety of interstellar neutral species, a high-sensitivity pickup ion spectrometer to measure the isotopic composition of the interstellar gas, and a high-energy spectrometer to measure ionic charge states to higher energies than is presently possible.

1.7 CONNECTIONS TO OTHER PHYSICS DISCIPLINES

Solar-heliospheric physics shares with other physics disciplines an interest in nuclear fusion, magnetic dynamos, magnetoconvection and turbulence, collisionless shocks, magnetic reconnection, energetic particle acceleration and transport, emission and absorption of electromagnetic radiation, instabilities (kinetic, fluid, and MHD), neutrino detection, multiwavelength spectroscopy, and other diagnostic techniques. Over the past several decades, cross-fertilization among these various disciplines has advanced our understanding of the Sun and heliosphere. The panel therefore recommends the continuation of this highly fruitful interchange throughout the next decade, with particular focus on three research areas: atomic physics, nuclear physics, and plasma physics.

ATOMIC PHYSICS

Spectroscopy is used to map the physical and dynamic properties of the solar atmosphere and has provided opportunities for fundamental discoveries such as the existence of the element helium. Recent advances in computational technology are enabling more accurate multilevel atomic-physics calculations, which will improve line identification and interpretation, and more detailed calculations of the time-dependent ionization balance essential for deciphering the sources of chromospheric and coronal heating. As high-cadence, high-resolution observations become available, a substantial effort will be required to incorporate these new theoretical capabilities and laboratory measurements into analysis and interpretation of the data.

The interpretation of certain types of particle observations depends critically on our knowledge of several complex processes, including EUV and collisional ionization, recombination, charge exchange, and electron stripping of high-energy particles. Accurate modeling of these processes, including the effects of charge-to-mass ratio dependent fractionation during particle acceleration and transport, is essential for interpreting the measured ionic charge state and elemental composition of thermal, suprathermal, and more energetic particles in the solar wind. For example, comparisons between ionic charge-state observations and models of shock acceleration near the Sun, in interplanetary space, and at the termination shock have been used recently to interpret the time scale for such acceleration processes. The Sun-heliosphere system provides unique opportunities for observing these processes, but proper interpretation of energetic ion composition measurements requires cross sections that have been measured in the laboratory or calculated with high precision.

NUCLEAR PHYSICS

Attempts to resolve the “solar neutrino problem” have produced a very close interplay between particle and nuclear physics, stellar structure theory, and helioseismology. Helioseismic inversion of p-mode frequencies, coupled with theoretical models for stratification in the nuclear burning core, has placed fairly tight bounds on neutrino production, thus ruling out many novel suggestions for how that production could be reduced. Along the way, substantial improvements in the theoretical calculation of opacities and equations of state for the high-temperature, high-density plasma at intermediate depths within the solar radiative interior were devised in order to bring solar structure models and helioseismic inferences into close agreement. The recent apparent detection of other solar neutrinos, possibly arising from electron neutrinos changing flavor during flight, may offer another path for resolving this major challenge.

PLASMA PHYSICS

Our solar system, from the Sun’s core to the heliopause, is mostly composed of magnetized plasma—that is, ionized gas threaded by magnetic fields. Consequently the Sun and the heliosphere behave on microscopic scales largely as a collection of ions and electrons whose dynamical properties are shaped (if not dictated) by the magnetic field, and on macroscales as a magnetofluid. Our proximity to this vast and variable plasma laboratory has yielded enormous benefits for both astrophysics and laboratory astrophysics. The Sun and heliosphere routinely “perform” experiments in parameter regimes that cannot be enacted on Earth, providing insights into the basic physical mechanisms that govern the behavior of plasmas. For example, magnetic reconnection (see, in Section 1.3, “Understanding the Active Sun and the Heliosphere” and in Section 1.5, “Primary Recommendations (Prioritized)”)—a process equally important to space plasmas and magnetic-confinement devices—was first hypothesized as a way to explain solar flares and magnetospheric convection. In turn, our understanding of the Sun and heliosphere has been revolutionized over the past 50 years by an ongoing influx of plasma theory and numerical modeling techniques developed for and applied to fusion devices and other laboratory experiments. Laboratory experiments, such as the Princeton Plasma Physics Laboratory Magnetic Reconnection Experiment, can probe fundamental MHD and kinetic phenomena with a combined temporal and spatial resolution that exceeds what can presently be achieved in space. The funding agencies have acknowledged the benefits of this mutually beneficial relationship by supporting interdisciplinary studies through such programs as International Solar-Terrestrial Physics (NASA) and the Science and Technology Centers (NSF). The panel enthusiastically encourages the continuation and expansion of this cross-fertilization among all branches of plasma physics, which is making an essential contribution to our understanding of the physical laws governing the plasma universe.

1.8 RECOMMENDATIONS (NOT PRIORITIZED)

POLICY AND EDUCATION

Solar and heliospheric physics is a large and complex enterprise involving academic institutions, federally funded centers, observatories, and space missions, as well as commercial interests. To sustain a vigorous research community it is necessary to achieve the proper level and balance of resource allocation and investment in the above. The research community, in turn, is needed to attain long-term strategic goals and to maintain U.S. leadership in science and technology. Despite the present overall health and vitality of the solar and heliospheric physics community and the impressive advances of the past decade, there are serious concerns about the way in which solar and heliospheric physics is currently supported.

- Ground-based national research facilities for solar and heliospheric physics are funded primarily by NSF, which is divided between the AST and ATM divisions under the Mathematical and Physical Sciences and the Geosciences Directorates, respectively. On the other hand, the ground-based solar facilities operated by universities are primarily supported by NASA through its SR&T program, with secondary support from NSF. Space-based programs are funded mostly by NASA, with much smaller roles played by NOAA and DOD. Data analysis and theory efforts are mostly supported by relatively modest grants at NSF and by mission-specific programs, the SR&T program, and the Sun-Earth Connection theory program at NASA. Solar and heliospheric physics, perhaps more than any other discipline, relies extensively on multiwavelength observations from both ground-based facilities and space-based missions that are funded by multiple agencies. Interagency planning and coordination is therefore of critical importance. The excellent coordination between NSF, NASA, and other agencies, developed in recent years for joint ventures such as the National Space Weather Program, needs to be maintained in the future.
- NSF has a large investment in ground-based facilities that it builds and operates. It invites scientists to use those ground-based facilities but does not routinely sponsor the data analysis activity required to maximize the science return from its facility investment. Those analysis efforts often represent the largest share of an individual scientist's research effort. In addition, in many cases NSF does not fully fund the travel required for a scientist to use its facilities.
- NASA generally has been effective in supporting mission-specific research, although support for theory associated with missions has been uneven. To enable the systems approach of the LWS program, NASA has recognized the need to integrate theory and modeling into all phases of mission development and deployment and has created a targeted theory, modeling, and data analysis "mission" within LWS. The panel encourages similar approaches to integrate theory and modeling into future mission programs.
- Ground-based solar and heliospheric facilities routinely provide critical data that enable space-based missions to meet their science goals and/or enhance their scientific return. They also serve as important training grounds for the scientists, technicians, and instrument builders upon which NASA relies. Yet NASA's support of and investment in ground-based facilities through its SR&T program is fragile and uneven despite the fact that a number of present and future NASA programs depend on auxiliary support from ground-based facilities for mission success.

Both NASA and NSF rely heavily on university programs as training grounds for scientists, engineers, technicians, and instrument builders. The continued success of space-based missions and ground-based facilities requires that university programs thrive. In fact, however, university-based programs have faltered in recent years, at least in part owing to short funding cycles, lack of outreach to promising students, and a shift of some technical activities, for example instrument building and development, to commercial enterprises and NASA centers. Increasingly the universities find it difficult to come up with the resources to develop new instruments and concepts. The net effect has been a decreased rate of production of educated professionals having solar and space physics expertise, with the shortfall of trained experimentalists being particularly acute. It is widely recognized that the national shortfall of qualified personnel will intensify during the next 5 years. For example, in that period over half of NASA center scientists and engineers will become eligible for retirement. The scientific community and the funding agencies need to be proactive in attracting qualified students to the scientific enterprise.

As a consequence of the above concerns, the panel makes several recommendations. Some of these recommendations, although independently arrived at by the panel, closely parallel recommendations in the 2001 NRC report *U.S. Astronomy and Astrophysics: Managing an Integrated Program.*

- **The panel strongly encourages NASA, NSF, and other agencies that fund solar and heliospheric physics to continue interagency planning and coordination activities that will optimize the science return of ground- and space-based assets. It encourages a similar high level of planning and coordination between the AST and ATM divisions of NSF.**
- **The panel recommends that NSF plan for and provide comprehensive support for scientific users of its facilities. This includes support for data analysis, related theory efforts, and travel.**
- **The panel recommends that NASA support instrumentation programs, research programs, and software efforts at both national and university ground-based facilities where such programs are essential to**

the scientific aims of specific NASA missions and/or to the strategic goal of training future personnel critical to NASA's mission.

- **The panel recommends that both NSF and NASA study ways in which they could more effectively support education and training activities at national and university-based facilities. This support is particularly needed for training scientists with expertise in developing experiments and new instruments. The national laboratories have capabilities that could be better exploited by the universities. The panel recommends that both NSF and NASA study the idea of forming Centers of Excellence with strong university connections and tied to national facilities as a means of sustaining university-based research efforts and of educating and training the scientists, technicians, and instrument builders of the next generation. These centers should have lifetimes of 10 to 15 years and should be reviewed every 2 to 3 years to ascertain they remain on track.**

OTHER

The past decade has witnessed an explosion in collaborative efforts aimed at understanding the Sun-heliosphere-Earth connection. A most positive development has been the growth of broad community organizations, such as RISE, CEDAR, GEM, and, more recently, SHINE, which is concerned almost exclusively with the panel's third science priority. These organizations span what had been a deep gulf between pure scientific research and applied space weather studies. Although all of these efforts originated as grass-roots coalitions of researchers, they have flourished with support from NSF. The panel strongly recommends that NSF continue its support for these groups, in particular SHINE.

ADDITIONAL READING

A strategy for the conduct of space physics research has been set down in a number of reports by the NRC's Space Studies Board and its predecessor the Space Science Board. These reports include the following:

Space Science Board, National Research Council. 1985. *An Implementation Plan for Priorities in Solar-System Space Physics*. National Academy Press, Washington, D.C.

Space Science Board, National Research Council. 1983. *The Role of Theory in Space Science*. National Academy Press, Washington, D.C.

Space Science Board, National Research Council. 1980. *Solar-System Space Physics in the 1980's: A Research Strategy*. National Academy of Sciences, Washington, D.C.

Space Studies Board, National Research Council. 1995. *A Science Strategy for Space Physics*. National Academy of Sciences, Washington, D.C.

Space Studies Board and Board on Atmospheric Sciences and Climate, National Research Council. 1991. *Assessment of Programs in Solar and Space Physics—1991*. National Academy Press, Washington, D.C.

The research in this field is summarized in both textbooks and conference proceedings, including the following:

J.L. Kohl and S.R. Cranmer (eds.). 1999. *Coronal Holes and Solar Wind Acceleration*. Kluwer Academic Publishers, Dordrecht.

S. Habbal (ed.). 1997. *Robotic Exploration Close to the Sun: Scientific Basis*. AIP Conference Proceedings 385, Woodbury, New York.

T. Bastian, N. Gopalswamy, and K. Shibasaki (eds.). 2001. *Solar Physics with Radio Observations*. NRO Report 479.

N.U. Crooker, J.A. Joselyn, and J. Feynman (eds.). 1997. *Coronal Mass Ejections*. AGU Monograph 99. American Geophysical Union, Washington, D.C.

R.A. Mewaldt, J.R. Jokipii, M.A. Lee, E. Mobius, and T. H. Zurbuchen (eds.). 2000. *Acceleration and Transport of Energetic Particles Observed in the Heliosphere*. AIP Conference Proceedings 528. American Institute of Physics, Melville, N.Y.

A. Balogh, R.G. Marsden, and E.J. Smith (eds.). 2001. *The Heliosphere Near Solar Minimum: The Ulysses Perspective*. Springer, Chichester, U.K.

S.R. Habbal, R. Esser, J.V. Hollweg, and P.A. Isenberg. 1999. *Solar Wind Nine*. AIP Conference Proceedings 471. American Institute of Physics, Woodbury, N.Y.

Note added in proof: As a result of selections made in August 2002, SDO is now in development.

P.C.H. Martens and D.P. Cauffman (eds.). 2002. *Multi-Wavelength Observations of Coronal Structure and Dynamics*. COSPAR Colloquia Series Vol. 13. Pergamon, New York, N.Y.

O. Engvold and J.W. Harvey (eds.). 2001. *Physics of the Solar Corona and Transition Region*. Proceedings of the Monterey Workshop, August 1999. Kluwer Academic Publishers, Dordrecht.

P. Song, H.J. Singer, and G.L Siscoe (eds.). 2001. *Space Weather*. Geophysical Monograph 125. American Geophysical Union, Washington, D.C.

E.R. Priest and T. Forbes. 2000. *Magnetic Reconnection: MHD Theory and Applications*. Cambridge University Press, Cambridge, U.K.

2

Report of the Panel on Solar Wind and Magnetosphere Interactions

SUMMARY

The study of solar wind–magnetosphere interactions at the turn of the 21st century finds itself engaged in exciting exploration of exotic extraterrestrial environments and consolidating a comprehensive, fundamental understanding of the terrestrial magnetosphere. To capitalize on these discoveries, we need both classic missions of exploration to the planets and modern multispacecraft probes in the near-Earth environment. This report summarizes what we now know about planetary magnetospheres and the processes within them, what we need to know, and how we should proceed in obtaining this knowledge.

MAGNETIC FIELDS

Magnetic fields play a crucial role in governing Earth's space environment. They organize the heliospheric and magnetospheric plasmas, shield planetary bodies, such as Earth, from bombardment with charged particles, couple energy from one plasma regime to another, store that energy and later release it rapidly. Moreover, they guide the motion of charged particles to regions where they can cause visible displays such as solar flares on the photosphere or the polar lights in the atmosphere. Partners in these processes are the plasmas, energetic particles, waves, and electromagnetic emissions from radio to x-ray wavelengths in the solar wind and the planetary magnetospheres.

Solar and planetary magnetic fields organize space into normally well-separated regions. The principal plasma regimes are the corona, where the solar wind originates; the solar wind, the outward streaming plasma that carries the Sun's magnetic field to the outer heliosphere; and the magnetospheres of planetary bodies, intrinsic or induced. The magnetospheres may act as flexible shields that deflect the solar wind and thereby protect the planet and its atmosphere from most of the direct impact of the solar wind particles. However, these shields are not impenetrable.

One of the principal processes by which the shield is penetrated is called magnetic reconnection. This process is strongly controlled by the relative orientation of the magnetic fields in adjacent regions, leading to connection between the magnetosphere and the solar wind. Magnetic reconnection not only breaches the boundaries between different plasma and magnetic field regions, it is also the main process involved in the rapid release of magnetic energy in eruptions in the solar atmosphere and Earth's magnetosphere, in laboratory plasmas, and, presumably, in astrophysical settings.

Other processes can breach the magnetic shield. In the case of weakly magnetized bodies such as comets, Venus and Mars, and the moons Io and Titan, neutral particle transport across plasma boundaries occurs, with subsequent ionization. In magnetically noisy environments, particles can be scattered across the boundaries, and for small bodies finite gyroradius effects allow penetration.

An important aspect of the plasmas in most of space is that the magnetic fields that guide the motion of the charged particles are, in turn, created by the motion of those very same particles. Thus the magnetized plasma can be quite nonlinear, enhancing, deflecting, or annihilating the original magnetic field.

MAGNETOSPHERES

Planetary magnetospheres are particularly accessible settings for studying the processes occurring in magnetized plasmas, providing unique insights into basic physical processes not amenable to direct probing, processes such as particle acceleration, shock formation, and magnetic reconnection. The solar wind interaction with a magnetosphere produces thin boundaries, separating large regions of relatively uniform plasma. Within these thin boundaries microscale processes couple to the meso- and macroscale processes, affecting the stability and dynamics not only of the thin boundary layer but also of the entire coupled magnetosphere system. The magnetospheric shields of planets and moons vary considerably. Some weakly magnetized planetary bodies like Earth's moon routinely lose their atmosphere to the solar wind, while others such as Venus and Mars have had their atmospheres significantly altered, as indicated by the isotopic ratios of their atmospheric constituents, but not completely removed. Magnetospheres also exhibit rapid reconfigurations, such as the ejection of magnetic islands, or plasmoids, while the inner region collapses, as seen routinely in the tail regions of Earth and Jupiter. Overall planetary magnetospheres are complex, coupled systems, connected on one end to a supersonic flowing magnetized plasma, the solar wind, on the other end to a cold dense planetary atmosphere and ionosphere, and sometimes to embedded plasma sources such as satellites and rings.

While each planetary magnetosphere presents great intellectual challenges and its behavior provides insight into diverse astrophysical solar and laboratory systems, the terrestrial magnetosphere is of particular practical interest. It provides a home to many technological systems that are increasingly sensitive to magnetospheric disturbances. Such disturbances affect the quality of communications, our ability to navigate, the capacity of power transmission lines, the orbits of low-altitude satellites, and the operation of geosynchronous spacecraft carrying TV broadcasts, relaying phone calls, and monitoring our weather. Both astronauts and flight crews on polar air routes can receive undesirable levels of radiation from energetic particles controlled by the magnetosphere. Thus, understanding and predicting the response of the magnetosphere to varying interplanetary conditions, i.e., space weather, has become a particular concern.

THE TERRESTRIAL MAGNETOSPHERE

The study of Earth's magnetosphere began with ground-based measurements of the time variations of the magnetospheric magnetic field. These observations revealed not only the existence of the magnetosphere but also its variable state of energization. The International Geophysical Year initiated an era of discovery in which single-spacecraft missions throughout the magnetosphere provided an overview of the characteristic regions, boundaries, and plasma conditions, with some evidence of the processes therein, but they did not elucidate how the processes in the magnetosphere work. Therefore, current and future exploration of the terrestrial magnetosphere concentrates on the use of multispacecraft missions complemented by ground-based arrays of magnetic, radar, and optical sensors to characterize plasma behavior in a dynamic environment and to probe cause and effect in a complex system at various scales. At the other planets, with few exceptions, we remain in the discovery phase since thus far we have generally been restricted to single-spacecraft missions, often single flybys, not orbiters.

There are many success stories in magnetospheric exploration as well as continuing puzzles. The standing bow shock is well understood, but it is only the fastest of three waves that should stand in the solar wind flow. The other two waves—the intermediate mode, which rotates field and flow, and the slow mode, which "stretches" field lines—could also lead to standing structure. Reconnection is now known to provide a time-varying interconnection of the terrestrial magnetosphere with the magnetized solar wind, driving the circulation in the magnetosphere, but in a manner that is as yet not well understood. Reconnection is recognized to be the principal mechanism for the violent release of stored magnetic energy and for magnetic flux return from the tails of the magnetospheres of both Jupiter and Earth. Nevertheless there is not agreement on what triggers the rapid onset of magnetotail reconnection.

Radial diffusion and pitch angle scattering of energetic particles apparently produce many of the observed features in the radiation belts of planetary magnetospheres, but the driver of the radial diffusion remains elusive, and the sources and acceleration mechanisms for the involved energetic particles are not always clear. At unmagnetized planets the mechanism for the formation of induced magnetospheres is relatively well understood but the atmospheric loss is poorly understood.

INTRINSIC AND INDUCED

Magnetospheres can be divided into two types: induced, if any intrinsic magnetic field of the body is so weak that the ionosphere is directly exposed to the flowing solar wind plasma, and intrinsic, if the body has an internal magnetic field sufficiently strong to deflect the plasma that flows against it. Induced magnetospheres form around highly electrically conducting obstacles if the conductor, generally an ionosphere, can stave off the solar wind flow. Induced magnetospheres also form in strong mass-loading environments such as at a rapidly outgassing cometary nucleus. Comets, Venus, Mars, and some of the moons of the gas giants have magnetospheres induced by the rotating magnetospheric plasma. Mercury, Earth, Ganymede, and the gas giants have intrinsic magnetospheres. Circulation inside the intrinsic magnetospheres can be driven by the externally flowing plasma or by an internal source such as plasma derived from the volcanic gases of Io, accelerated by the rapidly rotating jovian magnetosphere. The centrifugal force of this plasma drives a massive circulation pattern in the jovian magnetosphere, powering a massive magnetospheric "engine." Thus Jupiter acts as a bridge in our understanding of the terrestrial and astrophysical magnetospheres.

For both intrinsic and induced magnetospheres the supersonically flowing solar wind is deflected by the magnetosphere, forming a bow shock. Behind the bow shock, the decelerated shocked plasma flows around the obstacle in a region known as the magnetosheath. In intrinsic magnetospheres, the boundary between the flowing plasma of the solar wind and the plasma, connected by the magnetic field to the planet, is called the magnetopause. In an induced magnetosphere, the analogous boundary is called an ionopause. Often the magnetopause and the ionopause are thin layers of current. Behind the magnetosphere proper the magnetic field and plasma are stretched by the solar wind flow, forming a long magnetotail. Inside the magnetosphere, differing regions of plasma can be found, such as the plasmasphere in Earth's magnetosphere and the Io torus in Jupiter's. In an induced magnetosphere, the plasma is generally relatively cold and affected by the external flow in ways much different than in an intrinsic magnetosphere. For these induced magnetospheres, the solar wind interaction acts to scavenge the atmosphere and may be responsible for the loss of water from the atmospheres of Venus and Mars and for alteration of isotopic ratios over the eons since the formation of the solar system.

THE PRESENT PROGRAM

The present program of studies of magnetized space plasmas is robust. There is a vigorous program of ground-based measurements, theory, modeling, and data analysis, supported jointly by NASA, NSF, and, to a lesser extent, by other agencies. Data are being returned from the solar wind, magnetotail, magnetosphere, and low Earth orbit. Galileo has recently completed its exploration of the jovian magnetosphere and Cassini is on its way to Saturn. The data are being analyzed promptly, and significant scientific discoveries are being made. Several important projects are under development and moving toward their launch opportunities. Nevertheless, there is still much to do.

UNIFYING THEMES

The outstanding questions that need to be addressed in planetary magnetospheres can be divided into three themes: the creation and annihilation of magnetic fields; magnetospheres as shields and accelerators; and magnetospheres as complex, coupled systems. The first theme includes the formation of the major magnetospheric current systems: the magnetopause, the tail current, the ring current, and the field-aligned currents. This theme also includes the disruption of some of these currents and reconnection of the magnetic field across current layers, at the magnetopause, in the magnetotail, and in planetary magnetodisks.

Under the second theme is the role that induced and intrinsic magnetic fields play in deflecting the solar wind and the energetic particle populations coming from the Sun. These magnetospheres also store energy for later release, leading to sudden energization of the plasma in the magnetosphere and acceleration of magnetospheric energetic particles. In the inner magnetosphere, trapped charged particles are also accelerated slowly to high energies by stochastic processes. None of these processes is well understood. Even less well understood are the interactions of flowing magnetized plasma with the remanent fields of bodies like Mars.

The third theme encompasses some of the most difficult areas of magnetospheric research: the interactions among the disparate plasma regimes within a magnetosphere. The bow shock interacts with the incoming solar wind upstream and the magnetosheath and magnetopause downstream. Reconnection changes the topology of magnetic field lines, connecting interplanetary and terrestrial magnetic field lines so that the plasmas from the two regimes mix, and allowing momentum and energy to flow into the magnetosphere from the solar wind. The ionosphere interacts with the polar magnetosphere and the magnetospheric regions at lower latitudes. Planetary magnetospheres have their own unique twists on these processes. In the jovian magnetosphere the ionosphere enforces co-rotation of the plasma over enormous scales and a giant circulation pattern is set up within the magnetosphere. At the unmagnetized planets there is direct coupling of the solar wind with the neutral atmosphere.

RECOMMENDATIONS

The discipline of space physics and the subdiscipline of solar wind–magnetosphere interactions have experienced an explosion of knowledge and understanding in recent years. Still there are some very basic processes that we do not understand, especially at a predictive level. If we cannot predict the rate of reconnection at our own magnetopause or in the magnetotail (and today we cannot), we have little hope of extending our knowledge to planetary and astrophysical systems. Thus we recommend that the future exploration of the terrestrial and extraterrestrial magnetospheres should be directed toward the deeper understanding of the fundamental physical processes and the global coupled systems, supported and guided by theoretical investigations and simulation efforts. This requires multisatellite missions and the optimal use of simultaneous, coordinated, and overlapping spacecraft missions. The global coupled system extends all the way down to the upper atmosphere and ionosphere. Thus in the terrestrial magnetosphere ground-based facilities play an important part in the exploration of the coupled system.

In planning for the next decade of studies of solar wind–magnetosphere interactions we have been guided by four essentials. We must understand the physical processes involved and therefore need measurements with high resolution, capable of studying three-dimensional structure with support from theory and modeling. Our models must be predictive from knowledge of external conditions. This requires global, multipoint observations and is best achieved with deep, theoretical insight rather than empirical models. We must investigate how regions couple, not simply how they work in isolation, and we must continue to explore new settings to develop greater understanding.

Critical scientific objectives in the future exploration of solar system magnetospheres include the following:

- *A deeper physical understanding of fundamental plasma processes, such as particle acceleration, magnetic reconnection, and the role of turbulence*. Achievement of this objective should be at the core of present and future space exploration, and the panel endorses the planned Magnetospheric Multiscale mission.
- *Understanding the scale sizes of the solar wind structures that power Earth's magnetosphere*. Achieving this objective, which is needed for predictive purposes, will require multispacecraft missions near 1 astronomical unit (AU) with spacecraft separations measured in tenths of astronomical units.
- *Understanding the dynamics of the coupled magnetospheric system and of space weather*. Achievement of this objective requires arrays of instruments in space as well as on the ground (just as readings from ground weather stations are complemented by readings from space). A magnetospheric constellation of up to 100 spacecraft to monitor a significant volume of the magnetosphere is strongly recommended, along with complementary ground-based measurements.
- *Understanding the complex interaction between the solar wind and the polar ionosphere*. Achievement of this objective requires the establishment at high latitudes of the long-awaited Advanced Modular Incoherent Scatter Radar (formerly known as the Relocatable Atmospheric Observatory). This facility could be enhanced by many possible space missions, such as a stereo imager or a polesitter auroral imager.
- *Measurement of the density of the invisible populations within the magnetosphere*. To achieve this objective, the panel recommends the establishment of magnetometer arrays that can perform magnetoseismology, in analogy to terrestrial and solar seismology, recording transient waves and the ringing of the magnetosphere.
- *Understanding the energization of the radiation belts*. This long-sought objective requires knowledge of the radial swaths of the particle and field environment simultaneously at different local times and under different geomagnetic conditions to learn how and why particle populations intensify and decay.
- *Understanding the complex interactions of the solar wind and planetary magnetospheres and atmospheres*. To achieve this objective, particles and fields instruments will need to be flown on both Discovery-class and major space missions.
- *Understanding planetary magnetospheres*. The exploration of planetary magnetospheres is in its infancy, yet comparisons between these magnetospheres and the terrestrial magnetosphere and with each other are critical to fully understanding the processes taking place. Missions to study atmospheric loss from Venus and Mars, the occurrence of lightning at Venus and Jupiter, the dynamics of Mercury's magnetosphere, and the joint control of the jovian aurora by Io and the solar wind are some of the many missions that could contribute to our understanding of planetary magnetospheres.

TECHNOLOGY DEVELOPMENT

While some of these objectives are already technically within our grasp, additional technology development is needed for others. For example, several missions could be undertaken most effectively with a solar sail. Improved ion propulsion, nuclear-powered propulsion, and mid-size expendable launchers would also increase access to space. Smaller spacecraft systems and instruments would enable the constellation missions that are planned and would allow greater return from resource-limited planetary missions. Finally, attention needs to be given to the entire data chain, from operations to data transmission to their assimilation in models to reduce manpower and the total expense of the data chain.

CHANGES IN POLICY

Many of our programs would be enabled and enhanced with some simple changes in policy. In some cases, this simply requires better coordination between or even within agencies. Sometimes data are obtained but funds are required for data access or archiving. We need to have processes to determine when a technique has moved from the research arena into the space-forecasting arena. We need to coordinate opportunities for access to space so that all such opportunities are utilized, and we need to ensure that funding for space experiments is available when possible flight opportunities arise. Presently, missions of opportunity are solicited far too seldom and on time scales incongruent with the duration of the opportunity. We also have to guard against using the space science budget to cover shortfalls in other programs such as the space station. Budget raids can devastate smaller programs. Moreover, we need to find ways to reduce regulatory burdens, such as International Traffic in Arms Regulations (ITAR) and information technology security regulations, which have led to more and more obstacles to international collaboration and to university participation. These policies often have results much different than originally intended. High-level communication and coordination between regulatory agencies and NASA are needed to achieve reasonable implementation standards and procedures.

SYNOPSIS

In short, the research enterprise in solar wind–magnetosphere interactions is strong, and much has been accomplished. Nevertheless, some very fundamental understanding is still needed to reach the quantitative level of a fully predictive science. Fortunately, the means to attain this understanding now exist. In some cases an investment in technology will bring us to the threshold of the needed breakthroughs. The next decade of this discipline, launched with the momentum of the last decade's discoveries, fueled by an exciting series of new observations, and supported by a strong program of theory and modeling, promises to usher in a new, quantitative level of understanding of the Sun-Earth connection.

In the next section, the panel provides an overview of the workings of planetary magnetospheres. This overview is followed in Section 2.2 by a detailed discussion of current understanding of the processes in the terrestrial magnetosphere and the environments of the planetary magnetospheres. This description is needed to understand why the panel has chosen the paths it recommends, but it may be skipped by those seeking only to learn the recommendations. Section 2.3 is an attempt to provide three unifying themes that order the remaining tasks. Section 2.4 summarizes the existing program and presents recommendations. Sections 2.5, 2.6, and 2.7 describe, respectively, the recommended future program, the recommended technology developments, and the recommended policy changes that will enable the progress needed in this field.

2.1 SOLAR WIND–MAGNETOSPHERE INTERACTIONS: THE REALM OF MAGNETIZED PLASMAS

INTRODUCTION

Our increasingly technological society relies more and more on assets launched into space. In addition, our investigations extend well past the local space plasmas into those of the solar and astrophysical systems. Understanding the behavior of magnetized plasmas has become increasingly important. We must understand the environment in which our satellites operate. We need to predict how the solar wind affects the terrestrial magnetosphere. We require insight into how the Sun generates explosive events, and we desire to comprehend the workings of distant astrophysical systems that are clearly also affected by magnetic processes.

The solar wind's interaction with the terrestrial and planetary magnetospheres allows us to treat a problem of much practical importance and learn how these plasmas work in a most general manner. We can then extend this knowledge to other plasma systems in regions we cannot probe directly.

For users of this report who are not familiar with the physics of space plasmas, this section offers some brief insight into the basic plasma processes that occur in space. This section also provides a preview of the themes introduced in Section 2.3 and the rationale for the recommendations made in subsequent sections.

THE FOURTH STATE OF MATTER

Plasmas are often referred to as the fourth state of matter. The behavior of this state, especially of magnetized plasmas, can be nonintuitive. We are most familiar with the other three states—solid, liquid, and gas—whose dynamical properties are governed by the differing intermolecular forces of each state. Our intuition usually serves us well here. In an ideal gas, the forces between the molecules are transmitted through collisions. The random motions of the gas are characterized by a temperature and, in collisional equilibrium, all constituents come to the same temperature. The pressure in the gas is proportional to the product of the density and the temperature. A pressure gradient exerts a force. For example, in Earth's atmosphere we know that the pressure decreases with altitude. The force associated with this pressure gradient acts on a parcel of air to support it against the force of gravity.

In space plasmas there often are no collisions in the usual sense. Thus, different components of the plasma can have different temperatures. Further, temperatures along the magnetic field and across it can differ. Still, and counterintuitively, a plasma can exert pressure forces not only through the thermal motion of its particles but also through its magnetic (and electric) fields. These fields do have pressure (proportional to the square of the field strength) and, as in the case of a gas, the gradient of that pressure exerts a force. In a magnetized plasma, the magnetic field orders the charged particle motion, the energy of the gyrating particles provides the plasma pressure perpendicular to the field, and the parallel thermal motions provide pressure along the field.

An example of the interplay between these pressures is provided by the boundary between the solar wind flow and the magnetosphere. This region, the magnetopause, is often treated as a boundary between a plasma with at most a weak magnetic field (the shocked solar wind) and a strong magnetic field (Earth's magnetosphere) containing very little plasma. The pressure on the solar wind side is contained in the thermal motions of the plasma. At the boundary of the plasma the thermal-pressure gradient exerts a force toward the magnetosphere. The pressure in the magnetic field similarly exerts a force into the solar wind plasma, where the magnetic pressure decreases. Thus there is force balance, and the magnetosphere and the solar wind establish a pressure equilibrium in the absence of collisions.

The ratio of the proton mass to the electron mass is 1,836. Thus an electron at the same temperature as a proton moves at 43 times the speed of the proton, and for this reason electrons can communicate rapidly in a plasma. Protons, though, have all the inertia and momentum, and electrons tend to follow the dynamics of the protons, setting up small ambipolar electric fields to maintain quasi-neutrality in the plasma. As a result, except on the microscale, the electric field seldom builds up to such a degree that its pressure is important. However, when electric fields do arise parallel to the magnetic field they can be very important to the processes in the plasma, so much so that it is critical to be able to observe such generally small electric fields. The panel notes that these electric fields are frame independent, while the electric field in the direction perpendicular to the magnetic field is frame dependent, so that the perpendicular electric field detected depends on the velocity of the observer relative to the magnetic plasma. Thus the flowing solar wind has an electric field as seen in Earth's reference frame. In short, while a plasma has many similarities to fluids and gases, it is different enough that our physical intuition is often ill-prepared to understand processes that occur therein.

RECONNECTION

The very large mass ratio between the proton and electron affects their gyromotion as well as their typical speeds. Because of the nature of the Lorentz force, which keeps a charged particle in orbit about a magnetic field, the radii of gyration of different charged particles are proportional to their mass times their velocity and inversely proportional to their charge. If protons and electrons have the same energy perpendicular to the magnetic field, the radius of the gyrating electron is 2.3 percent of that of the proton. In a collisionless plasma, the gyrating particles define the magnetic field lines. Particles in orbit about a magnetic field line stay with that magnetic field line. The ability of a charged particle to orbit a magnetic field line depends on the scale size for changes in the field. A small change can cause drift motion, and too large a change in the field on the scale of a gyroradius can cause a charged particle to become unmagnetized and move to orbit another field line. Owing to their smaller gyroradii, electrons can follow small-scale field variations to sizes roughly 43 times smaller than protons. One might think that this is moot for a system as large as Earth's magnetosphere, because its scale sizes are vast compared with those of the gyroradii. In fact, it is standard practice to average over the gyromotion and treat the magnetized plasma as a magnetic fluid. This formulation is known as magnetohydrodynamics (MHD). However, the vast scale of the magnetosphere does not allow us to completely ignore the kinetic motion of its particles. It just reduces the region in which that kinetic motion is crucial to small areas called neutral points. Close to these points the protons first become unmagnetized, and then closer yet the electrons become unmagnetized.

This process in which charged particles lose their ability to define a magnetic field line is called reconnection. If they are antiparallel, two neighboring magnetic field lines, say one that starts and ends on Earth and another that starts and ends on the Sun, can become connected so that two new field lines are created, both of which have one end on Earth and one end on the Sun. This topological change enables the plasmas in the two regions (terrestrial and solar in this case) to mix. It also allows momentum and energy to be supplied from one plasma to the other. Figure 2.1 illustrates the geometry of this situation. The magnetic field at the reconnection point forms an x-type configuration with plasma flowing into it from the left (Earth's magnetospheric plasma) and from the right (Sun's solar wind plasma). Field lines switch partners at the x-point, and plasma and cojoined field lines flow rapidly outward (top and bottom). Since an electric field is frame dependent, these moving magnetized plasmas have electric fields in the frame of the reconnection point, as sketched in Figure 2.1.

Collisions, either particle-particle or wave-particle, can also demagnetize orbiting charged particles, and in numerical simulations numerical dissipation can mimic the reconnection process. Thus it is not always clear how the magnetosphere undergoes this most critical process. Much continued in situ study with high temporal and spatial resolution is required, as well as investigation with state-of-the-art numerical codes.

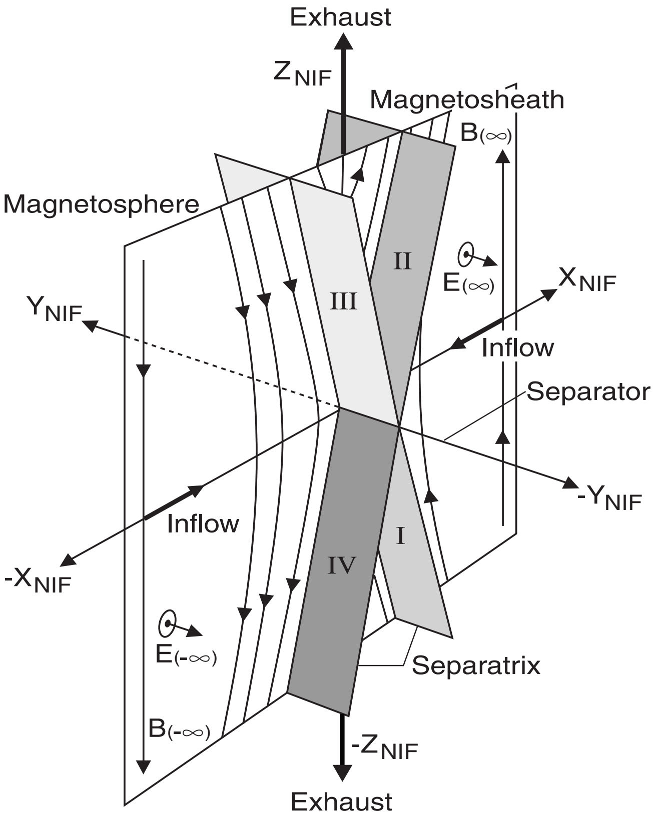


FIGURE 2.1 The geometry of reconnecting antiparallel magnetic fields at a neutral point. Courtesy of J.D. Scudder.

From the above it is obvious that magnetic reconnection is the crucial process enabling plasma (and momentum, energy, and magnetic flux) to cross magnetospheric boundaries. In addition, the partner-swapping process in reconnection can dramatically alter the stress balance in a plasma, leading to catastrophic energy release processes, such as solar flares and magnetospheric substorms, discussed below.

FLOWING MAGNETIZED PLASMAS

A solid can support both compressional and transverse oscillations but a normal liquid and a gas cannot. Thus the dynamics of the flow around an object in a flowing gas is dominated by compressions and rarefactions. However, in a magnetized plasma that otherwise resembles a gas, there are transverse oscillations as well as two compressional waves. These three waves are usually called fast, intermediate, and slow. They are all necessary to transmit an arbitrarily shaped perturbation through a magnetized plasma. For example, in the interaction of the flowing solar wind with Earth's magnetosphere, the fast mode slows, heats, and deflects the flow and magnetic field, but in general the intermediate mode is needed for additional field and flow deflection, and the slow mode is needed to prevent a density pileup at the subsolar point. Just as in a solid or gas, perturbations travel at a finite velocity, and in the magnetized plasma as in many other situations the velocity of each wave mode is different.

When it arrives at each of the planets the solar wind flow is supersonic, moving faster than the speed of the compressional (fast mode) wave that could deflect it around the planetary obstacle. The momentum flux of the solar wind represents a dynamic pressure that confines the planetary magnetic field, but in order for it to be applied to the magnetosphere, the flow must pass through a bow shock that slows, deflects, and heats the flow, making it subsonic. Then the three wave modes (fast, intermediate, and slow) can act on the plasma to cause the deflection of the flow and alter the plasma conditions at the boundary of the magnetosphere.

It is very important to magnetospheric processes that there are finite propagation times and finite transport times in the magnetosphere. When the solar wind magnetic field reconnects with the subsolar magnetospheric magnetic field, it begins to transport magnetic flux to the geomagnetic tail, as illustrated in Figure 2.2. The tail may increase in size for about an hour. Then, when

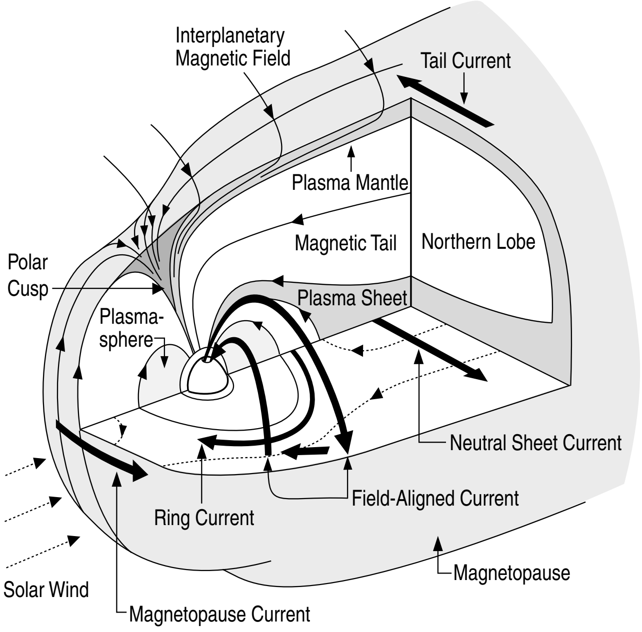


FIGURE 2.2 A cutaway diagram of Earth's magnetosphere. Courtesy of C.T. Russell, University of California, Los Angeles.

reconnection begins in the tail, it may require another hour to transport the magnetic flux out of the tail, but the signals denoting the onset of each reconnection event can travel through the magnetosphere within minutes. Other manifestations of the finite travel time of perturbations in the plasma include the resonant behavior of individual dipole flux tubes, with waves bouncing back and forth in the magnetic tube and the propagation of shock-initiated disturbances through the magnetosphere. These signals can be used to probe the magnetosphere much as seismology uses waves triggered by earthquakes to probe the structure of Earth.

The waves discussed above have very long wavelengths, generally a large fraction of the dimension of the system. Waves at shorter wavelengths are also present in magnetized plasmas. Often these waves are responsible for releasing free energy from the plasma, ultimately into heating of the system. Examples of such waves include those caused by the upstream ions reflected at the bow shock and ion cyclotron waves produced both in the solar wind interaction with comets and in the flow of the Io torus past the mass-loading region at Io.

When a flowing magnetized plasma interacts with an unmagnetized planet there are important similarities and differences from the magnetized case. First, if the unmagnetized planet has an atmosphere, then it will be ionized by the solar UV and EUV radiation. The solar wind magnetic field will be draped across this electrically conducting ionosphere and will pile up in front of it, as illustrated in Figure 2.3. This pileup region deflects the solar wind particles around the ionospheric obstacle while the magnetic field begins to diffuse into the ionosphere. However, the solar wind magnetic field is quite variable in direction, and the long-term (days) vector average field is close to zero. The diffusion time into the interior of the ionosphere, high in the collisionless exosphere, is long. Thus, the deep ionosphere does not generally become strongly magnetized by this external magnetic field. Second, the neutral atmosphere often has a great enough extent that the neutral density is significant on solar wind stream tubes that are flowing rapidly. When the neutral atoms and molecules of the atmosphere become ions they are accelerated by the solar wind and lost to the planet. This loss can lead to significant changes in a planetary atmosphere over the age of the solar system. Hence the magnetic field is both an accelerator and a shield at unmagnetized planets.

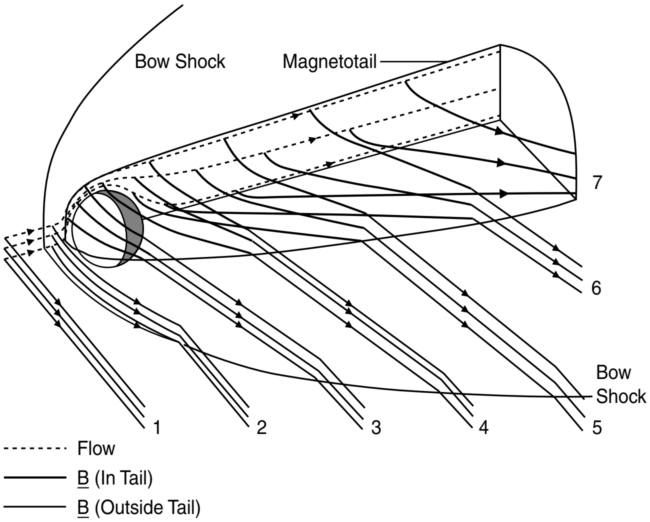


FIGURE 2.3 The solar wind interaction with an unmagnetized planet, illustrating the effect of the shock on the field and the stretching of the field to form an induced magnetotail. Courtesy of C.T. Russell, University of California, Los Angeles.

THE STORAGE AND RELEASE OF ENERGY

The interaction of the solar wind with a planet creates a magnetic cavity with a long magnetic tail. Since a magnetic field has an energy density, the formation of a planetary magnetotail or its expansion involves the storage of energy. The additional energy in a magnetotail can be provided externally when the magnetosphere reconnects with the interplanetary magnetic field at the dayside if the fields are in opposite directions. The stored energy is extracted from the mechanical energy of the solar wind flow as the magnetic field lines joining the magnetosphere to the solar wind slow its flow.

The two nearly antiparallel magnetic lobes of the tail can also reconnect in a manner similar to the fields at the magnetopause, and energy can be released rapidly from the tail lobes, in a process called a substorm. Some of this energy accelerates the bulk of the plasma, some heats the plasma, and some energizes a few particles to very high energies. On Earth these high-energy particles help populate the radiation belt. Over the auroral zones beams of particles are created that cause the auroral emissions during the reconfiguration of the night magnetosphere associated with these acceleration processes. On the Sun, solar flares are seen when highly energetic particles strike the solar surface after a reconnection event in the magnetic field above the photosphere.

The creation of a small number of energetic particles (a “high-energy tail,” as these energetic particles are often called) is another nonintuitive phenomenon in a collisionless plasma. In a collisional environment the number of particles as a function of energy follows a Maxwellian distribution in which there are very few particles that deviate much from the bulk of the distribution. However, in a collisionless plasma, the high-energy particles can receive a disproportionate amount of the energy being put into the plasma. Understanding the conditions under which this occurs is important, because these high-energy particles can be deleterious to a spacecraft system in orbit in the magnetosphere. In the broader astrophysical setting, we are interested in understanding how cosmic rays reach energies much, much higher than those of particles accelerated by processes in the solar system.

COUPLING IN A COLLISIONLESS PLASMA

The large proton-to-electron mass ratio not only affects the relative speeds of the two particles under normal circumstances and also their relative gyroradii, but also the charge separation. In general, electrons will stay close to the ions to maintain charge neutrality. Although in the absence of collisions charged particles will stay on a single magnetic field line, any charge imbalances that arise can generally be removed by motion along the magnetic field. Thus most plasmas are quasi-charge-neutral.

The magnetic field in most space plasmas is strong enough to divide space into different plasma regimes with little communication across the boundaries between them. When the magnetic fields in two adjacent regions are in nearly opposite directions, reconnection, discussed above, can occur, linking the two regions. If one of these regions is flowing past the other, this linkage can transfer momentum from the flowing plasma to the initially stationary plasma. This is the way in which Earth’s magnetosphere is stirred by the solar wind that flows past it. Surface waves can also transfer momentum across such a boundary if the system is dissipative. The waves may be generated at the bow shock and blown back against the magnetopause, or they may arise in the interaction due to a velocity-shear instability, such as the Kelvin-Helmholtz instability, that is akin to the process by which the wind creates ocean waves.

An important coupling occurs between the plasma and the neutral gas in Earth’s magnetosphere at the foot of the field line. The stress that is applied at the interface between the solar wind and the magnetosphere must eventually be taken up by Earth, and this occurs ultimately through collisional transfer between the ions and the neutrals. To get the ions moving at the feet of magnetic field lines and overcome the drag of the neutral atmosphere, the magnetosphere sets up a large current system that connects the outer magnetosphere with the ionosphere along magnetic field lines. The closure of the current across field lines in the collisional, electrically conducting ionosphere accelerates the low-altitude ions via the $\mathbf{J} \times \mathbf{B}$ force or ponderomotive force. This force is the macroscale manifestation of the Lorentz force, which maintains charged particles in their orbits around magnetic field lines. It can also transfer stress from the ionosphere to the magnetosphere such as to enforce co-rotation of the cold magnetospheric ions. This mechanism is especially important in the jovian magnetosphere. In the terrestrial auroral ionosphere, the chain of momentum transfer is completed when the moving ionospheric ions transfer their momentum to the neutral gas, generating high-altitude winds. The overall coupling from the solar wind down to the ionosphere is a very complex process, and it is fair to say that while we now understand this process much better than even a decade ago, it is not yet understood as well as we need to understand it.

IMPACT AND RELEVANCE

Planetary magnetospheres are important first and foremost because we live inside one. Earth is currently the only habitable planet, and that habitability may in part be due to the shield provided by Earth's magnetosphere. That shield also protects our spacecraft and communication systems. At times, however, the shield is overwhelmed and it is not as effective. It is becoming more and more important to be able to predict when these times of stress occur.

Now the question of habitability extends beyond Earth. Other planets are magnetized and may once have had Earth-like fields even if they do not today. Could an early magnetic field on Mars, say, have protected the atmosphere and climate of Mars from the solar wind and solar energetic particles so that life began, even if for only a short while?

Even if life were not an issue, many aspects of planetary behavior are affected by the properties of the induced and intrinsic magnetic fields. Atmosphere is being lost from Mars and Venus and in the past, from the Moon and Mercury. The moons of Jupiter interact strongly with the jovian magnetosphere, and one moon, Io, is responsible for turning that planet into a powerful radio source. Many of these processes may have counterparts on an astrophysical scale. Radio sources, magnetic flux tubes, x-ray production, shock waves, and cosmic ray acceleration may all have analogues in the more accessible terrestrial magnetosphere and provide some ground truth for astrophysical processes.

Paradoxically, the accessibility of space plasma is also important for laboratory plasmas. Laboratory devices often cannot be probed because of their size or operating conditions. When an analogous process occurs in space, an instrument can be placed in the space plasma without disrupting the system, whereas in the laboratory an instrument might interfere with the process, if it is able to operate at all. Most important, space plasmas provide ground truth for computer simulations. Codes can be validated on space plasma processes and then used in situations where there is no ground truth because of the dangerous conditions involved in testing the system.

SUMMARY

Magnetized plasmas are an exotic state of matter in which processes occur that are outside our usual experience. We need a combined program of observation, theory, and modeling to understand them, and we need to perform these studies under a variety of conditions. The processes that occur when a magnetized plasma flows past an obstacle are particularly important and challenging. Three important aspects of such plasmas are that the magnetic fields can act as both shields and accelerators, that they can generate and annihilate magnetic fields, storing and releasing energy in the process, and that the coupling between the various plasma regimes occurring in planetary magnetospheres is complex. Understanding the physics of these magnetospheres is important to planetary scientists, to astrophysicists, to laboratory physicists, and to the inhabitants of this planet.

2.2 MAGNETOSPHERES AND THEIR PARTS

OVERVIEW

Magnetospheres consume huge amounts of power and extend over vast volumes of space. They are powered by both the solar wind and internal energy sources. Jupiter, for example, has a very strong internal energy source, extracting energy from the rotation of the planet. At the other extreme, the dynamics of Earth's magnetospheric plasma is largely driven by the solar wind. The size of a magnetosphere is determined by the balance between the pressure exerted by the magnetosphere outward and the pressure of the solar wind confining the magnetosphere. The vast extent of the jovian magnetosphere is due to the strong magnetic dipole moment of the planet and the strong outward centrifugal force in its dense, nearly co-rotating magnetodisk, balanced by a relatively weak solar wind. In contrast, Earth's magnetosphere is smaller because its magnetic moment is much weaker than that of Jupiter, the solar wind pressure is stronger, and there is very little internal plasma pressure. Venus has essentially no internal magnetic field, but it does possess a thick atmosphere and a significant ionosphere. The solar wind induces a "magnetosphere" that is not much larger than the diameter of the planet but otherwise similar in shape to Earth's magnetosphere.

The solar wind is magnetized and mainly flows around planets rather than being absorbed by them, whether they have intrinsic or induced magnetospheres, and the interaction between the solar wind and planetary magnetospheres is quite complex. While there is a direct correlation between the size of an intrinsic magnetosphere and solar wind dynamic pressure, the momentum transfer between a magnetosphere and the solar wind depends also on the strength and relative orientation of the planetary and solar wind magnetic fields, the latter generally called the interplanetary magnetic field (IMF). Understanding the coupling of the variable solar wind to a planetary magnetosphere requires fundamental knowledge of the properties of magnetized plasmas.

The more than 40 years of research on planetary magnetospheres has yielded a wealth of information about the general properties of the solar wind–magnetospheric interaction. The characteristics of Earth's magnetosphere serve as well-documented examples of the complex interaction between the solar wind and the planet's magnetic field. As illustrated in Figure 2.4, detached from the obstacle in the solar wind is a collisionless bow shock that acts to slow the solar wind to subsonic speeds and partially deflect it around the magnetosphere. Between the bow shock and the magnetopause is the magnetosheath, where plasma undergoes further slowing and deflection around the magnetosphere. The magnetopause is the location where the inward pressure of the solar wind is balanced by the outward pressure of the magnetosphere. On the dayside, the magnetosheath magnetic field is enhanced at the magnetopause and the density depleted. The degree to which the field is enhanced and the density depleted depends on the external solar wind conditions and the magnetic field direction.

On the nightside, as illustrated in Figure 2.1 above, the solar wind flow past Earth stretches Earth's field lines into a long tail. The dynamics of this tail is very strongly correlated with the solar wind magnetic field orientation and the solar wind pressure. Inside the magnetopause, other plasma regions are formed as a result of the solar wind interaction. In general, these plasma regions also have distinct properties and are separated from each other by thin boundaries. This cellular structure is a distinct characteristic of the interaction of collisionless magnetized plasmas.

At high latitudes are the polar cap, polar cusp, plasma mantle, and lobe regions. These are generally regions of open magnetic field lines (i.e., one end of the field terminates in Earth's ionosphere and the other terminates on the Sun) created by magnetic reconnection or by interconnection of solar wind and magnetospheric magnetic fields. A relatively dense plasma sheet is formed around the neutral sheet in Earth's magnetotail.

Closer to Earth, plasmas are energized, creating the ring current and radiation belts (or Van Allen belts, as they were originally named). Finally "cold" plasma from Earth's ionosphere (~1000 K plasma relative to a ring current temperature of ~10^8 K) populates a torus near Earth, called the plasmasphere. Similar structures are seen in other planetary magnetospheres such as those of Mercury and Jupiter, but they are modified there by the presence of other sources of plasma (such as satellites and rings) and by the strength of these plasma sources and internal magnetic field relative to the surrounding solar wind conditions.

Each of these regions is discussed in detail in the subsections below. Significant achievements are discussed first, followed by outstanding questions. In general, the similarities and differences between Earth's magnetosphere and other planetary magnetospheres are treated in each particular subsection. However, some properties unique to the interaction of the solar wind with other planetary magnetospheres and small bodies in the solar system are discussed separately.

BOW SHOCK

Achievements

The investigation of Earth's bow shock spans much of the history of space physics in the spaceflight era. Initially it was thought that collisionless shock waves could not exist, because the mean free path of a solar wind particle was larger than 1 AU. Later, it was realized that waves in plasmas could provide the dissipation that would have been provided by particle-particle collisions in a collisional shock. However, waves were not the complete answer. It soon became apparent that the heating across a shock (dictated by the magnetohydrodynamic jump conditions across the discontinuity) could not be provided by wave dissipation above a certain critical Mach number. Above this Mach number, a fraction of the incident solar wind ion beam is reflected from the shock by a combination of the increased magnetic field and the shock electric field that is along the normal in the direction to slow the ions. The reflected ions return to the solar wind. The ultimate fate of these ions depends strongly on the angle between the solar wind magnetic field and the normal to the (curved) shock surface (θ_{bn}). Later it was appreciated that these processes are present even for subcritical shocks.

For $\theta_{bn} < 45$ degrees, particles that reflect from the quasi-parallel shock or that "leak" from the downstream region do not return to the shock (see Box 2.1 on dissipation). The backstream ion (and electron) beams found in the foreshock (see Figure 2.4) interact with the solar

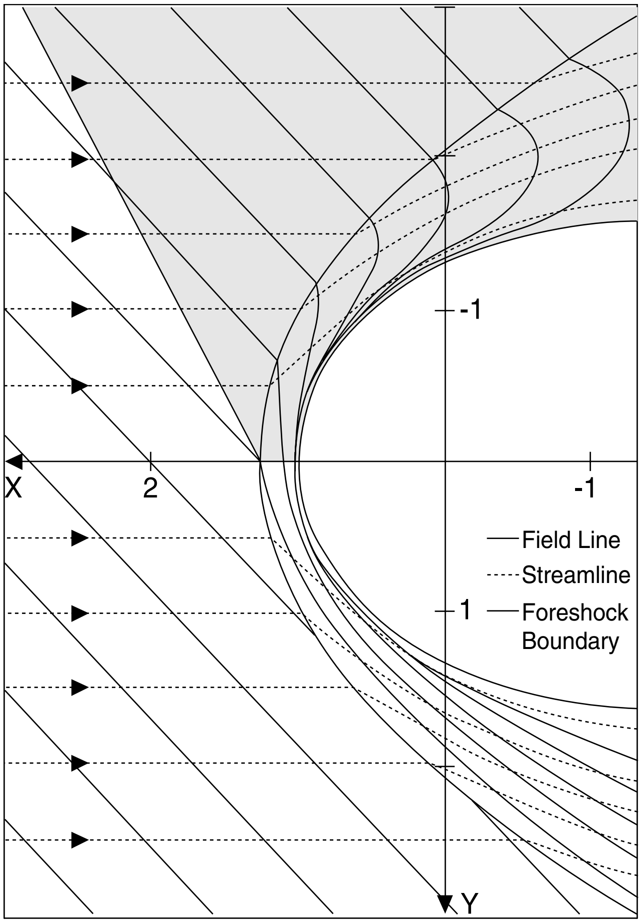


FIGURE 2.4 The interaction of the solar wind and the interplanetary magnetic field with a magnetospheric obstacle. Shaded is the foreshock region, where reflected ions cause disturbed solar wind, which is convected into the magnetosheath. Courtesy of C.T. Russell, University of California, Los Angeles.

wind, producing a wide variety of plasma wave modes. The group velocity of many of these large-amplitude waves is less than the solar wind velocity, and the waves are convected back into the shock. These waves play an important role in particle acceleration, especially of the high-energy tail of the ion distribution. In particular, owing to growth and saturation of the wave modes and generation of modes that compress the magnetic field and plasma populations, the convection of the waves back into the shock creates a set of converging "mirrors," where ions are accelerated in the first-order Fermi process.

Observed high-energy populations in the foreshock region agree extremely well with predictions from Monte Carlo simulations of this acceleration process. As ions become energized in the converging mirrors, their probability of escape from the system increases. Thus, the maximum energy attained by first-order Fermi acceleration depends simply on the size of the interaction region. At Earth, the maximum energy gain is about 150 times the original energy of the solar wind. Evidence of this scaling is seen when observations near Earth's bow shock are compared with observations at the much larger bow shock at Jupiter.

Another type of acceleration can occur near the shock. Ions and electrons can reflect off the shock or escape from the downstream region near the region where the convecting solar wind magnetic field is nearly tangent to the shock interface. These particles can "surf" or drift along the shock front, in the direction of the tangential electric field seen in the shock frame, gaining (or losing) energy while staying on the convecting magnetic field that is nearly tangent to the shock surface. This time-stationary unipolar electric field exists because an observer in the shock rest frame sees plasma and magnetic field advancing through the shock surface. Through this process, sometimes called fast Fermi acceleration, a small fraction of the incident solar wind particles can gain ~30 to 100 times their original energy.

A paradox soon was realized in the electron behavior: namely, that the same electric potential that slowed the ions should accelerate the electrons, whereas the electrons were heated less by the shock than were the ions. The resolution to this paradox is that electrons remain tied to the magnetic field as they cross the shock and move along that field and across equipotentials imposed by the solar wind in such a way that they see a minimal change in electric potential.

Finally, the study of Earth's bow shock has led to a better understanding of the preconditioning of the plasma and magnetic field that ultimately interact with Earth's magnetosphere. This understanding goes well beyond the prediction of bulk properties of the shocked solar wind plasma, which are well reproduced in global MHD codes. Quasi-parallel and quasi-perpendicular shocks, where the IMF is parallel to and perpendicular to the shock normal, respectively, have different effects on the magnetosphere. For example, mirror-mode waves produced at the bow shock and in the magnetosheath interact with the magnetopause when the magnetosheath in the vicinity of the Sun-Earth line is downstream of the quasi-perpendicular shock (this is the typical situation for the Parker spiral angle of the solar wind

BOX 2.1 DISSIPATION AT THE BOW SHOCK

The development of our understanding of the mechanisms responsible for dissipation at collisionless shocks such as Earth's bow shock is an excellent example of the interplay between analytic theory, in situ observations, and computer modeling. Early in shock research, it was apparent that the structure of the shock depended strongly on the Mach number of the solar wind flow and the angle between the solar wind magnetic field and the normal to the (curved) shock surface (θ_{bn}). In particular, above a critical Mach number, resistivity alone was shown to be inadequate to provide the change from upstream to downstream that was required to satisfy the magnetohydrodynamic jump conditions across the shock. Determining the dissipation mechanism for supercritical shocks had to wait for advances in space plasma instrumentation and improvement in computing facilities.

In the late 1970s and early 1980s, these two technological advances were achieved, and it was found that the dissipation at a supercritical shock occurs through reflection of a portion (~10 to 20 percent) of the incident solar wind ion beam off the shock front. For shock normal angles greater than 45° ($\theta_{bn} > 45°$), these reflected ions execute approximately one half of a gyro-orbit in the upstream region, gain energy in the solar wind electric field, and return to the shock. Computer simulations showed that the reflection required both electric and magnetic forces at the shock and led to distinctive signatures in the magnetic field and phase space distributions of the particles near the shock. These features were conclusively identified in high-time-resolution, in situ observations near Earth's bow shock. Thus, by the mid 1980s, the dissipation at Earth's quasi-perpendicular bow shock (i.e., the part of the bow shock where $\theta_{bn} > 45°$) and other perpendicular bow shocks was reasonably well understood. However, there was a fundamental problem in applying this same dissipation mechanism to quasi-parallel shocks (i.e., the part of the bow shock where $\theta_{bn} < 45°$).

In contrast to the quasi-perpendicular shock, simple particle trajectory calculations showed that ions reflected off the quasi-parallel shock in a uniform upstream magnetic field would not return to the shock to provide the necessary dissipation. Once again, detailed computer simulations and in situ observations combined to provide an answer. The nonuniformity of the upstream magnetic field and the instability of the shock itself are critical in understanding the dissipation mechanism at this type of shock. Computer simulations showed that the quasi-parallel shock undergoes a reformation process whereby the shock front steepens and forms a structure similar to that seen at the quasi-perpendicular shock, but continues to steepen and overturn. The overturning process creates a turbulent transition between the upstream and downstream states that does not resemble a uniform shock transition. Detailed in situ observations downstream of the quasi-parallel shock showed evidence of a periodic, quasi-perpendicular-like dissipation mechanism. These periods were interspersed with a hot, isotropic downstream population consistent with a turbulent transition from the upstream to downstream states. Thus, by the beginning of the last decade, the dissipation mechanisms at planetary bow shocks (and interplanetary shocks) were understood from a theoretical perspective and confirmed from high-resolution, in situ measurements at Earth's bow shock.

---

magnetic field). When this region is downstream from the quasi-parallel shock, the magnetosheath conditions are dramatically different. Large fluctuations in the plasma beta (ratio of the perpendicular pressure of the plasma and the magnetic field) and significantly increased magnetic turbulence produce a very different interaction with the magnetosphere. Also, particle distributions produced at the shock can be observed directly in the magnetosphere. A case in point is the recently discovered energetic ion population in Earth's magnetospheric cusps. This population has all the characteristics and solar wind dependencies of the energetic ion population created by first-order Fermi acceleration at Earth's bow shock.

Outstanding Question

How Are Shocks Modified by Multiple Components?

The development of a clear understanding of dissipation at Earth's bow shock has led to the possibility of understanding other types of collisionless shocks. In particular, Earth's bow shock represents a relatively clean example of a planetary bow shock where the supersonic solar wind is decelerated in response to the presence of Earth's magnetosphere. Even in this relatively clean example, the modification of the quasi-parallel shock due to the presence of a small (1 percent) population of "reflected" ions is profound. In other collisionless shocks, additional modifications occur because of the presence of additional plasma populations. For example, at comets, the outgassing of the comet and subsequent ionization of its extended atmosphere creates a population of ions that are picked up by the solar wind. In fact, mass loading is so strong at comets that the accompanying deceleration produces an obstacle to the solar wind, which produces the bow shock.

Plasma observations from spacecraft flybys of comets have not been sufficient to determine completely the nature of the changes in the shock due to the addition of heavy ions. These in situ investigations are important because other shocks of astrophysical interest (e.g., the heliospheric termination shock and supernova shocks) probably have significant effects attributable to the presence of secondary components such as high-energy particles (e.g., galactic cosmic rays). From the recent experience with Earth's bow shock, understanding the dissipation mechanisms at cometary bow shocks will require a strong interplay between computer simulations and high-resolution, in situ measurements.

MAGNETOSHEATH

Achievements

The magnetosheath is the region of shocked solar wind that is bounded by the thin boundaries of the bow shock on the upstream (or sunward) side and the magnetopause on the downstream (or earthward) side. Because the solar wind is supersonic, moving faster than the speed of the compressional wave in a plasma, the shock is the first place at which the solar wind is decelerated, with the possible exception of interactions with backstreaming ions, discussed in the preceding section. Once downstream from the shock in the subsonic magnetosheath, the plasma can be further slowed and deflected continuously by pressure forces. Eventually the plasma reaccelerates as it moves past Earth, reaching speeds comparable to those of the solar wind.

Magnetohydrodynamics dictates that there are two other low-frequency waves that can stand in the magnetosheath downstream from the fast mode shock and upstream of the magnetopause. These waves are the intermediate wave, which rotates the field and the plasma flow, and the slow mode, which can increase the field strength while decreasing the density. Currently, one of the controversies is how sharp these transitions must be. For example, a slow mode transition is predicted and possibly observed just upstream of the magnetopause, but other interpretations of and reasons for the absence of this effect have been offered. This is an active area of research, and it is important because it involves understanding the basic physics of a magnetized plasma. What is more, this mechanism determines how strongly the solar wind will interact with the magnetosphere.

The processes occurring at the inner edge of the magnetosheath, the magnetopause, are affected by the modification in the plasma as it moves through the magnetosheath. One effect is the formation of a plasma depletion layer adjacent to the boundary. Accompanying this reduction in the plasma density is an increase in the magnetic field strength. When the IMF has a southward-directed component (at Earth), reconnection in the subsolar region is more prevalent than when the IMF has a northward component and the plasma depletion is weaker. The understanding of the process of plasma depletion and piling up of the field has resulted in a successful prediction of the temperature anisotropy of the ions in the magnetosheath and plasma depletion layer (see Box 2.2). The existence of a depletion layer has important consequences for reconnection at the magnetopause. These consequences are discussed in the next section, "Magnetopause, Cusp, Boundary Layers."

The magnetosheaths of other planets have properties similar to Earth's magnetosheath. In particular, similar magnetic field turbulence is observed, as are similar plasma transitions. The magnetosheaths of active comets are interesting in that a strong region of piled up magnetic flux is formed outside the contact surface (the region where the solar wind magnetic field is excluded). This region is similar to the pileup of magnetic field at Earth's magnetopause. Also, comets and some planets can have different interactions in their magnetosheaths. At Earth, relatively little plasma escapes from the magnetosphere and enters the magnetosheath. This plasma usually consists of a relatively high-energy population having a density that is significantly less than the density of the shocked solar wind plasma. Other planets and comets have stronger internal sources. These produce significant particle populations in the magnetosheath, which affects the wave-particle interactions in the region, especially at comets. The mass loading produced by these ions also slows the flow and further increases the size of the induced magnetosphere.

Outstanding Questions

What Are the Specific Wave Mode Transitions in the Earth's Magnetosheath and How Sharp Are They?

As mentioned in the discussion above, the MHD description of the magnetosheath suggests that there

BOX 2.2 ION ANISOTROPY-DRIVEN WAVES IN THE MAGNETOSHEATH AND PLASMA DEPLETION LAYER

Spacecraft observations in Earth's magnetosheath show a distinct inverse correlation between the ion temperature anisotropy and the ratio of the perpendicular plasma pressure to the magnetic field pressure (plasma beta). Theoretical models of the magnetosheath reproduce this inverse correlation through wave-particle interactions. As plasma convects toward the magnetopause, the magnetic field increases, the perpendicular temperature of the shocked solar wind ions increases, and, because of losses along the magnetic field, the ion parallel temperature decreases. The increase in the magnetic field is accompanied by a decrease in the shocked solar wind density, which results in a reduction of the plasma beta as the shocked solar wind approaches the magnetopause.

The increase in the temperature anisotropy as the magnetopause is approached cannot continue unabated. Above a certain critical anisotropy, which depends on the plasma beta, the electromagnetic ion cyclotron instability is excited in the convecting plasma. Like all wave instabilities in a plasma, the waves interact with the particles and reduce the free-energy source of the instability. Thus, the temperature anisotropy decreases and the instability is quenched. As the plasma continues to approach the magnetopause, the anisotropy grows again until the ion cyclotron instability is once again excited. The net result is a temperature anisotropy that is maintained near the threshold of the ion cyclotron instability. This threshold depends inversely on plasma beta, so that the observed anisotropy-beta relation in the magnetosheath is reproduced by application of wave-particle interactions. This understanding is not incorporated in global MHD models but relies on kinetic and hybrid (massless electron) approaches.

should be specific wave mode transitions between the fast mode transition at the shock and the magnetopause. It is still an open question how sharp (and indeed if) these transitions occur. The understanding of these features in the magnetosheath will lead to a better understanding of the plasma that interacts with the magnetosphere at the magnetopause.

What Are the Plasma and Field Conditions Adjacent to the Magnetopause?

Surprisingly, an accurate description of the plasma and magnetic field parameters in the magnetosheath adjacent to the magnetopause is still lacking. Current state-of-the-art studies of the solar wind–magnetosphere interaction use either gas dynamic modeling (e.g., the Spreiter and Stahara gas dynamic model) or isotropic, single-fluid MHD. As a result, the total magnetic field, the ion and electron temperatures, and the temperature anisotropies (and therefore the division of plasma pressure into perpendicular and parallel components) have not been described accurately on a global scale. As will be seen in the next section, these input parameters are critical for understanding the most important plasma transfer process at the magnetopause, magnetic reconnection. Because the boundary is dynamic and has an a priori unknown geometry, including nonplanar structure, resolving this issue will require multipoint measurements of the plasma and magnetic field in the magnetosheath as well as three-dimensional, hybrid computer models to investigate the physical processes leading to the observed features discussed in Box 2.2.

MAGNETOPAUSE, CUSP, BOUNDARY LAYERS

Achievements

The magnetopause is where the external plasma and magnetic field pressure of the magnetosheath meet the internal plasma and magnetic field pressure of the magnetosphere, establishing a pressure balance. It is usually considered to be the outer boundary of the magnetosphere. The structure of Earth's magnetopause is complicated because the plasmas on both sides of the discontinuity are magnetized, so that when the fields on opposite sides of the boundary are nearly antiparallel they can reconnect. This connection spoils the simple topological distinction between solar wind and magnetospheric plasma by creating a region that contains both plasmas on the same magnetic field lines. Under these circumstances it is best to think of the magnetopause as a region rather than a thin discontinuity.

For typical solar wind conditions, the distance from Earth's center to the magnetopause along the Earth-Sun line, called the subsolar distance, is about 10 times the radius of Earth (1 R_E = 6,371 km). The shape of this boundary can be approximated by an ellipsoid of revolution about the Earth-Sun line, with the magnetopause at the terminator about 50 percent further from Earth than it is at the subsolar point. In the magnetotail, this approximation is not accurate. The magnetotail continues to expand in cross section and extends many hundreds of Earth radii in the direction away from the Sun.

Since the magnetopause is in pressure balance, changes in the internal or external pressure cause the location and shape of this boundary to change. The solar wind dynamic pressure (that is, the solar wind mass density times the square of the solar wind velocity) is far from constant. This dynamic pressure is converted to plasma thermal pressure and magnetic field pressure in the magnetosheath. Thus, variations in the solar wind dynamic pressure cause the magnetopause to move toward and away from Earth. The subsolar distance can move in to about half its nominal distance for extremely high solar wind pressure and can move out to about twice its nominal distance for extremely low solar wind pressure. The size of planetary magnetospheres varies greatly in absolute and relative scales. At Mercury the magnetosphere's nose does not extend much above the planet's surface on the dayside. In contrast, at Jupiter many of the planet's moons orbit inside the magnetosphere. If the plasma and magnetic field in the magnetosheath and the magnetopause interacted only at this basic level, then the physics of the interaction would be already solved. However, the major difference between this simple interaction and that at Earth's magnetopause is that it neglects the external magnetic field in the magnetosheath. Although the magnetic field in the magnetosheath is weaker than its magnetospheric counterpart, it is the presence of this magnetic field that causes most of the interesting phenomena associated with Earth's magnetopause.

In 1961 J.W. Dungey suggested that the solar wind and terrestrial magnetic fields would interconnect at a neutral point when the magnetospheric and solar wind magnetic fields were nearly antiparallel (see Box 2.3). Under certain assumptions, the change in plasma parameters at the magnetopause can be predicted, allowing the reconnection hypothesis to be tested. Furthermore, the kinetic treatment of the magnetopause, in which particle motions are followed, leads to specific predictions for the ion and electron distributions that can be observed in the various magnetopause layers. Spacecraft- and ground-based observations in the last decade have firmly established that reconnection is the

BOX 2.3 MHD ASPECTS OF RECONNECTION (OUTSIDE THE DIFFUSION REGION)

Although predicted by J.W. Dungey in 1961, the details of magnetic reconnection and its possible application to the magnetopause were not worked out until several years later. As discussed above, reconnection interconnects magnetized plasmas on different sides of a boundary, often leading to accelerated flows. This signature of magnetic interconnection accompanied by fast flows has become the major indicator of the presence of reconnection. Predictions for antisotropic MHD were developed in 1970 under the assumption that the magnetopause is locally a one-dimensional rotational discontinuity whose properties remain steady on time scales long compared with the transfer of a single-fluid plasma element across the discontinuity. Under these assumptions, the MHD equations yield inequalities at the magnetopause that are related to the reconnection rate. Specifically the normal component of the magnetic field at the magnetopause, B_n, is not zero across a rotational discontinuity but is zero across a tangential discontinuity across which the fields on the two sides are not connected. While difficult to quantify, this nonzero normal component has been observed at the magnetopause. Spacecraft observations (initially two-dimensional plasma observations in the 1970s and then three-dimensional observations in the mid-1980s) became detailed enough to provide additional observational verification of predictions from single-fluid MHD theory. Even with the improved observations, some additional inequalities such as the normal component of the velocity and the tangential component of the electric field are exceedingly difficult to measure because they are small compared with the motion of the boundary in response to the ever-changing solar wind.

The existence of a normal component of the magnetic field across the magnetopause requires the magnetized plasma on both sides of the discontinuity to flow together. This leads to the existence of a reference frame, called the de Hoffmann-Teller (HT) reference frame, which slides along the magnetopause with the tangential velocity of the interconnected field lines. By itself, the existence of an HT frame is a necessary (but not a sufficient) condition for reconnection. Two important relationships develop out of the tangential Maxwell stress conditions and the existence of an HT frame at the boundary. These relationships also represent necessary, but not sufficient, conditions for the existence of magnetic reconnection. The first is related to mass conservation across the magnetopause and is difficult to test without high-resolution mass-resolved measurements at the discontinuity. The second, often referred to as the Walen relation, indicates that plasma on either side of the discontinuity flows at the local Alfvén speed in the HT frame. Spacecraft observations have demonstrated that ion (and more accurately electron) flows at the magnetopause often have this Alfvénic flow property. The verification of these predictions and those at other scale lengths firmly establishes reconnection as an important transfer process at the magnetopause.

primary process for the transfer of mass, energy, and momentum across the magnetopause and intimately controls the properties of the polar cusps (see Box 2.4).

When magnetic reconnection occurs for directly southward interplanetary magnetic fields (at Earth), the magnetosheath and magnetospheric magnetic field lines interconnect in a relatively small region (often called the diffusion region) near the subsolar point. Because the diffusion region is very small and (possibly) moving, the probability that a spacecraft will cross through the diffusion region is thought to be small. The understanding of the physics of this region is critical to understanding how reconnection works, and the understanding of reconnection is important to astrophysics and solar physics as well as to planetary magnetospheres.

Outstanding Questions

What Is the Global Reconnection Rate at the Magnetopause and How Does It Change with IMF or Solar Wind Conditions?

The MHD tests of magnetic reconnection at the magnetopause fall into two separate categories. Inequalities such as the nonzero normal component to the magnetic field relate to the reconnection rate. Tests of the magnitude of the tangential flow velocity, such as the Walen test, are independent of the reconnection rate. Unfortunately, because the reconnection rate is very low, quantitative measure of this rate is exceedingly difficult using, for example, the normal component to the magnetic field or the normal component to the flow velocity. Thus, the inflow rate is not well known. The total extent of the reconnection region is also not well known. Typically, single spacecraft observe reconnection features at a single point on the magnetopause or two or more spacecraft separated by a small distance (when compared with the total magnetopause extent) observe reconnection signatures. Ground-based radar observations are very important in this regard, especially the SuperDARN radar network.

Both the rate and the total extent of the reconnection region are required to determine the total inflow rate into the magnetosphere. Without a measure of this rate, it is difficult to assess the effectiveness of reconnection in providing the mass, energy, and momentum transfer into the magnetosphere. Certainly this transfer is a function of the external solar wind conditions, especially the IMF orientation. It is also a function of geometry. Reconnection generally occurs in a three-dimensional geometry while theoretical treatments are generally done in two dimensions.

What Is the Interplay Between the Microscale and Mesoscale Aspects of Reconnection? Is the Dissipation Driven by External Boundary Conditions or Internal Microscale Instabilities?

Despite the small size of the region where magnetic field lines interconnect (the diffusion region), it has global consequences. The diffusion region remains one of the last unexplored regions of the magnetopause, and its study is a matter of very high priority. It consists of transitions on several scale lengths. Two of the more important transitions are the regions where the ions decouple

BOX 2.4 DISCOVERY OF THE DIRECT CONNECTION OF THE BOW SHOCK TO THE IONOSPHERE THROUGH THE CUSP

Following his prediction of magnetic reconnection at Earth's subsolar magnetopause during periods when the IMF had a southward (or negative B_z) component, J.W. Dungey predicted that there would be reconnection at the high-latitude magnetopause during periods of northward IMF. Since magnetic reconnection provides a means to interconnect solar wind magnetic field lines with those in the magnetosphere, these two predictions indicate that there could be a nearly continuous connection between the magnetosphere and Earth's bow shock.

Nowhere is this connection more apparent than in Earth's magnetospheric cusps. These high-latitude regions exhibit direct entry of solar wind plasma. Recent modeling and detailed comparison of these models with observations have shown that the properties of the cusp are consistent with magnetic reconnection at the magnetopause.

New observations made in the last decade showed the presence of an energetic ion component of solar wind origin that had gone undetected. It was first suggested that this energetic ion component was accelerated out of the magnetosheath ion population in the cusp. While some such acceleration may take place, analysis of the recent observations and comparisons with global MHD simulations indicate that this energetic ion component is a by-product of the nearly continuous connection between the magnetospheric cusps and the solar wind. Solar wind ions energized at Earth's quasi-parallel bow shock have direct access to the cusp along reconnected field lines.

from the magnetic field and the much smaller region where the electrons decouple. Ion decoupling, in a location called the Hall region, has recently been reported for reconnection in Earth's magnetotail. The lack of observations and the difficulty in theory to deal with very disparate scale lengths has led to some ambiguity about what controls reconnection. Certainly boundary conditions must play an important role in modulating the rate of reconnection, but so too the conditions at the reconnection point must be important, perhaps controlling where reconnection takes place. This control in turn will affect the transfer of magnetic flux from closed to open field regions. Thus the determination of the role of microphysics in reconnection is critical to understanding of magnetospheric dynamics.

Is Reconnection Patchy or Quasi-continuous, Inherently Unstable or Quasi-static? Does It Occur When Magnetic Fields Are Exactly Antiparallel or When Only One Component Is Oppositely Directed?

The preceding question dealt with the "why" of reconnection. These questions relate to the "when, where, and how much" of reconnection. However, they also have important implications for the "why" of reconnection. All of them remain unanswered because of one fundamental limitation of previous observations and several fundamental problems with the reconnection process and observations at the magnetopause in general. Previous measurements at the magnetopause were mostly from single spacecraft or from two or more spacecraft separated by distances that were small compared with the possible scale sizes of patchy reconnection. Patchy reconnection could be occurring on scale sizes as small as a few thousand kilometers or as large as several Earth radii. Field-line convection away from the reconnection site can give the appearance of a large reconnection region when in fact the neutral line where the diffusion is occurring is relatively short. Thus, it is not sufficient to have large spacecraft separations; they must be separated in the appropriate direction.

Fundamental problems with the reconnection process and with observations at the magnetopause in general include the difficulty of measuring the reconnection rate using in situ measurements at the magnetopause, the constant motion of the magnetopause so that observations are limited to brief crossings of the boundary, changes in the location of reconnection that follow the variable IMF direction so that spacecraft positions at the boundary may be ideally suited for only a narrow range of IMF orientations, and the large variety of scales (from electron to ion to current sheet thicknesses) associated with reconnection, ranging from <1 km to many R_E. These limitations and problems cannot be overcome by any mission consisting of a single spacecraft. Multispacecraft missions carrying high-time-resolution instruments are needed.

MAGNETOTAIL

Achievements

The extended magnetotail is the direct manifestation of the interaction between Earth's magnetic field and the solar wind (a magnetotail is similar to cometary tails, whose study led to the discovery of the solar wind). The magnetotail is also relevant because it plays a major role in the dynamics of the magnetosphere, acting as an energy storage region and governing the sudden release of this energy in magnetospheric substorms. Progress in recent decades has led to a much more detailed understanding of activity in the tail and its connection with Earth.

Earth's magnetotail exhibits various modes of activity, including not only substorms but also pseudosubstorm onsets ("pseudobreakups") and steady magnetospheric convection (SMC) events. All of these may be accompanied by short-duration fast plasma flows, so-called bursty bulk flows (BBFs). Bursty bulk plasma flows in the plasma sheet (see Box 2.5) are fast flows typically lasting 10 minutes and having individual peaks of about a minute. Pseudobreakups are small activations, observed in both the magnetotail and the ionosphere prior to substorms, that do not lead to the global reconfiguration of the magnetotail and the expansion of the auroral oval as do full substorms. SMC events are characterized by overall stability under steady, driving solar wind conditions; numerous transient activations; and an absence of substorms.

Even in the absence of substorms, the tail plasma sheet is found to be highly variable, giving the appearance of a turbulent rather than a laminar flow state. Since the maximum scale size, the tail diameter, is only about two decades larger than the ion gyroradius, however, the MHD turbulence in the tail is far from homogeneous. Since the near-Earth plasma sheet is a source of other populations (e.g., the ring current and the radiation belts), its variability affects those populations as well.

The overall evolution of the magnetotail during a substorm is well understood. It consists of a growth phase, initiated by southward IMF, during which magnetic flux and energy are transported from the front side

BOX 2.5 DISCOVERY OF BURSTY BULK FLOWS

The introduction of the open magnetospheric convection model was one of the great conceptual inventions of the early 1960s. In this paradigm, magnetospheric convection is driven by dayside reconnection when the interplanetary magnetic field (IMF) has a southward component. Subsequently, interconnected field lines are dragged into the tail, where they again undergo reconnection, thereby trapping plasma on closed field lines. Plasma and magnetic flux are then transported around Earth to the dayside, where they again can undergo reconnection with the solar wind, in a seemingly steady circulation. This paradigm still governs our basic understanding of magnetospheric circulation today. However, it is particularly the steady-state aspect of this picture that has come under scrutiny, because of inconsistencies both observationally and theoretically.

In a steady convection state, the electric field should be a potential field. In ideal MHD, with negligible parallel electric field, this potential is constant along magnetic field lines. However, average electric fields in the tail plasma sheet are considerably smaller than those obtained by mapping the ionospheric potential to the tail. Furthermore, the average tail configuration is inconsistent with adiabatic, i.e., entropy-conserving transport from the tail toward Earth. Entropy conservation would lead to pressure increases that by far exceed the pressure necessary to balance the observed magnetic forces.

The solution to these discrepancies lies in the fact that the tail hardly ever assumes a steady state, and that entropy conservation on closed magnetic flux tubes is sporadically violated by the severance of plasmoids via magnetic reconnection. The manifestation of these features is that fast plasma flows in the tail plasma sheet, and correspondingly high electric fields, occur only in bursts of short (~10 min) duration with spikes of even shorter periods (~1 min). Despite their short duration and relatively sparse occurrence, these bursty bulk flows play the most important role in the transport of magnetic flux and energy in the tail.

Although the incidence of bursty bulk flows is generally correlated with geomagnetic activity, their occurrence is not confined to a particular substorm phase or even to substorms alone. They are frequently observed during steady magnetospheric convection events, characterized by the absence of substorms under steady solar wind driving conditions. They are apparently highly correlated with auroral intensifications.

---

of the magnetosphere to the tail, where they become temporarily stored in increasingly stretched and enhanced magnetic fields. This phase is followed by a release phase, during which some inner portion of the tail collapses toward a more dipolelike field, while an outer portion of the plasma sheet, a "plasmoid," becomes detached via magnetic reconnection and ejected antisunward. Tail observations have shown that this reconnection site on average forms between 20 and 30 R_E down the tail (see Box 2.6). A recovery phase restores the tail to its original state. However, not all accept this model, which predicts that the substorm process begins in the near tail, as illustrated in Figure 2.5. Some predict that the onset begins at lower latitudes and moves tailward.

The first two phases are well understood and have been successfully modeled on large scales, including the acceleration of particles (see Box 2.7). The processes that govern the transition from one phase to another are less well understood. Various observations, as well as plasma simulations using a variety of approaches, have demonstrated the important role of the formation of a thin current sheet in the near tail, embedded in the wider plasma sheet, prior to the substorm release phase. As at the magnetopause, the development of strong gradients with small characteristic scales is a necessary condition to break the frozen-in flux condition of ideal MHD, enabling current disruption and magnetic reconnection. Thus, a thin sheet with sufficiently small scales is a crucial step toward instability.

Advances in plasma simulations have also shed more light on the physics of reconnection, at least under laminar conditions. The simulations suggest a characteristic structure with threefold-embedded regions. The innermost region is governed by nonadiabatic electrons, which provide the dissipation mechanism. This region is surrounded by a Hall region, where electrons are coupled to the magnetic field but ions are still uncoupled. The outermost region then is governed by fluidlike behavior, where both particle species are coupled to the field.

Although reconnection underlies many of the processes that govern both the tail and the magnetopause, it is not sufficient to study just one or the other of these

BOX 2.6 NEAR AND DISTANT NEUTRAL POINT LOCATIONS

The steady magnetospheric convection paradigm for southward IMF involves two magnetic reconnection sites, at the dayside magnetopause and in the distant tail. The great conceptual invention of the 1970s was the realization that a further reconnection site should form sporadically in the near tail at a near-Earth neutral line. This solves not only the entropy and mass transport problem but also provides an explanation for the generation of fast plasma flows. Thorough statistical analyses on the basis of Geotail observations in the past decade have confirmed the existence of these neutral lines (reconnection sites) and provided their average locations. On the basis of the occurrence of fast earthward or tailward flows related to substorm onsets, it is found that the near-Earth reconnection site typically forms at 20–30 R_E distance down the tail. In contrast, the location of the distant reconnection site or neutral line appears to be much more variable. Its location apparently varies, typically from ~60 R_E to distances beyond 200 R_E. Occasionally it may also be located inside 60 R_E. Figure 2.6.1 illustrates the relative magnetic topology of the two neutral points and how the magnetic flux in the tail might respond to southward and northward turnings.

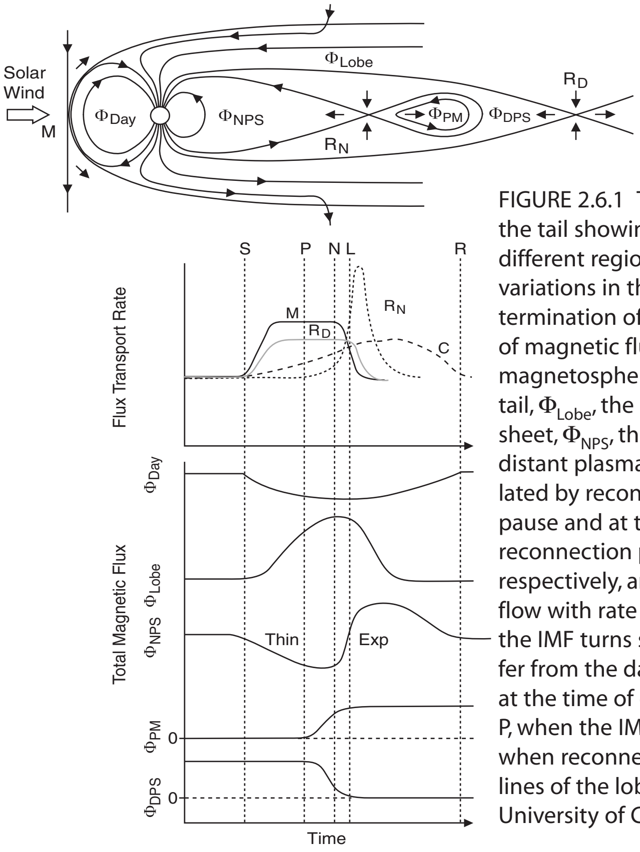


FIGURE 2.6.1 Two neutral point models of the tail showing how the magnetic flux in different regions of the tail responds to variations in the IMF and the initiation and termination of reconnection. The amount of magnetic flux in the closed dayside magnetosphere, Φ_{Day}, the open lobe of the tail, Φ_{Lobe}, the closed near-Earth plasma sheet, Φ_{NPS}, the plasmoid, Φ_{PM}, and the distant plasma sheet, Φ_{DPS}, are all modulated by reconnection at the magnetopause and at the near and distant reconnection points at rates M, R_N, and R_D, respectively, and by the return convective flow with rate C. These rates change when the IMF turns southward to initiate transfer from the dayside to the lobe at time S; at the time of onset of plasmoid formation, P, when the IMF turns northward, N; and when reconnection reaches the open field lines of the lobe L. Courtesy of C.T. Russell, University of California, Los Angeles.

regions. First, the plasma conditions are different at the two locations; second, we need to learn how reconnection produces the observed dynamics of the magnetopause and the tail. For example, at the magnetopause the interaction geometry produces a current layer across which the magnetic field is generally not antiparallel, and there is no interconnecting magnetic component. In contrast, in the tail the fields in the two lobes are nearly antiparallel but are generally linked by a strong interconnecting component. At the magnetopause, dynamic behavior occurs when a normal component arises. In the tail, dynamics occurs when the normal component reaches a very small value.

Outstanding Questions

Despite the many advances in understanding the magnetotail and its dynamics, major questions remain, such as the cause of substorm onset, and new questions have come up from new discoveries such as BBFs.

What Causes the Onset of Terrestrial Magnetotail Activity and, Particularly, Substorms?

This is one of the oldest questions. However, the competing views of the onset sequence now differ only in the timing of events and only by one or a few minutes. Nonetheless, since the processes that occur within

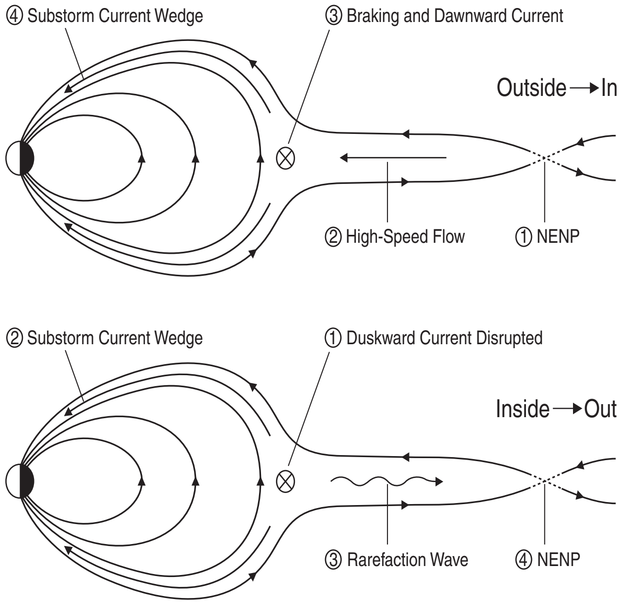


FIGURE 2.5 *(top)* The near-Earth neutral point (NENP) model predicts that the substorm begins with reconnection in the near tail. *(bottom)* In contrast, the current disruption model predicts that reconnection begins in the region around 8 R_E and moves tailward. Both models have currents deep in the magnetosphere and reconnection in the tail but differ on the relative timing of events. Courtesy of H.E. Spence, Boston University.

this time frame involve the microphysics of the onset, they are at the core of an understanding of how collisionless dynamical processes are initiated. Solution of the controversy requires high-temporal-resolution, multipoint measurements, both in space and on the ground, and a deeper understanding of the reconnection process.

What Causes the Localization of Activity, Such as Reconnection, Bursty Bulk Flows, Pseudo-onsets, and, Similarly, Flux Transfer Events at the Magnetopause?

Reconnection is thought to occur where the magnetic field configuration resembles an X. If two field lines merge at a point, then the reconnection location can be referred to as a neutral point. If the merging of two field lines extends some distance across the tail, it can be referred to as a neutral line, but in this case the field strength might not be zero along its entire length. While statistical studies of tail data have narrowed down the average location where the near-Earth neutral line forms, its cross-tail extent is much less well known. In addition, as three-dimensional plasma simulations show, the extent of an active reconnection site can be much smaller than the extent of a neutral line per se. Furthermore, how do we identify a neutral line and a reconnection site in general cases when the neutral line is not strictly “neutral” anymore? Is there a “guide field” parallel to the neutral line? Similar problems exist for mechanisms that do not involve reconnection.

In the case of BBFs in the tail, serendipitous multispacecraft observations have shown that large differences in flow activity are often observed even when the spacecraft are nearly colocated. Estimates of the amount of magnetic flux propagated earthward during the substorm sequence compared with the amount of flux contained in a BBF also require that the cross-tail dimension of the BBF be very small. Again, a constellation of globally distributed spacecraft (Magnetospheric Constella-

BOX 2.7 ENERGETIC PARTICLE INJECTIONS

Rapid increases of energetic ion and electron fluxes in the inner magnetotail, observed extensively at geosynchronous orbit, are an intrinsic part of the substorm process. The breakdown of adiabatic particle behavior at a reconnection site has for a long time suggested the near-Earth neutral line as a potential acceleration site to generate the particles with energies of tens to hundreds of keV that constitute these energetic particle injections. However, the successful identification of the average location of the reconnection site near 20–30 R_E downtail has made this interpretation inconsistent with observations of dispersionless injections at geosynchronous orbit. On the other hand, the fact that flux increases are typically limited to the energy range of tens to hundreds of keV seemed to rule out an adiabatic heating mechanism such as betatron acceleration. Various simulations now have solved this dilemma by demonstrating that the induced electric field associated with the collapsing inner tail is responsible for the major part of substorm acceleration. Reconnection is the process that enables a collapse, but acceleration at the reconnection site is not a major contributor to the enhanced energetic particle fluxes in the inner tail. The apparent nonadiabatic effects are explained by the finite cross-tail extent of the acceleration region. Particles at higher energies drift too fast across the tail through this region to participate in the collapse, so that they do not experience the associated betatron or Fermi acceleration.

tion) has been proposed to help answer this outstanding question.

What Causes Quenching of Reconnection and Other Activity?

There is evidence that reconnection, as the likely cause of fast flows, sometimes ceases before lobe field lines become involved. What is the cause for this cessation? When reconnection proceeds to the lobes, it apparently does not involve the entire lobe flux. Why is this, and when does it cease in this case? An analogous question exists at the magnetopause but may be more addressable in the tail, where the plasma is more stationary, at least initially.

What Is the Role of Boundary Conditions, such as the IMF North-South Direction, and What Are Some Possible External Triggers of Activity?

There is plausible evidence that substorms are triggered by a northward shift of the IMF in perhaps half of the observed cases, with plausible evidence also that at least some substorms are not triggered externally. How does the trigger work, and why does it sometimes work and at other times not?

What Is the Relationship Between Near/Midtail Reconnection on the One Hand and Substorm Onset and Substorm Auroral Features on the Other?

Currently there is debate on the causal relationship between processes in the midtail and the near-geosynchronous orbit region during substorm onset. The debate centers on what role midtail reconnection plays in the near-Earth reconfiguration at substorm onset. Does midtail reconnection drive substorm onset behavior in the near-Earth region or is it a consequence of it? As mentioned above, the relevant time period covers only a few minutes, but an understanding of the processes within this period is crucial for understanding collisionless plasma instability.

How Do the Mid- and Distant Tail Map to the Planet Magnetically and Dynamically?

Empirical and global field models provide general qualitative mapping between the various magnetotail regions and their ionospheric counterparts. But highly variable, large-scale current systems render mappings very uncertain except in a statistical sense. More importantly, a number of observations suggest that ionospheric and magnetospheric flows are not always well coupled, but we do not know the extent of this decoupling or its causes or consequences. What can ionospheric measurements really tell us about conditions in the magnetosphere?

What Determines the Terrestrial Plasma Sheet Density? What Is the Role of the Ionospheric and Plasmaspheric Plasma in Populating Earth's Plasma Sheet? How Does the Solar Wind Populate the Plasma Sheet?

The plasma sheet is populated by plasmas originating from various sources: the solar wind and Earth's ionosphere and plasmasphere. Changes in plasma composition after large geomagnetic storms show that Earth's ionosphere can contribute significant mass to the plasma sheet. However, the relative contribution of the ionosphere to the plasma sheet is not known, especially how it varies with geomagnetic activity and with radial distance. There are also outstanding questions about other possible pathways from the inner magnetosphere to the plasma sheet, such as through dayside reconnection and convection back into the tail.

INNER MAGNETOSPHERE

Achievements

Geomagnetic storms are the largest manifestation of energy coupling between the solar wind and magnetosphere and are directly driven by solar wind transients that have large solar wind dynamic pressure and IMF B_z south magnitudes. Intense ring currents are formed in the wake of these disturbances and are created by strong and variable convection, often reaching deep into the inner magnetosphere. The partial ring current (the current generated by energetic particles on open drift trajectories) has been recognized as the main contributor to the main phase ring current.

Recently, efforts to make hybrid simulations by coupling global MHD codes with kinetic models have advanced our ability to study magnetic storms and substorms in the midtail and near-Earth plasma sheet. This allows analyzing the behavior of non-Maxwellian distribution functions that are observed with in situ spacecraft and provides clues to the source mechanism and location for acceleration of energetic particles.

Empirical magnetic field and magnetospheric specification models that include observations made in the inner magnetosphere over a wide range of geomagnetic conditions have significantly improved the ability to describe the configuration of the magnetosphere during geomagnetically disturbed intervals. These models are aiding in the understanding of the relevant scale sizes of magnetospheric structure and contributing to the development of future multiprobe missions.

The energetic ions (also known as the ring current) present one aspect of the particle energization. The strong, storm-associated, penetrating electric fields drive plasma sheet ions deep into the inner regions and to low *L* values[1] during the main phase of magnetic storms. The composition of the ring current—that is, the partitioning of the energy between the different species and, ultimately, different plasma sources—has been an intense topic of discussion, research, and modeling during the past decade. The models have reached a high level of maturity, and their ability to characterize the ring current is now mostly limited by our inability to specify the phase-space distributions in the source region, the near-Earth plasma sheet. They are also limited by the inadequacy of current electric and magnetic field models for the inner magnetosphere. Only with the Combined Release and Radiation Effects Satellite (CRRES) and Polar observations from this past decade has it become clear that the electric fields penetrate very deep into the inner magnetosphere during storms, even to *L* values less than 2.5. The local time and radial distributions of the fields are not well known because the measurements have been single-point. The spatial picture we have of the equatorial electric fields and their variability is statistical. The empirical electric field models in use today are based on low-altitude, polar-orbiting satellite and ground-based observations. As such, they do not cover the inner magnetosphere well. They also do not account for potential drops between the equator and the ionosphere that exist during disturbed periods, nor do they include the highly time-dependent and strong inductive electric fields. What is needed is to determine the spatial/temporal structure of the electric and magnetic fields throughout the inner magnetosphere at or near the magnetic equator during magnetic storms and substorms. This knowledge is necessary if we are ever to understand the transport, energization, entrapment, and flow of energy into and through the inner magnetosphere. The energetic electrons that make up the radiation belts (electron ring current) are also often accelerated and transported deep into the inner magnetosphere during sudden impulses, as illustrated by the simulation in Figure 2.6. In addition, they can be accelerated by magnetic activity. It has been known for some time that increases in the flux of high-energy electrons are associated with high-speed solar wind streams. Recently it has been determined that the high-speed solar wind must be accompanied by IMF $B_z < 0$ for the enhancements. Yet those conditions do not guarantee that the electron enhancements occur (i.e., they are necessary but not sufficient conditions). The exact mechanisms that control the processes and the physics involved in the electron acceleration are not understood at this time. Several models have been proposed, from the classic radial diffusion picture, driven by stochastic electric and magnetic field fluctuations, to shock resonance acceleration, recirculation models, and ULF wave-electron interactions. Again, the major impediment to solving the electron acceleration and transport problem is the paucity of critical electron observations and poor knowledge of the background magnetic and electric fields during the acceleration events.

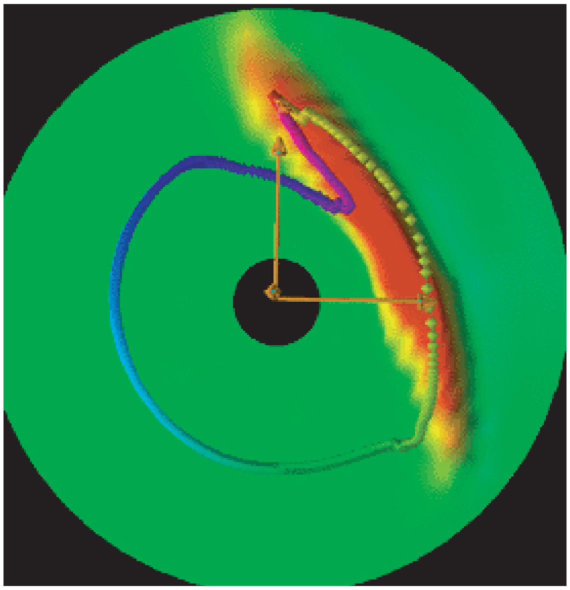


FIGURE 2.6 Simulation of electron acceleration in the magnetosphere during a shock compression event. Plasma pressure shown by color contours. Also shown is a distorted ring of energetic electrons. Courtesy of M.K. Hudson, Dartmouth College.

The current state-of-the-art radiation belt models now include fairly realistic inner magnetospheric plasma populations, including the dense, cold plasmaspheric plasma. The plasmasphere is an important region for the generation of waves and the location for significant radiation belt loss processes.

---

[1] *L* is the dipolar shell parameter, defined as $L = R_E/\cos2(M)$, where R_E is the number of Earth radii and *M* is the geomagnetic (dipolar) latitude. For an explanation of *L* value, see <http://pluto.space.swri.edu/IMAGE/glossary/units.html>.

Outstanding Questions

Which Processes Energize Ring Current and Radiation Belt Particles?

There are significant differences in the intensity of the ring current and the flux of the radiation belts from storm to storm. Which processes are involved in the generation of the two plasma populations and what differences in solar wind conditions determine the efficacy of the processes?

What Is the Nature of the Source Population for Inner Magnetosphere Electrons (and Ions) Before and During Events?

During the main phase of a magnetic storm the electron fluxes decrease dramatically. However, examining fluxes is not the appropriate way to interpret the data. One needs to examine the phase space densities at constant first and second invariant. This requires using magnetic field models. The present field models are wrong during the storm main phase because of the strong current systems that are operating and the fact that the magnetosphere is usually compressed. Accurate time-dependent models of the magnetic and electric fields are required to put the physics analyses on solid footing. Single-point satellite measurements and ground-based measurements cannot provide the data required to generate such models. New observation and analytical techniques are required.

One technique that has been suggested is to use energetic particle and field observations from small constellations of satellites to derive, self-consistently, the field models. This would be done by adjusting a field model to agree with the multipoint field observations. Then the particle trajectories would be calculated in phase space and the models further adjusted to bring the satellite-to-satellite mappings of the particles into agreement with the observations. Thus the multipoint magnetic field and energetic electron and ion observations can, in principle, be used to generate a self-consistent and dynamic field model on particle-drift-period time scales instead of satellite-orbital-period time scales.

How Does the Electric Field (Convection Pattern) Dynamically Evolve During Storms?

Similarly, combining the derived dynamic field model with multipoint electric field and plasma measurements may allow inferring an electric field model for the inner magnetosphere. To determine this evolution successfully requires a sufficiently dense and globally distributed set of field and particle observations throughout the inner magnetosphere such as would be provided by an inner magnetosphere constellation–type mission. This mission would require significant development of data assimilation and modeling techniques beyond what is presently possible and a range of new and developing technologies.

PLASMASPHERE

Achievements

Some of the recent achievements in understanding the plasmasphere include the first global images by the IMAGE spacecraft and their use to characterize cavities within the plasmasphere. IMAGE has obtained the first global images of the plasmasphere at time scales (10 min) sufficient to track its dynamics. The images are of He^+ emissions at 30.4 nm (resonantly scattered sunlight). The IMAGE data have verified the theory of plasma tails and revealed a shoulder-type feature that forms in the dawn sector following sharp northward turnings of the IMF and then corotates with Earth through the dayside. IMAGE EUV images have revealed density cavities that form within the plasmasphere and corotate with Earth; these cavities have been inferred from in situ observations. They have been shown to be the sites for generation of kilometric continuum radiation. EUV and ENA images from IMAGE have shown that the observed ENA emissions from the ring current are diminished inside the plasmapause and inside the plasma tails. Correlative measurements with Polar have shown that these effects are caused by changes in the pitch-angle distributions of ring-current ions in regions of higher cold plasma density.

Outstanding Questions

How Does the Plasmasphere Respond to Changes in Global and Local Electric Fields?

The large-scale structure and size of the plasmasphere depends on the interplay between the global convection and corotation electric fields and ionospheric refilling. We know that to the first order the plasmasphere expands during intervals of low, steady geomagnetic activity (i.e., when there is a low and steady convection electric field) and that the plasmasphere shrinks during intervals of high geomagnetic activity. However, statistical studies of plasmaspheric structure repeatably show large variations in plasmaspheric extent for essentially all levels of geomagnetic activity. In addition, significant structure as a function of local time is observed in response to both stormtime and substorm electric fields.

What Role Does the Plasmasphere Play in ULF Wave Generation, Modulation, and Propagation?

The plasmapause is thought to act as a natural cavity for global ULF wave oscillations and as a sharp boundary for ULF waves propagating into the inner magnetosphere. Unfortunately, there have been few in situ simultaneous observations of both the ULF waves and the cold thermal plasma of the plasmasphere. In a few cases, such as from CRRES, large-amplitude ULF waves were observed with corresponding density fluctuations, suggestive of the plasmapause boundary moving back and forth across the spacecraft. The modulation of ULF frequency and amplitude has also been observed across the boundary. However, no systematic study has been made of the ULF wave environment near the plasmapause. The advent of ground-based ULF resonance techniques that are able to identify the plasmapause location from the ground should address the role of the plasmapause in modulating and generating ULF waves.

What Is the Role of the Terrestrial Plasmasphere in Modulating Ring Current and Radiation Belt Particles?

Through Coulomb collisions and wave-particle interactions near the plasmapause, ring current and radiation belt particles can be lost. In addition, it has been suggested that ULF wave drift resonance may be responsible for the energization of relativistic electrons. Again, there have been very few observations of simultaneous ULF, VLF, energetic particle, and thermal plasma populations, so the role of the plasmasphere in modulating ring current and radiation belt particles is not clearly known.

What Is the Mass Composition of the Plasmasphere, and How Does It Change?

Measuring the thermal mass composition is experimentally very difficult, so little information exists for the mass composition of the plasmasphere. Observations of energetic particle composition and thermal plasma data from instruments on the Dynamics Explorer spacecraft have given inconsistent mixing ratios for helium and hydrogen, though ratios from 0.01 to 0.1 are most common. Essentially there are no statistical studies of the dynamics of the mixing ratios of the heavy ions as a function of geomagnetic activity, so it is generally assumed that the mixing ratio is slowly varying and nearly constant. This assumption is needed to interpret the IMAGE EUV observations, which measure resonantly scattered sunlight in the He^+ 30.4-nm line. The development of paired, ground-based magnetometer chains and the ability to routinely infer inner magnetospheric mass density should be able to address this fundamental question.

SOLAR WIND INTERACTIONS WITH WEAKLY MAGNETIZED BODIES

The solar wind interaction with weakly magnetized bodies exhibits a wide range of individuality and richness in its physical processes. Moreover, one cannot generally separate the aeronomy or—sometimes—the geology of these bodies from the solar wind interaction. Each member of this class, including Venus, Mars, the Moon, Pluto, asteroids, comets, and interplanetary dust, constitutes a different realization of the parameter space of space-plasma interactions.

We are fortunate that a few spacecraft missions have been dedicated to the detailed exploration of the plasma interactions at these bodies. Without the comprehensive earlier views provided by Explorer 35 and the Apollo 15 and 16 subsatellites at the Moon, the Pioneer Venus Orbiter (PVO), and the cometary missions ICE, Suisei and Sakegake, VEGA 1 and 2, and Giotto, our ability to infer what is happening at Mars, Titan, and Pluto from limited new observations and to plan future exploration efforts would be greatly compromised. Nor would we have appreciated, except by conjecture, the comparisons and contrasts to Earth's solar wind interaction—such as the effects the solar wind interaction may have had on the evolution of the terrestrial planet atmospheres. These alternative worlds give us insight into the breadth of plasma interactions that must be present at the still unexplored solar system bodies and in any extrasolar planetary systems harboring terrestrial planets. They also provide insight into the mechanisms of interaction between the solar wind and the terrestrial magnetosphere and ionosphere. Some of these mechanisms are universal and operate in magnetospheres of quite different characteristics. Others are specific to particular objects. Hence both the similarities and the dissimilarities between different magnetospheres can provide deeper insight into the operation of such mechanisms.

The basic parameters determining the features of the plasma interactions of weakly magnetized planets are the presence or absence of a substantial atmosphere and ionosphere; the presence or absence of significant remanent magnetization, or magnetic fields induced in the body by the plasma interaction; the nature and properties of the incident magnetized plasma flow; and the size of the body relative to the size of the incident particle gyroradii. Here the panel provides a brief summary of current knowledge, outstanding questions, plans, and recommendations for each of the weakly magnetized bodies.

Venus

Achievements

Venus exhibits a foreshock that has the general characteristics of the foreshock of Earth, scaled for Venus. The bow shock itself is the shape and scale expected for the solar wind conditions at Venus and an obstacle slightly larger in cross section than the solid body since the effective obstacle is the ionosphere. The bow shock has an elliptical cross section, with an average terminator radius of about 2.4 planetary radii. The relationship between the bow shock radius and the nature of the obstacle became an issue when it was found that the shock moved inward (by ~0.3 planetary radii at the terminator plane) at solar minimum. That amount of change could not be wholly explained by the apparent changes in the Venus ionosphere over the solar cycle.

Pioneer Venus Orbiter observations showed that the ionosphere's upper boundary ranges from average altitudes of ~300 km subsolar to ~800–1,000 km at large solar zenith angles around solar maximum, a height and shape determined by pressure balance between the normal component of the incident solar wind pressure and the thermal pressure of the ionospheric plasma (see Box 2.8). In the inner magnetosheath, at the boundary with the ionosphere, the initially dynamic solar wind pressure is transformed to magnetic pressure in a layer of enhanced magnetic field and depleted solar wind plasma known as the magnetic barrier. The name ionopause is appropriate for the ionosphere boundary because ions created from the atmosphere at higher altitudes are picked up and removed by the solar wind. The ionopause moves up and down with changing solar wind pressure, reaching a minimum altitude of ~225 km, where it extends into the photochemically dominated region of the ionosphere. The ionopause thickness also varies with altitude, exhibiting a local oxygen ion gyroradius scale of tens of kilometers when the ionopause is high in the collisionless region, but a thicker, more structured and diffuse character at low altitudes, where collisional diffusion and ionospheric convection determine the force balance in the boundary layer. This is in effect an extension of the Venus magnetosheath into the ionosphere (see the sister report by the Panel on Atmosphere-Ionosphere-Magnetosphere Interactions), a situation unique to the weakly magnetized planets.

The magnetosheath of Venus is well described by magnetized fluid models of flow interaction with a blunt obstacle or conducting sphere. In fact, because the interaction system is large compared with solar wind plasma kinetic scales but small compared with the scale of solar wind variations, because the ionospheric obstacle is less comprehensible than that of a magneto-

BOX 2.8 PIONEER VENUS MISSION

The Pioneer Venus Orbiter (PVO) mission is responsible for most of our current knowledge of the Venus–solar wind interaction. This spacecraft operated in orbit for ~14 years (1978–1992), covering both solar maximum and solar minimum conditions at altitudes suitable for observing the foreshock, bow shock, magnetosheath, and magnetotail. The upper atmosphere and ionosphere were regularly sampled with in situ instrumentation only during the early high solar activity phase of the mission owing to PVO's orbit evolution. PVO space physics-related instruments included a magnetometer, solar wind plasma analyzer, thermal ion mass spectrometer, retarding potential analyzer, Langmuir probe, and plasma wave antenna. Local solar EUV fluxes were derived from the Langmuir probe photoemission. The combination of the mission duration and comprehensive payload—together with a substantial period of well-supported data analysis, including a healthy guest investigator program—resulted in a revolutionary view of the solar wind interaction with an obstacle that contrasts sharply with the Earth because of its (then confirmed) almost complete lack of a planetary magnetic field of either internal dynamo or remanent nature. PVO demonstrated how an unmagnetized ionosphere was formed and transported and how it interacted with the solar wind. Many unexpected phenomena were discovered such as magnetic ropes in the otherwise magnetic-field-free ionosphere.

sphere, and because there are no boundary effects from reconnecting planetary and external magnetic fields, Venus may provide the most ideal example of a magnetosheath in our solar system. Complications that have been observed include (1) occasionally turbulent magnetosheath character when the IMF orientation places the quasi-parallel bow shock, with its associated waves and turbulence, near the nose of the bow shock and the stagnation streamline and (2) possible effects of planetary ion production in the ionopause vicinity and throughout the upper atmosphere, which extends well into the dayside magnetosheath and solar wind.

The subject of cometlike ion production in the solar wind interaction region, reinvigorated by PVO observations of the wake of escaping planetary O^+ pickup ions, has proven particularly important for present and future space physics investigations at the weakly magnetized planets. As at comets, the induced magnetotail of Venus, made up of highly draped magnetosheath flux tubes that sink into the wake created by solar wind flow divergence around the dayside ionosphere, rotates with the IMF from which it is largely constructed. In addition to its role in induced magnetotail geometry, the so-called mass loading of the solar wind plasma by planetary ions is suspected of having both large-scale effects on the magnetosheath flow and field and small-scale effects such as the production of plasma and MHD waves. What has attracted most interest, however, is the appreciation that solar wind erosion of the planetary atmosphere has had long-term, evolutionary effects.

The mainly atomic oxygen pickup ions at Venus have large gyroradii (~1 planetary radius) relative to Venus, with the consequence that their trajectories intersect the exobase (~200 km altitude), where they interact collisionally with the atmospheric gas. This interaction leads to both energy deposition and sputtering. The patchy, weak UV (130.4 and 135.6 nm) aurora observed on the nightside of Venus could be related to planetary pickup ion precipitation, but there are also other candidate explanations (see the report of the Panel on Atmosphere-Ionosphere-Magnetosphere Interactions). Model extrapolations of this process into the past indicate that the combination of ion pickup and sputtering-induced escape, together with photochemical escape, has contributed to losses of atmospheric constituents like oxygen over time. In a striking recent observation that in some ways confirms the historical view that Venus is in effect a comet, the Ulysses spacecraft detected the Venus pickup ion trail at a distance of over 4×10^7 km (~0.3 AU). These processes at Venus also have direct analogies at Mars.

Outstanding Questions

The Venus solar wind interaction was so well characterized by a single, well-instrumented mission, PVO, that Venus has not been a high-priority target for space physics measurements for many years. However, PVO raised, and left unanswered, a number of key questions about the Venus solar wind interaction.

What Is the Rate of Ion Pickup from Venus?

No PVO instrument was designed for pickup ion composition measurements over a sufficiently broad energy range. Pickup ions are originally neutral atoms and molecules that become ionized in the flowing plasma surrounding the planetary obstacle. These ions generally represent a significant loss of atmosphere integrated over the age of the solar system. Their energy ranges up to four times that of the solar wind protons times the mass of the ion relative to the proton mass. While some low-energy pickup ions may have been detected by the thermal ion mass spectrometer near the ionopause, and the solar wind plasma analyzer inferred the existence of O^+ up to ~8 keV, picked-up O^+ energies are expected to range from ~0 to 60 keV. Thus the extent of the pickup ion population at Venus, together with its composition and variations with solar and solar wind conditions, remains undetermined.

How Much Sputtering Occurs from the Venus Atmosphere?

The sputtering mechanism, whereby neutral atoms are knocked out of the atmosphere by energetic pickup and solar wind ions, has been postulated to exist only through modeling and has not been observed. It may be an important atmospheric loss mechanism. A survey during a range of solar activity conditions with an energetic (~100 eV to 100 keV) ion mass spectrometer, complemented by solar wind and IMF measurements, is necessary to fill the gaps in our knowledge of the direct solar wind erosion of the atmosphere of Venus. A low-energy, highly sensitive neutral particle detector or UV spectrometer measurements could resolve the issue of the presence of and importance of the sputtering mechanism. Complementary in situ measurements of the ionosphere below would significantly enhance the depth of interpretation by providing the information needed to model the photochemically produced exosphere that both seeds the pickup ion population and feeds photochemical escape. At this writing no Venus missions carrying such space physics measurements are under development, although missions are now being planned in Japan and Europe.

Mars

Achievements

Mars exploration in general has suffered from mission mishaps, including the recent losses of Phobos 1 (with the mission of Phobos 2 not being fully realized), Mars 96, Mars Observer, Mars Climate Orbiter, and Mars Polar Lander, although the latter carried no space-physics-related instruments. Most of our current knowledge of the Mars solar wind interaction comes from the Soviet Phobos 2 (see Box 2.9) and our own Mars Global Surveyor (MGS). The interpretation of the observations obtained on these missions has been enhanced by observations of the upper atmosphere and ionosphere, by the Mariner 9 and Viking missions, by data from the Soviet Mars mission series, and by Mariner 4 observations of the close-in location of the bow shock and magnetosheath with magnetometers and plasma analyzers. These latter observations from Mariner 4 and Soviet Phobos 2 data were sufficient to establish the weakness of any global magnetic field and magnetosphere of Mars.

Despite its limited period of operation, Phobos 2 provided a detailed and accurate picture of the solar wind interaction with this weakly magnetized planet and its ion loss processes. In addition, several hints at the existence of the crustal magnetic fields of Mars and their effects on the solar wind interaction in the pre-MGS era can be identified in retrospect. The martian obstacle seemed wider than its Venus counterpart, suggesting contributions to the obstacle pressure other than those present at Venus. The magnetotail boundary and bow shock also exhibited greater variability in their positions, although with the scale of the Mars subsolar magnetosheath approaching that of a solar wind proton gyroradius, kinetic effects on the solar wind interaction could not be ruled out as the cause.

Although MGS is primarily a remote-sensing, surface-mapping mission, the magnetometer and electron detectors it carries provided key solar wind interaction information that was improved by late changes in the early mission plan. A series of very low periapsis (~110 km), highly elliptical aerobraking orbits during which data were recorded revealed the presence of strong, localized crustal remanent fields strong enough to be significant at the inferred martian obstacle boundary (see Box 2.10). The uneven distribution of the crustal fields, with the strongest features in half the Southern Hemisphere, in the oldest exposed terrain is the likely explanation for much of the observed variability of the tail boundary and bow shock positions. Moreover, the low-altitude, dayside magnetosheath interface with the ionosphere is complicated by the remanent fields, which produce localized magnetospheric cusp or cleftlike features, and a hybrid magneto/ionopause that changes with the orientation of the interplanetary magnetic field, as well as with incident solar wind pressure and solar EUV flux. The MGS electron data confirm that the daytime upper ionosphere has a lumpy boundary that reflects the crustal field distribution, reaching higher altitudes over the regions where the field is strong enough to deflect the solar wind above it. Thus, Mars is a lumpy obstacle to the solar wind, and its interaction with the solar wind is probably the most complicated of all of the solar wind interactions sampled to date.

BOX 2.9 THE PHOBOS 2 MISSION

Phobos 2 carried several ion spectrometers, an electron analyzer, and a magnetometer through several highly elliptical transfer orbits that sampled the subsolar magnetosheath to ~850 km altitude. The ability of the ion spectrometers to distinguish planetary oxygen ions from solar wind protons made it possible to detect a transition in ion composition in the inner martian magnetosheath, which is thought to resemble the Venus situation at solar minimum. There were also a number of inferred “boundaries,” whose probable connection with the later-discovered crustal magnetic fields was not apparent at the time.

In its final, nearly circular orbit at ~2.7 Mars radii, data were obtained that led to better understanding of the bow shock positions near the terminator and the induced character of the martian magnetotail. As at Venus, ions of apparent planetary origin, including O^+ and other products from a largely CO_2 atmosphere, were observed escaping in the planet’s wake. Mars’s foreshock also included waves consistent with the presence of a population of picked-up planetary protons. Whereas Mars has an oxygen exosphere or corona analogous to that at Venus, unlike Venus its primary exospheric constituent is hydrogen, detected in Lyman-alpha emission during the early Mariner missions.

BOX 2.10 THE MGS MISSION DISCOVERS INTENSELY MAGNETIZED CRUST

The Mars Global Surveyor (Figure 2.10.1) was the first spacecraft to enter the martian crustal field magnetosphere. The Mars–solar wind interaction resembles a cross between the solar wind interaction with the lunar crustal magnetic fields and the Venus–solar wind interaction, which is purely ionospheric. The study of this solar wind interaction is tremendously complicated by the fact that the oncoming solar wind encounters different crustal field configurations as Mars rotates, as well as by the usual variations due to the changing solar wind and interplanetary magnetic field (with which the crustal fields reconnect).

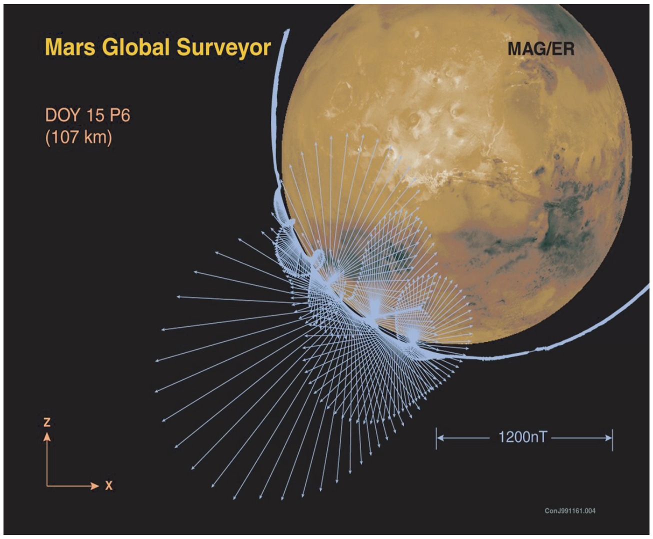


FIGURE 2.10.1 The orientation and magnitude of the measured magnetic field on an MGS pass through the martian crustal field magnetosphere. Courtesy of M. Acuña, NASA Goddard Space Flight Center.

Outstanding Questions

How Does the Solar Wind Interact with a Body Having an Atmosphere and an Ionosphere in the Presence of a Strong, Patchy Remanent Magnetic Field?

Different scenarios for the Mars–solar wind interaction must exist for each combination of solar wind and interplanetary field conditions and for each longitude of noon Mars local time. It is as if a different, complex magneto/ionospheric interaction exists for each. Without sophisticated models of the solar wind interaction, it will be extremely difficult to interpret the data, no matter how comprehensive. The Mars–solar wind interaction thus demands a new mission paradigm where modeling of the observations is not a discretionary choice but a necessity for obtaining results. It is also important in Mars studies to be able to use the models, together with the present-day observations, to infer the role the ancient magnetic field of Mars may have had on limiting Venus-like atmosphere escape by ion pickup and sputtering.

The ISAS Nozomi mission, on an extended cruise to Mars, where it will arrive in 2004, is expected to carry out an enhanced PVO-like solar wind interaction and aeronomy survey. Nozomi is equipped to undertake the needed observational reconaissance with its full complement of plasmas, fields, and remote sensing instruments. Like PVO, the measurements will cover the upper neutral and ionized atmosphere to ~150 to 200 km altitude, together with plasma, plasma wave, and magnetic field measurements throughout its elliptical orbit (up to its ~4 Mars radii apoapsis). While the lower inclination of Nozomi's orbit will limit its in situ perspective to the low-latitude solar wind interaction, the detailed measurements will give an unprecedented picture of the subsolar obstacle boundary. NASA's contribution to Nozomi of an ion-neutral mass spectrometer ensures at least some level of involvement by U.S. scientists in the analysis and interpretation of the Nozomi data.

ESA's Mars Express mission, launched in June 2003, is expected to arrive at Mars about the same time as Nozomi goes into orbit. The opportunity for collaborative efforts between the two missions has been made part of their science planning. This is especially fortunate because Mars Express includes an ENA detector and ion and electron analyzers to study the solar wind interaction through a combination of in situ measurements and remote sensing, but does not include supporting magnetic field measurements. On the other hand, Mars Express is in a high inclination orbit that will provide in situ ion and electron measurements to complement those from Nozomi's low-latitude sampling, including measurements of heavy pickup ions, which are expected to have asymmetric spatial distributions due to their very large gyroradii compared with the radius of Mars. NASA has contributed the electron analyzer and part of the ion analyzer system for Mars Express through the Discovery mission of opportunity option. Thus U.S. investigators will also be able to participate in the interpretation of Mars Express solar wind interaction data.

Together, Mars Express and Nozomi measurements can bring the state of knowledge of the Mars solar wind interaction to the level of, and in some respects beyond, our current knowledge of the Venus solar wind interaction. However, an important caveat is that the Mars solar wind interaction is much more complicated, as mentioned above. Beyond these missions, Mars Odyssey, which arrived in late 2001, includes a radiation environment monitor to measure the energetic particle and photon fluxes that are potentially hazardous to humans. However, this experiment is not designed to investigate which physical processes determine the radiation environment at Mars. Except through supporting measurements that may come from the MGS, Nozomi, and Mars Express instruments, the Odyssey experiment will not allow identification of the conditions (e.g., flares, coronal mass ejections, some unforeseen local effect) that give rise to radiation enhancements. Indeed, the question of what space weather is like around Mars has implications for more than just human exploration. The early Sun may have been more like today's active Sun, with the result that many more energetic particles were present in the solar wind than typically are today. The effects of such fluxes on the martian atmosphere and surface have not been determined. This area thus remains open to further enquiry.

What Is the Nature of Mars's Patchy Magnetic Field?

Opportunities are needed to better map the crustal magnetism of Mars and to send compact but well-conceived space environment packages on Mars orbiters in the Mars Surveyor and Mars Scout programs to reveal more about the complexity of the Mars obstacle and Mars space weather.

What Is the Particles and Fields Environment of the Surface and Lower Atmosphere, and How Does It Vary with the Solar Cycle?

As we learned from PVO at Venus, the solar activity cycle is an important factor in determining the solar wind interaction characteristics by virtue of its effects on both the ionospheric obstacle and the interplanetary conditions. In addition, landed magnetometers, electric field and radiation detectors, and photometers can reveal the extent to which the lower atmosphere and solid body of Mars are engaged in (or respond to) the solar wind interaction.

Moon

Achievements

The lunar solar wind interaction was explored to a first level of overall understanding during the Explorer 35 and Apollo 15 and 16 subsatellite missions. Those observations consisted of basic magnetic field and plasma measurements, together with suprathermal electron measurements that could be used to diagnose the near-surface field via magnetic reflection signatures in their pitch-angle distributions. Several landed solar wind experiments were also deployed during the Apollo years. As expected for a roughly spherical, solid rock body largely devoid of significant atmosphere, the Moon's solar wind interaction signature was generally that of an insulating absorber. This absorption left a wake in the solar wind with an only minimally perturbed interplanetary magnetic field but few solar wind particles in the immediate vicinity of the Moon. However, strong, localized crustal remanent fields, weaker than but akin to those at Mars, were observed to deflect the solar wind away from parts of the lunar surface, especially near the limb of the plasma interaction. The Moon thus represents another distinctive class of obstacle in the spectrum of solar wind interactions—Mars-like in some respects, but without a significant atmosphere.

Most recently, the Wind spacecraft, which serves as both a solar wind monitor for terrestrial magnetosphere interactions and a magnetotail probe, sampled the lunar wake at a distance of ~6.5 lunar radii with sophisticated ion and electron analyzers, an electric field instrument, and a magnetometer. This combination of measurements provided detailed diagnostics of the ion and electron distribution functions, highlighting their differences and kinetic aspects of their interaction with the Moon. In particular, the observations showed that the solar wind ion wake is approximately coincident with the optical shadow cylinder, while the faster, smaller gyroradius electrons are absorbed on the interplanetary flux tube penetrating the Moon. In the ion wake, charge-separation electric fields acting to refill the wake add to the production of anisotropic particle distributions that produce a variety of plasma waves. The Wind lunar observations clearly demonstrated the value of sophisticated plasma and field instrumentation in diagnosing the physics of solar wind interactions.

The Lunar Prospector mission provided close-in (~100 km altitude) detections of the solar wind and suprathermal electron behavior in the vicinity of the Moon's remanent magnetic field concentrations, together with measurements of the field itself. These data revealed that some of the perturbations in the lunar solar wind interaction, inferred earlier from limb features, resembled miniature magnetospheres abutting the lunar surface.

Outstanding Questions

How Does the Solar Wind Interact with the Small-scale Field on the Lunar Surface?

We understand little about the solar wind interaction with the small-scale fields of the Moon. It is important that we do achieve this understanding, for a similar interaction occurs at Mars and at asteroids. It is surprising that these regions act as mini-magnetospheres given that the scale size of the magnetic features is similar to the solar wind ion gyroradius. The scale of the solar wind proton gyroradius relative to the size of some of these features would seem to prohibit anything resembling the fluidlike interactions at their global magnetospheric counterparts. The variety of these local interactions is daunting, as each small magnetosphere is determined by the strength and orientation of the crustal feature, the feature's position with respect to the subsolar point, and the conditions in the solar wind, a situation not unlike that at Mars. In all the lunar plasma interaction cases, there is the additional variety provided by the different incident flow types in Earth's foreshock, the magnetosheath, and occasionally in the disturbed magnetosphere.

What Is the Nature of the Lunar Wake?

The Moon is within striking distance of Earth-orbiting spacecraft, as well as spacecraft heading off on interplanetary trajectories. Despite this accessibility we do not understand well the rather straightforward creation of the wake downstream of an atmosphereless, unmagnetized body. Given the absence of plans for further solar wind interaction exploration of the Moon, it is important to consider what opportunistic use could be made of other spacecraft with instrumentation for particles and fields. The Wind spacecraft has sufficient fuel to change its orbit periodically, for example, and possesses a proven instrument complement for probing the lunar wake. If it is simply a matter of timing transfers of the Wind orbit, possible lunar flybys deserve serious consideration in the planning. Similarly, the twin STEREO spacecraft will use the Moon for gravity assists to achieve their heliocentric orbits.

Pluto

Outstanding Question

What Is the Nature of the Solar Wind Interaction with Pluto?

We have no observations that pertain to the solar wind interaction with Pluto. Pluto is thought to be composed of an ice and rock mix more characteristic of comets than of asteroids or solid planetary bodies. Our experiences from exploring the solar wind interactions with these other obstacles, together with the assumptions about Pluto's size and extended atmosphere derived from remote sensing, suggest that the Pluto interaction may vary from cometlike to Venus-like to Moon-like, depending on a number of factors. A cometlike interaction is indicated if and when the hypothesized outflow of ~10^{27} particles per second of atmospheric neutrals occurs, perhaps near Pluto's perihelion at ~30 AU. On the other hand, if the ionosphere of Pluto is at these times sufficient to balance the incident particle pressure, some Venus-like deflection of the solar wind may occur, albeit with at least ion kinetic effects coming into play given the Moon-like size of Pluto and the relatively large solar wind proton gyroradius at this heliocentric distance. There is also expected to be a significant pressure contribution in the local solar wind from picked-up interstellar ions, including H^+, He^+, O^+, N^+, and C^+. These would supplement Pluto's own picked-up ion population from its extended atmosphere of molecular nitrogen, methane, and carbon monoxide (at least) in creating a diverse and highly anisotropic incident flow. As a result, Pluto's environment may be filled with plasma waves of various kinds, similar to the extended region of waves observed around the comets we have visited. Pluto must also be diminished by the constant cometlike erosion of its volatile content, even though its atmosphere will continue to be replenished for some time by sublimation, outgassing, and perhaps particle sputtering.

Pluto's orbital period of ~248 Earth years and its significantly elliptical orbit, with perihelion at ~30 AU and aphelion at ~40 AU, are expected to combine with the ~11-year solar cycle variations to modify the atmosphere, ionosphere, and thus solar wind interaction on long time scales. On shorter time scales, the interaction must be affected by the ~100 percent variations in solar wind density observed in the outer heliosphere by the Voyager and Pioneer spacecraft. In addition, the presence of some remanent magnetization of Pluto cannot be ruled out. If such magnetization were to exist, even at the ~10 nT surface level, it would also divert incident flows and add to the sensitivity of the Pluto obstacle the large excursions in solar wind dynamic pressure associated with the above density variations. In all cases, the presence of the satellite Charon, located at ~17 Pluto radii, is likely to contribute further to any asymmetries in the solar wind interaction, depending on its own nature as an obstacle as well as its effects on Pluto's atmospheric distribution.

NASA recently solicited and received proposals for a Discovery-style Pluto Express mission to carry out the first exploration. In spite of stringent constraints on mission cost and science payload, it is notable that particles and fields instruments were considered in several mission concepts. Because the likelihood of returning to Pluto in the decades to follow is so remote, the inclusion of space particles and fields measurements is essential to the completion of NASA's reconnaissance of solar system–solar wind interactions and their associated processes such as atmosphere escape. [See note, p. 123.]

Asteroids

Achievements

The solar wind interaction with asteroids represents a limiting case in the solar system. There is typically no significant atmosphere, with particle and/or photon sputtering and outgassing only weak sources. These solid rocky bodies, with their small (from tens to hundreds of kilometers) scales, provide almost no impediment to the solar wind flow. The existence of lunarlike wake features is unlikely given the object scale relative to the solar wind proton gyroradius. The generally nonspherical shapes and rotations of these bodies further prevent any orderly wake features from forming on a quasi-permanent basis.

A weak solar wind deflection by asteroids has been suggested by observations from Galileo's close flybys of Gaspra (radius ~14 km, closest approach ~1,600 km) and Ida (radius ~30 km, closest approach ~2,400 km) and Deep Space 1's close flyby of Braille. Perturbations in the magnetic field seen at the times of these flybys might be the signature of "whistler wings," formed by an obstacle with scale intermediate between the electron and proton gyroradius. However, the NEAR spacecraft magnetometer detected no apparent Eros-related perturbations at that body. Similar searches for signatures of solar wind interaction effects near the asteroid-like martian satellites Phobos and Deimos with the MGS magnetometer and electron detectors produced no clear evidence.

Outstanding Question

How Do Small Bodies Create Disturbances in the Solar Wind?

Three small asteroids—Gaspra, Braille, and possibly Ida—have been reported to cause magnetic perturbations in the solar wind flow. Our understanding of how these bodies create these disturbances, as in the case of the lunar magnetized regions, is poor. Further, it has been suggested that the asteroid Oljato produced a statistically significant set of signatures in the interplanetary field. In addition, the natural variability of the interplanetary magnetic field makes identifications of weak signatures from these obstacles difficult without significant supporting plasma measurements in specifically targeted regions such as the near-object wake.

The largest asteroids, like Ceres and Vesta, represent the best chance of observing a signature with limited space physics instrument capability. Ceres, nearly spherical and with a radius of ~480 km, and Vesta, with a radius of ~260 km, have recently been selected as targets of the Dawn Discovery mission, which will carry a magnetometer.

Kuiper Belt Objects

Achievements

The Kuiper Belt objects (KBOs) are probably Pluto-like or cometary nucleus-like bodies composed of ice, rock, and dust. Like Pluto, KBOs have not been observed in situ. Sixty KBOs have been detected since 1992,[2] though many others are expected to be present in a disk that extends through and beyond the outer solar system at ~30 to 50 AU. The Centaur family of KBOs includes those that cross the orbits of the outermost planets on their highly eccentric paths. It is frequently speculated that KBOs are leftover planetesimals from the early solar system, and that they are the reservoir for short-period comets. In this scenario, the KBOs become comets when their orbits are perturbed by occasional strong gravitational interactions with the outer planets. Other comets may come from the Oort cloud, which is more isotropically distributed around the solar system at much larger distances.

KBO sizes, derived from their brightness and assumptions of a comet-like surface character, are in the

---

[2] NRC. 1998. *Exploring the Trans-Neptunian Solar System*. National Academy Press, Washington, D.C.

asteroid size range, tens to hundreds of kilometers. An even smaller (~1 to 10 km) population is inferred from the short-period comets. Surveys suggest that there are up to a billion KBOs. Telescopic and spectroscopic observations of some of the Centaurs (e.g., Chiron) indicate there are short-lived comet-like outbursts of gases and dust from some KBOs due to processes such as sublimation.

Outstanding Question

How Do Kuiper Belt Objects Interact with the Solar Wind?

The plasma interactions with these bodies are expected to be as complex and changeable as the interaction at Pluto, depending, of course, on the plasma and field environment—which in the case of KBOs includes the outer solar system and possibly, on occasion, the region of the heliospheric termination shock—as well as on the occurrence of an outburst that produces a transient atmosphere. The interaction range is at its most extreme when the object is also on a highly eccentric orbit like the Centaurs.

While it is unlikely that a mission solely to the Kuiper Belt will be flown in the foreseeable future, planning for a flyby of a KBO has generally been a component of any Pluto-Charon mission plan. Solar wind interaction studies generally require a relatively close flyby, but ionization and pickup scale lengths are larger in the dim sunlight and weak magnetic field of the Kuiper Belt region.

Comets

Achievements

In the mid-1980s, the Soviet VEGA 1 and 2, ISAS Sakigake and Suisei, and ESA Giotto spacecraft were sent toward Comet Halley, destined to pass upstream and observe the solar wind interaction in detail, while imaging the cometary nucleus at perihelion. In the meantime, the ISEE-3 solar wind monitor of Earth's magnetospheric interaction was renamed the International Cometary Explorer (ICE) and diverted to encounter Comet Giacobini-Zinner near perihelion, on a trajectory that passed through the comet's wake. The data obtained by these missions provided a revolutionary view of comets and cometary processes at other bodies.

The comet-solar wind interaction appeared more or less as Alfvén had predicted decades before. The signature of a highly draped interplanetary magnetic field, produced by the near-stagnation of the solar wind by heavy cometary ion production in the outflowing extended cometary atmosphere, was clearly a major feature. This "mass loading" of the solar wind has counterparts in the inner magnetosheaths and magnetotails of Venus and Mars, and probably in those of Pluto. The picked-up planetary ions, moreover, extended over millions of kilometers. They were accompanied by a variety of wave populations generated by the highly anisotropic distribution functions of the picked-up ions. Indeed, the cometary environment provided an exceptional laboratory for pickup-ion-generated wave studies and wave particle interactions. Because of the large extent of the mass-loading regions, it was possible to observe the pickup ion distributions evolve from ringlike to shell-like as the particles were scattered by the waves in their path. An extended cloud of high-energy ions was also observed, evidently resulting from stochastic acceleration of a small number of the particles caught up in the cometary wave field.

Giotto came to within ~600 km upstream of Halley's nucleus, observing the pileup of the IMF, including the convected rotations of the magnetic field that Giotto had encountered earlier in the undisturbed solar wind. This close approach allowed sampling of the detailed features of the inner boundaries of a strong comet. A magnetic cavity was found surrounding the nucleus at ~4,500 km. Because the measurements were fairly comprehensive, it was possible to determine that this feature was not a Venus-like ionopause, but rather a boundary where the incident pressure was balanced by ion-neutral friction in the outflowing atmosphere. Apparent stagnation outside this boundary reinforced the early image of a comet as a source in a surrounding flow. Interestingly, the ion density profile remained smooth throughout this interface, being photochemically controlled. Because these observations occurred when Halley's atmosphere production rate was strong, at ~10^{30} particles per second, these observations are of course specific to a strongly outgassing comet at perihelion. The contrast with Giacobini-Zinner proved extremely valuable in this regard, as it had a perihelion gas production rate ~100 times less.

ICE flew through the wake of Giacobini-Zinner (G-Z) at a distance from the nucleus of ~7,800 km. Although both the G-Z and Halley encounters were characterized by large regions of cometary ions and their associated waves, the G-Z wake flyby was distinguished by the detection of a narrow plasma sheet (thickness ~2,000 km) of cold dense plasma in the center of the draped magnetotail field. ICE, VEGA 1 and 2, and Sakigake also had plasma wave instruments that allowed detection of

ELF (~10 to 2000 Hz) and VLF (~10^3 to 10^6 Hz) waves in addition to the ion cyclotron, mirror mode, and other low-frequency MHD waves observed with the magnetometers. These relate to different types of plasma instabilities that involve electron kinetics. Both whistler waves and probable lower hybrid waves were observed, often modulated by the larger-scale, lower-frequency waves from the pickup ion instabilities. The richness of the wave phenomena observed at the comets is still not fully appreciated.

Outstanding Question

How Does the Solar Wind Interaction with a Comet Vary with Heliocentric Distance?

In spite of the windfall of information obtained from the Halley and Giacobini-Zinner encounters described above, comets present a moving target in more than a dynamical sense. Their solar wind interactions must rapidly evolve as they approach the Sun, changing in the extremes from asteroid-like in the most distant reaches of their orbits, to Pluto-like, to something that can deflect the solar wind at an outflow boundary like Halley's, even after heavily mass-loading the oncoming plasma and field with its own atmospheric ions. Moreover, each comet has its own intrinsic character, including volatile content, shape and size, and structural detail. A recent discovery of soft x rays from comets suggested the possibility of using x rays as a new remote-sensing diagnostic of cometary interactions with the solar wind (see Box 2.11). The x rays are generated when charge exchange occurs between heavy solar wind ions and the cometary gases, leaving excited solar wind ions that radiate x-ray lines from a region of the coma sunward of the nucleus. However, the contributions of both the varying solar wind fluxes and the varying cometary gas production in producing the x rays makes it difficult to use this signature for detailed studies.

The future observations of the comet solar wind interaction largely rest with ESA's Rosetta mission. Rosetta is capable of a full complement of solar wind interaction measurements, including particles over a broad range of energy and composition, magnetic fields, and waves. Its mission is to follow the development of the comet Wirtanen from ~3.5 AU until perihelion, providing the first opportunity to observe the evolution of cometary features with heliocentric distance.

Dust

Achievements

The physics of solar wind and other plasma interactions with dust has numerous important consequences

BOX 2.11 COMETARY X RAYS

C.M. Lisse and co-workers found unexpected soft x-ray and extreme ultraviolet emissions (photons <2 keV) coming from Comet Hyakutake's sunward coma during its 1996 apparition. These emissions from a region ~100,000 km in extent, observed using ESA's ROSAT satellite, could not be easily explained by standard emission mechanisms. T.E. Cravens proposed that charge exchange between heavy (e.g., O, C, N) solar wind ions and the cometary neutrals could be responsible (Figure 2.11.1). The process leaves the participating highly charged solar wind ions in excited states that emit the energetic photons as they decay. This mechanism explains the observed morphology, intensity, spectrum, and temporal variability of the cometary x rays.

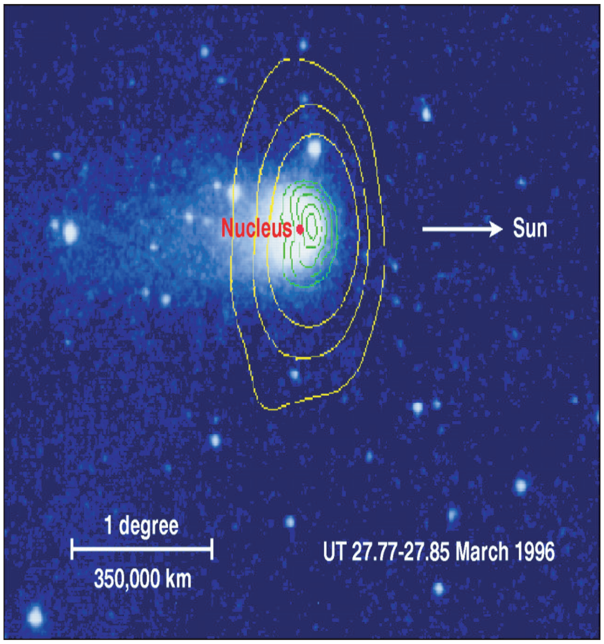


FIGURE 2.11.1 Isophotes of x-ray emission observed by ROSAT, superimposed on a visible light image of Comet Hyakutake, 1996 B2. Courtesy of T.E. Cravens, University of Kansas.

but is often less appreciated than it should be because of its exceptionally cross-disciplinary nature. Dusty plasmas are found near the surfaces of the Moon and asteroids; in cometary comas and tails; in the ionospheres of Earth and the planets; in planetary magnetospheres, where satellite volcanism and the sputtering of satellites and rings provide sources; and in interplanetary space, where both cometary and interstellar dust contribute to the zodiacal dust cloud. The plasma-dust interaction can take many forms and thus has many different consequences. Dust that becomes charged because of the effort of a grain exposed to plasma and solar UV to balance the currents to and from its surface (by incident particle fluxes countered by photoemission, for example) can act like a highly charged, superheavy ion component, mass loading the plasma, contributing significant additional gravitational and radiation pressure to the force balance in the plasma and producing a new class of possible waves and instabilities. This charged dust affects the appearance of comet tails, creates structures like spokes in planetary rings and levitated dust clouds on the Moon, and is accelerated in the outer planet magnetospheres by the co-rotation electric field to produce an escaping dusty planetary wind. It has been suggested that charging played a key role in grain adhesion and growth in the earliest stages of the planetary accretion process. It has also been suggested that interplanetary dust absorbs and re-emits solar wind ions as a singly charged population, adding to the heliospheric pickup ion sources. Together with the pickup ions from the ionization of interstellar gas, these ions are expected to provide the seed population for the acceleration of anomalous cosmic rays in the outer heliosphere.

Outstanding Question

How Do Dusty Plasmas Behave?

Dust detectors are now an accepted part of many planetary and interplanetary missions. The Galileo, Ulysses, Cassini, Rosetta, and Stardust missions all include them. Furthermore, charged dust is thought to be responsible for phenomena seen in Saturn's rings. Despite the importance of dusty plasmas and the great international interest in them, very little research has taken place in the United States. Given the broad applications and potential of a cross-disciplinary initiative on dust in planetary systems to provide important insights in space physics, it would be worth mounting such an initiative in the form of targeted investments by research and analysis programs across space physics, planetary science, and origins of planetary systems disciplines, including the development of dust detection technologies, laboratory research on dust-plasma interactions, remote sensing of dust in the solar system, the interaction of the solar wind with dust stream plasmas, and the role of grain size. It is important to learn the composition, distribution, variability, and charge state of dust in different solar system contexts and the roles that dusty plasmas and plasma-dust interactions play in solar system physics.

Epilogue on Solar Wind Interactions with Weakly Magnetized Bodies

From the above descriptions of past, current, and planned efforts, it is clear that the United States has ceded to Europe and Japan its previous leadership role in space missions dedicated to the solar wind interactions with weakly magnetized bodies. This was apparently a NASA management and advisory committee decision. Because there are scientifically important observations to be made (as in the examples described above) that the U.S. scientific community can contribute to and gain from, it is recommended that NASA seek and support collaborative ventures that both exploit the opportunities offered by our international partners in these endeavors and keep U.S. scientists and engineers involved in the solar wind interactions research at Venus, Mars, and some comets. It is also recommended that compact particles and fields packages be developed for inclusion on the smaller planetary missions now in vogue. While these missions will never be able to match a PVO in terms of gaining comprehensive knowledge, they open the door to important new insights and discoveries such as those obtained on MGS with the magnetometer and electron reflectometer. We must aggressively take advantage of the current preference for limited missions or forever lose our place in this fruitful area of space exploration.

OUTER PLANETS

The detection of strong radio signals from Jupiter quickly led to the realization that Jupiter had a strong intrinsic magnetic field and an extensive magnetosphere and radiation belt. Similar signals were not seen from Saturn, Uranus, and Neptune. In the 1970s the exploration of the outer solar system began with the launches of Pioneer 10, Pioneer 11, and then Voyagers 1 and 2. These spacecraft confirmed the existence of a strong jovian magnetic field and intense radiation belt and then revealed what ground observations could not, that Saturn, Uranus, and Neptune also had intrinsic magnetic fields and vast magnetic envelopes. In the following sections, we describe some of the achievements in the study of these magnetospheres and outline the outstanding questions.

Jupiter

Achievements

The first pass of Pioneer 10 through the jovian magnetosphere, whose noon-midnight cross section is shown in Figure 2.7, revealed a magnetosphere quite unlike that of Earth. The innermost part of the magnetosphere, like Earth's, was dominated by the strong intrinsic magnetic field, but the radiation levels far exceeded those of Earth, limiting the time that spacecraft could survive in that environment. Further out but inside the orbit of Io, a cold dense plasma (the "cold torus") was found, and outside the orbit of Io a hot torus was found. Both tori were derived from the interaction of Io's volcanic gases with the magnetosphere, but it was unclear how either torus was produced and why they differed so greatly in temperature. The torus plasma is accelerated to corotate with Jupiter by currents linking the torus to the ionosphere. This removes angular momentum and energy from the planet and transfers it to the plasma. The centrifugal force of the plasma pushing outward against the field then drives a giant circulation system that eventually takes the mass-laden flux tubes to the tail, where their ions can be released from the field lines by reconnection and the emptied flux tube can return to the neighborhood of Io. Jupiter provides our strongest example of a centrifugally driven magnetosphere and is the closest prototype we have to an astrophysical system that sheds angular momentum into its surroundings.

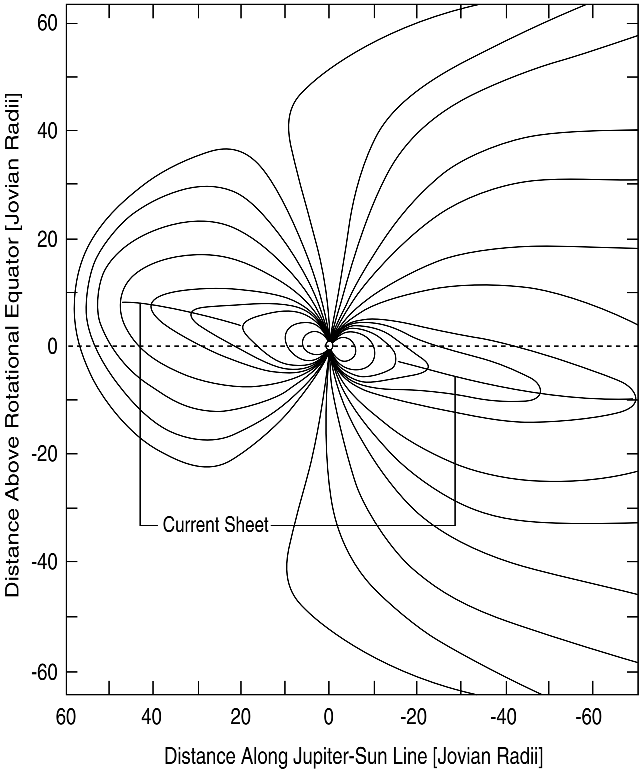


FIGURE 2.7 The noon-midnight cross section of the jovian magnetosphere illustrating the stretched out fields in the magnetodisk. Courtesy of C.T. Russell, University of California, Los Angeles.

Despite the addition of vast amounts of plasma into the magnetosphere by Io, the magnetic field of Jupiter is found to dominate the plasma forces until almost the orbit of Callisto near 25 R_J, where the magnetic field suddenly becomes stretched out in the equatorial plane dominated by the centrifugal force of the rapidly rotating cold plasma. This magnetodisk extends out to about 50 R_J in the dayside and down into the tail at night. A magnetic cushion region separates the magnetodisk from the magnetopause. Beyond the magnetopause, like at Earth, a magnetosheath, bounded on the sunward side by a strong bow shock, carries the shocked solar wind around the magnetospheric obstacle. A single pass of Ulysses in 1992 explored the higher latitudes of the magnetosphere and confirmed the picture built upon the Pioneer and Voyager low-latitude passes.

In 1995 a communication-crippled Galileo spacecraft proved the value of an orbiter for magnetospheric studies. Repeated passes of the spacecraft past Io revealed the extent of the mass-loading region. The various processes contributing to the dynamics of the magnetosphere were revealed, such as temporal variations in the volcanism on Io, reconnection in the magnetotail, and solar wind pressure variations. Optical measurements with the Hubble Space Telescope and other 1-AU observations and imaging with Galileo, as well as with the Cassini spacecraft as it flew by on the way to Saturn, show a dynamic auroral zone over both polar caps. Later the Galileo orbiter revealed the very interesting magnetosphere-within-a-magnetosphere of Jupiter's moon Ganymede (Box 2.12). Plasma circulation, driven by both internal and external processes, appears to be capable of causing the low-altitude acceleration processes leading to auroral forms; these are seen both in

BOX 2.12 GALILEO DISCOVERS GANYMEDE 'S INTRINSIC MAGNETIC FIELD AND MAGNETOSPHERE

Ganymede, one of the Galilean moons of Jupiter, is remarkable in many ways. An ice-covered body, it is the largest known moon in the solar system. In 1996, the Galileo Orbiter discovered that Ganymede has a permanent internal magnetic field large enough to form a magnetosphere (Figure 2.12.1). The magnetosphere shields the Moon from direct interaction with the flowing plasma of the jovian magnetosphere within which it is embedded. The discovery of an intrinsic magnetic field in a small planetary body that was thought to have solidified fully over its geological history is a surprise that has led to substantial rethinking of our ideas of planetary evolution. The small magnetosphere has been well characterized in multiple passes, at altitudes between 200 and 3,000 km. Analogies with and differences from the terrestrial magnetosphere have enabled us to test and extend our theories of magnetospheric processes.

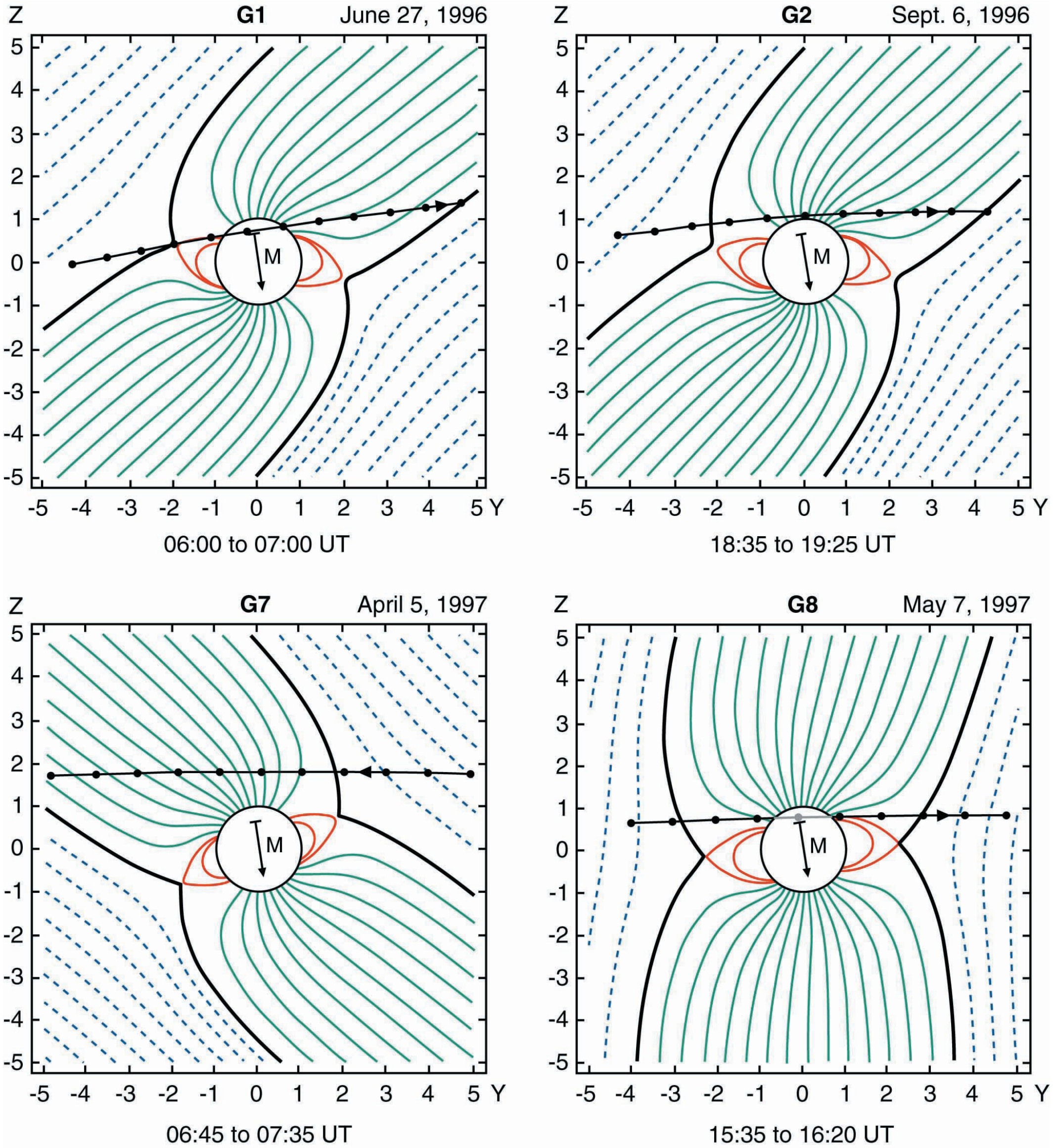


FIGURE 2.12.1 Cuts through the prime meridian of Ganymede representing the magnetic field and Galileo trajectories in the vicinity of the moon at the times of the first four close passes with different orientations of the magnetospheric field of Jupiter at the position of Ganymede. The heavy curves separate magnetic field lines that link to Ganymede at one or two ends from field lines that link only to Jupiter. These curves provide good estimates of the location of the magnetopause boundary. Courtesy of M.G. Kivelson, University of California, Los Angeles.

the closed magnetospheric regions and in the open tail regions.

Outstanding Questions

What Is the Relative Importance of the Solar Wind Interaction and Io-Powered Plasma Transport in Generating Magnetospheric Disturbances?

The observations in the jovian magnetosphere have seldom been obtained with simultaneous solar wind data outside the magnetosphere, and most of the measurements have been limited to the equatorial plane. This has left unresolved the important question of the relative importance of the solar wind interaction and Io-powered plasma transport.

What Is Responsible for the Observed Structure and Dynamics in the Plasma and Particles Deep in the Jovian Magnetosphere?

The formation of the cold torus is still poorly understood. The relative importance of scattering losses and radial transport from the hot torus and middle magnetosphere is not known. In addition, the dynamics of the radiation belts and their long-term and short-term variations are still mysterious.

How Does the Magnetic Flux Return to the Inner Magnetosphere from the Magnetotail?

The Io interaction loads magnetic flux tubes with heavy ions, and these tubes gradually spiral outward, eventually reaching the near-tail region. There, reconnection appears to produce magnetic islands of dense plasma that move downtail, leaving closed magnetic field lines connected to Jupiter and nearly devoid of plasma. It is not understood how these depleted flux tubes return to the region of Io against the outward flow of loaded tubes.

Saturn

Achievements

Saturn has been visited by three spacecraft, Pioneer 11 and Voyagers 1 and 2. At first glance these spacecraft reveal a magnetosphere much like that of Earth. While the dipole magnetic field of the planet is not tilted much with respect to the rotation axis, the rotation axis itself has a substantial inclination to its orbital plane, much as has Earth's dipole axis. The magnetic field at the surface of Saturn is similar to but slightly less than the terrestrial surface field. The size of the Saturn magnetosphere is huge, a third as large as that of Jupiter. The presence of a well-developed planetary ring system is quite different from the Jupiter magnetosphere but the rings absorb the radiation belt particles in the inner magnetosphere so that the flux of radiation belt particles is not like that of Jupiter but more like the terrestrial flux. Also like Earth, there are strong radio waves with kilometer-long wavelengths that appear to be controlled by the solar wind. So, too, the plasma wave population greatly resembles that of Earth's magnetosphere.

An important distinction between the magnetospheres of Saturn and Earth is the presence in Saturn's magnetosphere of many rings and moons. While the innermost rings may mainly act to absorb the energetic particles diffusing inward to Saturn, the E ring and the moons can supply mass to the magnetosphere, albeit at a lower rate than in the jovian magnetosphere. Nevertheless, this mass addition, accelerated to corotation, may also drive the radial circulation in the magnetosphere much as Io's plasma source can drive the jovian radial circulation. This circulation may also contribute to the generation of the aurora, as plasma circulation does at Earth and Jupiter. Titan is a very special moon with a very dense atmosphere. While—like Io—it provides much mass to the saturnian magnetosphere, it does not play the same role in the energetics of the magnetosphere as Io does at Jupiter because it is much further out in a region of very weak magnetic field.

Outstanding Question

How Is the Saturnian Magnetosphere Powered?

While Pioneer 11 and the two Voyagers gave us a good overview of the magnetosphere, they did not provide the quantitative data needed to determine how processes work. We do not know the basic plasma circulation system at Saturn, the strength of the ring and moon mass sources for the magnetosphere, or the relative strengths of solar wind driving and mass loading driving of the circulation pattern. We know nothing about the substorm process at Saturn, a process found to play a critical role in both the terrestrial and jovian magnetospheres. Does mass loading or solar-wind-driven convection lead to saturnian aurora? How important are dust-plasma interactions in the rings and in the magnetosphere? There is much left to resolve in the Saturn magnetosphere, but the well-instrumented Cassini orbiter on its 4-year-plus mission at Saturn should be able to resolve many of these outstanding issues.

Uranus and Neptune

Achievements

Voyager 2 flew by Uranus in 1986 and by Neptune in 1989, finding large magnetospheres supported by magnetic fields with high harmonic content. The dipole field was tilted at a large angle to the rotation axis for both planets. The unusual nature of these magnetospheres makes them ideal laboratories for understanding magnetospheric processes by comparing their behavior with that of the terrestrial magnetosphere. Overall, the Voyager 2 passes provided only a rudimentary picture of the sources of plasma, the presence of auroras, and the nature of plasma waves at these two planets.

Outstanding Question

How Do the Magnetospheres of Uranus and Neptune Compare with Those of Earth, Jupiter, and Saturn?

Our knowledge of the behavior of these two magnetospheres is at a very primitive stage. We know nothing about their dynamics, the relative importance of the solar wind interaction versus internal sources of plasma in these dynamics, or how the large tilt of the intrinsic magnetic field affects the dynamics of the planetary magnetosphere. How does the tilt of the dipole field affect the aurora? Do the moons and rings of these planets affect their magnetospheres? These very different magnetospheres could be very instrumental in teaching us about the terrestrial magnetosphere by contrasting the behavior of the various magnetospheric processes. Although no missions are in the planning stages for these two planets, there is much interest in exploring them further. The needed measurements could be made with a simple multiparameter particles and fields package, as proposed for other Discovery-class missions, such as those for Mercury and Pluto.

2.3 PROCESSES

INTRODUCTION

The anatomical investigation of the terrestrial magnetosphere is complete. We have probed all the important regions of the solar wind–magnetosphere interactions and we know their properties. We do not understand, however, how these regions work and interact. What are the processes that govern their behavior? In Section 2.2 the panel reviewed the outstanding questions surrounding these processes according to region within the terrestrial magnetosphere and at different bodies in the solar system and generated a perhaps daunting list. If instead we consider similar processes in different settings, we find much commonality. In general these processes can be grouped under three distinct themes: (1) the creation and annihilation of magnetic fields; (2) magnetospheres as shields and accelerators; and (3) magnetospheres as complex, coupled systems. By examining how the processes in each broad categorization interact and contribute to the overall system and by examining the processes acting under different plasma conditions at Earth and at the planets, we obtain a robust understanding of the process and of the system at a quantitative and predictive level. In this section the panel outlines how these three themes can unify and focus the recommended program. It emphasizes Earth as our most accessible magnetosphere, but certain phenomena, such as the rotationally driven magnetosphere, must be studied in a planetary setting. Moreover, all planetary data are important because the different conditions under which a particular process takes place invariably lead to a deeper understanding of the process as it occurs on Earth.

THE CREATION AND ANNIHILATION OF MAGNETIC FIELDS

The solar wind interaction with a planetary magnetosphere is an efficient generator of magnetic fields. This can be seen by imagining the situation without the solar wind. If the Sun suddenly went dark, stopped producing the light that warms Earth and ionizes its upper atmosphere and stopped producing the solar wind that confines Earth's magnetic field in the cavity we call the magnetosphere, Earth's field would become dipolar, the radiation belt particles would disappear, and the space around Earth would become a vacuum. All the external current systems would also disappear—those set up by the solar wind momentum flux that arrives at 1 AU and the electromagnetic energy that Earth captures. A major objective of space physics is to understand the creation of these magnetic fields and their causative currents. We understand some aspects of the currents well, but many not at all. When the interplanetary field is northward and there is no reconnection with the dayside closed field lines, we know where the magnetopause current will flow and how strong it will be. But when the interplanetary magnetic field turns southward the current system moves and it changes its strength. We can predict neither yet from first principles even though we have some empirical models.

If the interplanetary magnetic field is strong and remains southward for many hours, then the plasma deep inside the Earth's magnetosphere becomes energized and a new current system arises called the ring current. This current reduces the field on the surface of Earth and expands the magnetosphere. The two main theories on the origin of the ring current are diametrically opposite. Some maintain that the ring current arises when the magnetosphere is subjected to strong time variations in external conditions, while others maintain that steady conditions are needed for deep injection of particles into the magnetosphere.

The largest and most powerful magnetosphere in the solar system, that of Jupiter, also has a ring current. This one is associated with the acceleration of ions produced from the volcanic gases of Io. While no one doubts the origin of these ions, there is an analogous problem at Jupiter about the origin of magnetospheric activity. Which processes are energized by the solar wind, as happens at Earth, and which processes are energized by the rotation of the planet? The resolution of these questions has minor consequences for Earth but major consequences for astrophysical systems.

Jupiter and Earth also are major proving grounds for models of the reconnection process. The physics of the reconnection process centers around the *x* point, or neutral point, as illustrated in Figure 2.1. Here the problem is the annihilation of magnetic fields and not so much their creation. Magnetic fields can store energy. Earth's magnetotail and, even more so, the jovian magnetotail are giant storehouses of magnetic energy. Energy can be extracted from the solar wind and the planet and stored in these regions for later release, a release that can be quite rapid. But what triggers this release? Is it internally triggered or externally triggered? What controls the time scales? Are microscale processes critical to reconnection, or is reconnection governed principally by large-scale conditions and processes? The strong, global disturbances, known as geomagnetic storms, and the weaker localized events, called substorms, are produced by different solar wind conditions. How do these conditions alter the internal workings of the magnetosphere to create these phenomena? Why is there not more of a continuum of disturbance strengths? Reconnection takes place at both the dayside magnetopause and in the tail, but we generally treat the reconnection at the magnetopause as driven and in the tail current sheet as triggered. Why do the two regions appear to behave so differently, or is our understanding at fault? In short, we have much to learn about both the creation and the annihilation of magnetic fields in planetary magnetospheres.

MAGNETOSPHERES AS SHIELDS AND ACCELERATORS

One of the fascinating lessons from magnetospheric research is that magnetized bodies (intrinsic or induced) have a Jekyll-and-Hyde behavior regarding high-energy charged particles. On the one hand, the magnetic field forms a shield around the planet that, through magnetic forces, excludes cosmic rays and other energetic particles from much of the surface and prevents the solar wind from scouring the upper reaches of the atmosphere. On the other hand, the dipolelike magnetic field forms a trap for charged particles and enables their acceleration to very high energies, through a number of intriguing and cosmically important physical processes. This dual personality—shield and accelerator—raises a number of compelling questions, questions for which answers are now on the horizon.

As described in the previous section, Earth's magnetosphere is a bubble in space that largely deflects and excludes the supersonic solar wind flow impinging on it. The processes by which this exclusion occurs are themselves fascinating subjects of study, and our observations and theoretical investigations to date have raised important questions: How leaky is this shield? Can the solar wind momentum and energy fluxes be efficiently transferred into the system even if solar wind matter is largely excluded? How does the efficiency of the shield depend on the properties of the impinging solar wind? Answering these important questions requires both a detailed understanding of the processes by which the shield is breached and a global understanding of where and under what conditions the transfer into the magnetosphere takes place.

Similarly, Earth's magnetic shield prevents many cosmic rays and highly energized solar particles from entering much of near-Earth space. Yet such particles do have access to the polar regions of Earth and to the outer layers of the magnetospheric bubble, posing questions of practical significance for humans and technology: How deeply can such particles penetrate into our atmosphere? How low in latitude can they reach? What solar wind conditions allow them easier access? How can we anticipate the levels, spatial distribution, and duration of energetic particle enhancements? These and similar questions underlie accelerating scientific efforts to understand the physics that produces space weather. Our growing vulnerability to space weather disturbances makes it increasingly urgent to answer such questions, allowing us to predict and mitigate the effects of potentially damaging space radiation.

How Earth's magnetic shield works, and how well it works, are significant and challenging problems. An equally interesting problem is what would happen if this magnetic shield did not exist. Indeed, it may not have existed during episodic reversals of the geomagnetic field over Earth's history. What happened to Earth and its atmosphere during those episodes? How much of our atmosphere could have been lost to solar wind scouring? What processes would have come into play? What levels of energetic particle fluxes would have had access to Earth's surface? What consequences would these effects have had for the viability and evolution of life? The answers to these profound questions will lie in a better understanding of how the solar wind interacts with the atmosphere and its ionized upper reaches under conditions of weak or absent intrinsic magnetic fields. To help us to achieve this understanding, nature provides an array of solar system bodies with widely differing intrinsic fields and atmospheres—from unmagnetized, airless bodies that are exposed to the direct impact of the solar wind, to comets and planets that "capture" the solar wind magnetic field to form their own induced shield. Already, the brief glimpses we have obtained of these exotic systems show us the value of comparative magnetospheric studies. Extending our scientific scrutiny to other such bodies will greatly aid us in answering the questions about the past and future of Earth's shield.

The other side of the magnetospheric Jekyll-and-Hyde personality is the clear propensity for magnetically governed systems to store charged particles from various sources and accelerate some of them to very high energies. The cosmic problem of particle trapping and energization is typified by Earth's radiation belts. We have been aware of this region of energetic trapped particles since the flight of Explorer 1 in 1958, and much early observational work delineated its extent, its basic properties, and some hints about how it formed and how it varied with the solar cycle. More recent satellite observations have revealed a completely unanticipated degree of temporal variability to the radiation belts, showing them to be a fundamentally dynamic feature exhibiting formation and decay over a wide range of time scales, including the sudden formation of an entire new belt of radiation within a few hours.

These more recent observations have left us with a host of compelling questions about the formation and decay of trapped energetic particle populations: We know that particles can be accelerated suddenly during episodic events as well as more slowly over time through a process of diffusion, but we do not know the relative importance of these processes. Moreover, we know some of the aspects of diffusive acceleration, but we do not know just where and under what conditions nonreversible processes deposit energy into the trapped population. Nor do we know which solar wind properties drive the diffusive transport or how it is driven or what determines the time scale. Furthermore, there remain fundamental questions about the source population, about its possible preconditioning, and about how radiation belt particles are lost—through precipitation into the atmosphere, loss to interplanetary space, wave-particle interactions, or particle-particle interactions such as charge exchange.

MAGNETOSPHERES AS COMPLEX COUPLED SYSTEMS

A major obstacle in understanding the behavior of Earth's magnetosphere and other magnetospheres as well and an important challenge for years to come is that magnetospheres do not exist in isolation and that they show an internal complexity comparable to that of a human body. They are formed by the interaction with the solar wind and are hence strongly coupled to their surroundings. Perhaps even more important, however, are the internal couplings between the different parts of magnetospheres and the couplings between various processes that operate on vastly different scales.

As in a human body, electrical currents and mass flows govern the transport of information and energy. In contrast to the human body, however, these flows are not confined to channels like blood vessels or nerves. Although currents and plasma flows in magnetospheres are often concentrated in thin layers or tubes also, these layers or tubes form dynamically within the three-dimensional plasma and field distributions; they can change their locations and they can form newly or disappear. Owing to this complex behavior, we still understand only poorly the response of the magnetosphere to the equivalent of a local artery blockage—that is, the local disruption of the electric current flow across the geomagnetic tail. In some circumstances it may just create its own local bypass; in others it may cause a major collapse, a magnetospheric substorm. There are indications that a combination of the large-scale structure and the local conditions determines the magnetospheric response to such clogging, but the mechanism is not well understood.

Geomagnetic storms and substorms are the prime examples of the complex coupled behavior of magnetospheres, involving all of its parts and processes on all scales. During past decades, we learned a lot about the morphological changes associated with substorms, but problems remain unsolved. We know that the magnetic field orientation in the solar wind plays a crucial role: A sustained southward interplanetary magnetic field enables magnetic reconnection at the magnetopause and energy transfer to the magnetosphere and, particularly, to the magnetic tail. The release of this energy constitutes the main part of the substorm. It involves the severance of the magnetotail plasma, in the form of a plasmoid ejected down the tail, and the collapse of the inner tail, with changes of the coupled current system that also includes the ionosphere. Induced electric fields in the collapsing inner tail cause particle energization. The coupled current systems also involve currents and electric fields that are aligned with the magnetic field and cause bright auroral features.

However, we still do not know how a substorm is triggered. There are strong indications that in perhaps half of the cases a northward shift of the IMF might act as a trigger. But that conclusion actually increases the complexity. Why does a northward shift act as a trigger in some cases but apparently not in others? The ejection of a plasmoid requires magnetic reconnection to operate in the tail. But what are the processes leading to reconnection? It is clear that these processes must be collisionless plasma processes involving gradients on very small scales. The formation of parallel electric fields associated with auroral forms also requires plasma processes on very small scales. Thus small-scale processes appear strongly coupled with the large-scale dynamics of the magnetosphere. There have been a number of suggestions for such processes but no process has been uniquely identified.

The connections between the tail and the inner parts of the magnetosphere and the ionosphere raise major questions about the coupled system. What is the role of substorms in the transport of plasma and magnetic flux from the tail to the inner magnetosphere? Magnetospheric substorms may occur in isolation, but they are also an important part of big storms. What is their relationship? What is the relationship between tail dynamics and auroral features? The concept of space weather intrinsically concerns the global coupled system, involving the impact of coronal mass ejections, magnetospheric storms and substorms, particle energization, flows, and currents and their closure through various parts of the magnetosphere/ionosphere system. Global MHD simulations of this coupled system have been surprisingly successful in modeling certain aspects of this complex interaction but unsuccessful in others.

The magnetospheric system is a high-Reynolds-number system. Flows in such a system are typically highly turbulent. What is the role of this turbulence in magnetospheric transport? How does it possibly obstruct some forms of transport or enable others? What determines the scale sizes of flows? We have also learned that the magnetospheric plasma ultimately comes from two sources, the solar wind and the ionosphere. However, it is not well understood how these sources contribute under various circumstances, and how the plasma composition influences the plasma behavior on small and large scales.

A resolution of these problems obviously requires a coordinated investigation of the coupled system, involving both local and large-scale views, in close combination with theoretical models that address both small and large scales and their coupling. Significant insight also comes from the study of nonterrestrial magnetospheres. Learning whether and how analogous processes, such as storms or substorms or auroras, operate under different conditions will provide valuable clues to promote their physical understanding.

2.4 CURRENT PROGRAM

INTRODUCTION

To accomplish the varied and ambitious scientific objectives described in Section 2.3, a balanced program of theory, modeling, observations, and data analysis is needed. To answer questions regarding the structure and physics of various regions of the solar wind–magnetosphere system, in situ measurements of a number of important physical parameters must be made; this means that NASA, which is primarily responsible for the nation's civilian space platforms, is the dominant agency in the current program. However, it is a hallmark of the space science disciplines that other agencies also contribute key elements to the overall program. Examples include ground-based magnetometer and radar measurements, digital all-sky images, and photometric data, which are crucial to identifying the magnetospheric response to various solar wind stimuli, and magnetic field and particle measurements obtained by instruments on operational DOD, NOAA, and DOE satellites, which provide long-baseline and multipoint monitoring of magnetospheric activity. Moreover, a number of different agencies have ongoing programs to support the scientific analysis of relevant data, as well as related theory and numerical modeling.

In general, the activities of the various agencies have not been formally coordinated, although the recent focus on space weather has introduced an increased measure of communication, coordination, and collaboration among the agencies. Rather, it is the scientific community (which itself is a blend of participants from universities, NASA centers, government laboratories, and industry) that provides the glue to connect the independent programs of the various agencies. The research community provides advisory support to the different agencies, facilitating informal interagency communication. More importantly, the research community draws on all the available resources, combining them as needed, to carry out the scientific process. In recent years this ability to combine observations from disparate sources has been greatly facilitated by rapidly expanding use of the Internet. Data distribution centers, such as those at NASA/GSFC and NOAA/Space Environment Center, as well as the more limited databases at other institutions, have made it possible for individual researchers to access and combine data from many different sources. This development is discussed more extensively in Chapter 4, "Report of the Panel on Theory, Computation, and Data Exploration."

PROGRAMS

Tables 2.1 to 2.4 list current programs that address problems related to solar wind–magnetosphere interactions. For each program the tables give the responsible agency, the status, a brief description, and a summary of the science objectives. Table 2.1 lists the non-NASA programs. These programs are mainly funded by four agencies: NSF, USAF, NOAA, and DOE. Table 2.2 lists NASA's nonflight and suborbital programs. Table 2.3 lists NASA's nonplanetary missions relevant to solar wind–magnetosphere interactions. Table 2.4 lists NASA's missions relevant to planetary magnetospheres, both intrinsic and induced. The funding levels for these various programs are not given. In general, the flight programs represent the largest dollar commitments, but a large fraction of the funding for flight programs goes to spacecraft contractors, launch services, and operations.

An examination of Tables 2.1 to 2.4 also reveals important aspects of the current program of research in solar wind–magnetosphere interactions:

1. As noted above, the program is multiagency and, indeed, international in scope. NASA maintains a very active engagement with its partner space agencies in Europe, Japan, and Russia, producing opportunities for non-U.S. contributions to American space missions, as well as U.S. participation in programs devised and led by the other agencies. Such collaborations are extremely valuable and scientifically productive, and for some planet–solar wind interaction goals they provide our only option. Thus it is a concern that they are subject to political difficulties; in recent years, new State Department rules regarding the International Traffic in Arms Regulations have made international collaborations significantly more difficult.

2. The current endeavor is a broad-based mix of small and large programs, ground-based and space-based, observations, and theory. It addresses a wide range of phenomena and physics involved in solar wind–magnetosphere interactions. And it builds on previous programs, using earlier results to refine the questions to be addressed and the techniques that should be employed.

3. Much of the current program of solar wind–magnetosphere research depends on the opportunistic use of resources that were originally dedicated to other objectives (note the "Science Objectives" column), representing a strong leveraging of other investments. This includes measurements obtained by non-NASA agencies (e.g., DOE), as well as observations from NASA missions targeted at other science objectives (e.g., ACE, which is a mission to study the mass composition of cosmic rays but which has become the mainstay of solar wind monitoring). It also includes planetary missions to a very limited extent; such opportunities for making comparative magnetospheric studies are of particularly high scientific value. The scientific advances that have emerged from such opportunistic programs demonstrate that, although there is strong pressure within NASA for scientifically focused missions, application of creativity and modest additional investments can often reap major dividends in ways that were not part of the original objectives.

4. Many of the operating flight programs are well past their prime. That such missions are still returning very valuable—and in some cases, unique—observations demonstrates the cost-effectiveness of extended missions. Moreover, the overlap of these earlier programs with more recent missions greatly enhances the scientific return of both, by providing additional perspectives and complementary information about various phenomena.

5. There exists a conscious effort to exploit existing data sets, both at NASA and at some of the other agen-

TABLE 2.1 Non-NASA Programs in Solar Wind–Magnetosphere Interactions

Program	Agency	Status	Description	Science Objectives
Base program, magnetospheric and solar terrestrial	NSF	Ongoing	Supports solar-terrestrial and magnetospheric research, primarily theory and modeling studies.	Elucidate the full range of physical processes operating in the solar, solar wind, magnetospheric, and ionospheric environments.
Upper Atmospheric Facilities (Division of Atmospheric Sciences)	NSF	Ongoing	Supports four incoherent-scatter radar facilities and the SuperDARN coherent scatter radar system.	Promote basic research on the structure and dynamics of Earth's upper atmosphere.
Geospace Environment Modeling (GEM)	NSF	Ongoing	Supports basic research into the dynamical and structural properties of the magnetosphere, primarily in theory and modeling.	An end-to-end capability to model predictively the full chain of geospace environment dynamics, including the construction of a global geospace general circulation model (GGCM).
Space Weather Program (part of the interagency National Space Weather Program)	NSF (with contributions from the Air Force)	Ongoing	Supports basic research into processes and phenomena with space weather consequences; includes theory, modeling, and analysis of existing data sets.	Identify and pursue research into space weather phenomena that has the greatest potential for service to society and to facilitate the transition of research results to practical utilization at forecast centers.
Defense Meteorological Support Program (DMSP)	Air Force	Ongoing	Multiple satellites in low-altitude, Sun-synchronous orbit; remote sensing and particles and fields instrument complement.	Provide continuous visual and infrared imagery of cloud cover, as well as measure the atmospheric vertical profiles of moisture and temperature; to measure local charged particles and electromagnetic fields to assess the impact of the ionosphere on ballistic-missile early warning radar systems and long-range communications; to monitor global auroral activity; and to predict the effects of the space environment on military satellite operations.
Base program/Directorate of Mathematics and Space Sciences	Air Force Office of Scientific Research	Ongoing	Research grants to universities, industry, and AFRL technical directorates.	Sponsor and sustain basic research; transfer and transition research results; support Air Force goals of control and maximum utilization of air and space.
Geostationary Operational Environmental Satellites (GOES)	NOAA	Ongoing	Multiple geosynchronous satellites; payload includes remote sensing, energetic charged particle detectors, and a magnetometer.	Various meteorological objectives, plus monitor the near-Earth space environment.
TIROS satellites	NOAA	Ongoing	Multiple (normally, two) satellites in low-altitude polar orbit; remote sensing plus energetic charged particle sensors.	Various meteorological objectives, plus monitor the near-Earth space environment.
Space Environment Center; National Geophysical Data Center	NOAA	Ongoing	Access to current and archived geomagnetic and satellite data and indices.	
Los Alamos National Laboratory geosynchronous space environment sensors	DOE	Ongoing	Multiple satellites in geosynchronous orbit; magnetospheric plasma and energetic charged particle instrumentation.	Monitor the space environment, including the plasma sheet, the plasmasphere, outer radiation belts and ring current, and dynamical events such as geomagnetic storms and substorms.
Los Alamos National Laboratory GPS space environment sensors	DOE	Ongoing	Multiple satellites in inclined, 12-hour circular orbit; energetic charged particle instrumentation.	Monitor the space environment, especially the outer radiation belts and dynamical events such as geomagnetic storms and substorms.
Various	Other U.S. government agencies	Occasional	Rides of opportunity on programmatic launch vehicles.	Depends on payload.

TABLE 2.2 NASA Nonflight, Suborbital, and Related Astronomy Programs

Program	Agency	Status	Description	Science Objectives
Hubble Space Telescope (HST)	NASA	Ongoing	Spaceborne telescope complement (visible, IR, UV).	To obtain extremely high-resolution images of space objects, including the planets of our solar system. HST has imaged auroras on Jupiter and Saturn.
Infrared Telescope Facility	NASA and NSF[a]	Ongoing	Ground-based telescope established by NASA in 1979 primarily to provide infrared observations in support of NASA's programs.	The facility is available for all types of infrared observations, but it emphasizes solar system studies. Of particular interest for this report is the use of IRTF for planetary aeronomy.
Living With a Star, targeted research and technology program	NASA	Ongoing	Funds basic and applied analysis aimed at developing the scientific understanding necessary to enable the United States to effectively address those aspects of the connected Sun-Earth system that affect life and society.	To exploit data from past and present space missions for scientific analysis, theory, and modeling efforts as well as technology improvements that lead to operational answers in areas relevant to societal needs.
Low-cost access to space (LCAS)	NASA	Ongoing	Funds payload development and launch on suborbital platforms (balloons and rockets) and low-cost orbital opportunities (shuttle-based carriers, the International Space Station, and other flights of opportunity).	Same as in geospace SR&T program, below.
Sun-Earth Connection (SEC) Guest Investigator Program	NASA	Ongoing	Provides grants to support basic research and analysis, utilizing existing databases obtained by SEC missions.	To understand the solar interior and the solar atmosphere, including the evolution of mass and energy ejected from the solar atmosphere; to understand the propagation of disturbances in the three-dimensional as well as the distant heliosphere; to investigate the flow of mass, energy, and momentum throughout the near-space environment of Earth; and to investigate the effects of solar activity on Earth's atmosphere and ionosphere.
SEC Theory Program	NASA	Ongoing	Provides grants to support critical-mass theory and modeling groups addressing problems in solar physics, heliospheric physics, magnetospheric physics, and ionospheric, thermospheric, and mesospheric physics.	Same as in the SEC Guest Investigator Program, above.
Supporting research and technology (SR&T), especially the geospace and the solar and heliospheric physics clusters	NASA	Ongoing	Provides grants for basic research and analysis, including theory, modeling, data analysis, and instrument concept development that support NASA's flight programs, especially future programs.	To understand the region of space that surrounds and is influenced by Earth and its magnetic field, beginning with the neutral upper atmosphere and extending outward through the ionosphere, into and beyond the magnetosphere; to understand the origins of solar variability, its effects on the solar atmosphere and the heliosphere, and the transport and dissipation of matter and energy through the solar atmosphere to the outer edges of the heliosphere.

[a] The Infrared Telescope Facility (IRTF) is managed and operated by the University of Hawaii's (UH) Institute for Astronomy. NASA provides the costs of operation, and NSF provides support for new focal plane instrumentation based on grant applications from IRTF-supported astronomers. Observing time is open to the entire astronomical community, and 50 percent of the IRTF observing time is reserved for studies of solar system objects.

TABLE 2.3 NASA Flight Programs

Program	Agency	Status	Description	Science Objectives
Advanced Composition Explorer (ACE)	NASA	Operating (launched August 1997)	Explorer mission in L1 halo orbit.	To study mass composition of energetic solar particles and cosmic rays. Includes magnetometer and solar wind instrument, which serve as upstream monitors of solar wind conditions imposed on the magnetosphere.
Cluster II	ESA/ NASA	Operating (launched July and August 2000)	Part of ESA's Solar Terrestrial Science Program (STSP). Four-spacecraft cluster, launched in pairs and reunited in orbit; inclined elliptical orbit.	To determine the physical processes involved in the interaction between the solar wind and the magnetosphere by visiting key magnetospheric regions (solar wind and bow shock, magnetopause, cusp, magnetotail, and auroral zone) and to study the three-dimensional plasma structures in space and time in these regions.
Fast Auroral SnapshoT Explorer (FAST)	NASA	Operating (launched August 1996)	Small Explorer mission in low-altitude, elliptical orbit; well-instrumented for fast measurements of particles and fields.	To study Earth's auroral regions with very-high-time-resolution measurements. Combined with upstream solar wind measurements and geomagnetic indices, it also aims to explore the influence of solar wind variations on ionospheric outflows and electrodynamic coupling.
Geotail	ISAS/ NASA	Operating (launched July 1992)	Part of the ISTP fleet; near-equatorial orbit with variable apogee; particles and fields instrumentation.	To study the dynamics of Earth's magnetotail over a wide range of distance, extending from the near-Earth region (8 Earth radii (R_E) from Earth) to the distant tail (about 200 R_E).
HESSI	NASA	Operating (launched February 2002)	Small Explorer in low-Earth orbit; high-resolution imaging and spectroscopy of solar flares from 3-keV x rays to 20-MeV gamma rays with high time resolution.	To determine the frequency, location, and evolution of impulsive energy release in the corona; study the acceleration of electrons, protons, and heavier ions in flares; study the heating of plasma to tens of millions of degrees and determine its relationship to particle acceleration; study the propagation and evolution of energetic particles in flares; determine the relative abundances of accelerated and ambient ions in flares; plus astronomy science objectives.
IMAGE	NASA	Operating (launched March 2000)	Medium-class Explorer in high-inclination elliptic orbit; magnetospheric remote-sensing payload.	To produce the first comprehensive global images of the plasma populations in the inner magnetosphere in order to observe, in a way never before possible, the large-scale dynamics of the magnetosphere and the interactions among its constituent plasma populations.
Polar	NASA	Operating (launched February 1996)	Part of the ISTP fleet; near-polar elliptical orbit; particles and fields instrumentation, plus UV, x-ray, and optical imagers.	To learn how the solar wind plasma energy enters into the magnetosphere through the polar cusp on the dayside of the magnetosphere; to determine the mechanisms that cause ionospheric plasma outflow; to discern the importance and characteristics of various processes that accelerate aurora-producing particles; to investigate the ways in which energy and momentum are exchanged between collisionless plasmas and electromagnetic fields in the magnetosphere; and to determine the rate of energy input into the atmosphere from auroral particles and their effects on the atmosphere.

SAMPEX	NASA	Operating (launched July 1992)	Small Explorer in low-altitude, near-polar orbit.	To study the ionization state and isotopic composition of the anomalous component of cosmic rays; to observe precipitating magnetospheric electrons that interact with the middle atmosphere; to investigate the isotopic composition of particles originating in energetic solar flares in different phases of solar activity; and to investigate the origin and transport of galactic cosmic rays.
SOHO	ESA/ NASA	Operating (launched December 1995)	Part of ESA's STSP. Solar remote sensing instruments in L1 halo orbit.	To study the internal structure of the Sun, its extensive outer atmosphere, and the origin of the solar wind; to monitor the state of the solar atmosphere, including the occurrence of coronal mass ejections and solar flares.
Solar-B	ISAS/ NASA/ PPARC	In development (launch 2005)	Single spacecraft in Sun-synchronous Earth orbit, with a coordinated set of optical, EUV, and x-ray instruments.	To study the creation and destruction of the Sun's magnetic field; observe the modulation of the Sun's luminosity with a resolution, wavelength coverage, and time span adequate to determine the mechanism responsible for the modulation; study processes such as magnetic reconnection and wave dissipation that are believed to be responsible for the conversion of magnetic energy into UV and x-ray radiation; with accurate measurements of magnetic fields, electric currents, and velocity fields, reveal the root causes of the Sun's plasma eruptions.
TWINS	NASA	Small Explorer mission of opportunity (U.S. government satellites)	Suite of energetic neutral atom imagers on two satellites in high-altitude, high-inclination orbits, to provide stereo neutral atom imaging of magnetospheric plasmas.	Three-dimensional visualization and resolution of large-scale structures and dynamics within the magnetosphere.
Ulysses	ESA/ NASA	Operating (launched October 1990)	A single spacecraft in polar orbit about the Sun, providing the first in situ observations of the inner heliosphere at high heliographic latitudes; instrumented primarily with particles and fields instruments.	To characterize the solar wind and interstellar magnetic field inside 5 AU as a function of solar latitude and solar activity level; to study cosmic rays and interstellar and interplanetary neutral gas and dust.
Wind	NASA	Operating (launched November 1994)	Part of the ISTP fleet; various orbits for different mission phases; originally served as a solar wind monitor, but has also provided magnetospheric measurements.	To monitor the upstream plasma, energetic particle, and magnetic field conditions for magnetospheric and ionospheric studies; to determine the magnetospheric output to interplanetary space in the upstream region; to investigate basic plasma processes occurring in the near-Earth solar wind; and to provide baseline ecliptic plane observations to be compared with out-of-the-ecliptic observations by Ulysses.
Yohkoh	ISAS/ NASA/ U.K.	Operating (launched December 1995)	Part of ESA's STSP. Solar remote sensing instruments in L1 halo orbit.	To study the internal structure of the Sun, its extensive outer atmosphere and the origin of the solar wind; to monitor the state of the solar atmosphere, including the occurrence of coronal mass ejections and solar flares.

TABLE 2.4 Planetary Missions with Magnetospheric Science Potential

Program	Agency	Status	Description	Science Objectives
Bepi-Columbo Mercury orbiter	ESA/ ISAS	Approved; concept definition under way (launch ~2010)	ESA cornerstone mission to Mercury. Bus carries two science satellites: Mercury planetary orbiter for remote sensing and Mercury magnetospheric satellite (built by ISAS) for in situ measurements, plus a Mercury surface element penetrator that will be dropped to the surface. The magnetospheric satellite includes a magnetometer, wave instruments, and particle spectrometers.	Overall mission is to investigate Mercury's interior, geology, atmosphere, the possibility of surface water ice, the origin and character of Mercury's magnetic field, and the solar wind interaction.
Cassini Orbiter	NASA	In cruise phase (launched 1997, arrival at Saturn 2004)	Single large spacecraft carrying the Huygens probe provided by ESA. Comprehensive orbiter payload includes imagers for visible, IR, and UV wavelengths, radar, in situ particles and fields experiments.	After dropping off the Huygens probe at Titan and relaying Probe data, to carry out a full reconnaissance of the Saturn system in the style of Galileo at Jupiter, with a special focus on Titan.
Galileo	NASA	Operating (launched 1989, arrived at Jupiter 1995)	Jupiter orbiter with atmospheric probe. Comprehensive orbiter payload including imaging in visible, IR, and UV wavelengths, in situ particles and fields experiments.	Full reconnaissance of the Jupiter system, with a special focus on Io. Extended mission focus on Io and Europa, to monitor Io's variability and assess Europa as a potential candidate for future exobiology investigation.
Mars Express	ESA	In cruise phase (launch 2003, arrival at Mars 2004)	Three-axis stabilized spacecraft designed for remote sensing. High-inclination, elliptical orbit with periapsis ~250 km and apoapsis ~11,000 km. Carries a space physics package including an energetic neutral atom (ENA) detector and NASA-funded electron spectrometer and ion mass analyzer.	Main objectives are to search for subsurface water and drop off the Beagle-2 Lander. The space physics experiment will study the Mars-solar wind interaction. Planned cooperative activities with Nozomi.
Mars Global Surveyor (MGS)	NASA	Operating (launched 1996)	Spacecraft designed primarily for remote sensing. Early mission phases included an elliptical orbit aerobraking phase with a periapsis that sometimes reached ~110 km. Circular mapping phase orbit is near-polar at ~400 km altitude in a Sun-synchronous orbit. Space physics-related instruments include a magnetometer and an electron analyzer/reflectometer.	To map the surface geology and mineralogy, probe the topography and interior, and measure the planetary magnetic field of Mars.

Mars Odyssey	NASA	Operating (launched 2001)	Single spacecraft designed primarily for remote sensing. A space physics-related instrument is the Martian Radiation Environment Experiment, which detects energetic particles and photons with an instrument designed for potential biological effects diagnostics. Also carries a communications relay for future Landers.	To remotely sense and map the mineralogy and morphology of the martian surface, and to characterize the radiation environment related to human exploration.
Mercury Surface, Space Environment, Geochemistry, and Ranging (Messenger)	NASA	Under construction (launch 2004, in orbit 2009)	Discovery class spacecraft with Sun shield. Two flybys (2007, 2008) before orbit insertion. Elliptical, high-inclination orbit, 200 km periapsis, 15,000 km apoapsis. Includes several instruments for space physics measurements: magnetometer, plasma and energetic particle spectrometer, UV spectrometer.	Global mapping and characterization of the surface, interior, atmosphere, and magnetosphere of Mercury.
Nozomi	ISAS	In extended cruise phase (launched 1998, arrival at Mars 2004)	Spinning spacecraft in a low-inclination orbit with periapsis ~150 km and apoapsis ~50,000 km. Comprehensive space physics and aeronomy payload includes a magnetometer, electron analyzer, ion analyzer, electron and ion spectrometer, plasma wave instrument, ion composition analyzer, UV spectrometer, EUV scanner, Langmuir probe, and ion-neutral mass spectrometer (NASA-provided).	To study the martian upper atmosphere and its interaction with the solar wind.
Rosetta	ESA	Under construction (launch 2004, arrival at Comet 67P/ Churyumov-Geasimenko 2014)	ESA cornerstone mission to Comet 67P/Churyumov-Geasimenko. Space physics-related instruments include a magnetometer, ion/electron analyzer, ion composition analyzer, Langmuir probe, and mutual impedance probe. The Lander carries amagnetometer and plasma detector.	To study the nucleus of Comet 67P/Churyumov-Geasimenko and its environment continuously as the comet approaches the Sun. To land on the cometary nucleus.
SELENE	ISAS, NASDA	Under construction (launch 2003)	Main spacecraft plus a relay subsatellite. Main spacecraft to go into near-polar lunar orbit at ~100 km altitude, subsatellite to go into elliptical orbit with apoapsis ~2,400 km. Instruments include a magnetometer and a plasma instrument whose data can be used for space physics studies. Most instruments for remote sensing of surface, structure, and mineralogy. Technology development goal for future lunar exploration.	To obtain information on lunar origin and evolution.

cies. It is by data interpretation and related theory and modeling that we strategically build on current knowledge, so it is crucial that such efforts be adequately supported. This topic was considered in the NRC report *Supporting Research and Data Analysis in NASA's Science Programs: Engines for Innovation and Synthesis* (Space Studies Board, National Research Council, 1998) and is addressed at greater length in "Report of the Panel on Theory, Computation, and Data Exploration," Chapter 4 of this volume.

CRITICAL NEEDS

Tables 2.1 through 2.4 above describe a robust program that is undertaking significant studies on a broad array of topics. The success of this program is due to many factors. Some of these factors need repeating so that the future program can continue the productivity of the present program and that past lessons need not be relearned. Other recommendations concern changes that should be made.

Support for the Analysis of Existing Data

Such support would provide for the documentation and archiving of the data and for making the data and results widely available and easily usable. The attention to this topic varies across disciplines. In the planetary arena the Planetary Data System (PDS) has been funded to provide discipline-wide standards and discipline-wide oversight in archiving and data access. In the terrestrial arena the responsibility has devolved to the project level because there is no space physics or Sun-Earth Connection data system. Missions have finite lifetimes centered on the period over which the spacecraft operates. When missions end, their financial support soon ends as well. This puts their data sets at risk. The National Space Science Data Center is the ultimate archive for PDS and SEC data sets, but it does not provide the level of documentation or of archival and scientific expertise for SEC data that the PDS provides for solar system exploration data. The experience of the PDS indicates that a distributed, community-based archiving organization is the most effective means of providing long-term access to data.

Interplay Among Space- and Ground-Based Observations and Data Analysis, Modeling, and Theory

In a program such as we have at present new observations can be incorporated into the existing paradigms rapidly and progress is rapid. There are supporting data from space for ground-based observations, and supporting data from the ground are often available for space-based studies. Models have been developed to test current theories and to allow global pictures to be extrapolated from more localized data. This successful interplay provides an important lesson for the future. Progress in science requires a holistic approach whereby the entire system is examined. Seldom will one new observation in isolation be sufficient to permit the achievement of a breakthrough in this field.

Communication Between Programs and Agencies

The present strong communication has been due in part to the involvement of a number of the key players and in part to coordination of the agencies, both domestic and international, the latter coordinated by the Inter-Agency Consultative Group. This communication helps avoid duplication of effort and services, develops synergies, and leverages opportunities. The panel strongly recommends keeping the communication lines between future projects in this field as open as they currently are.

Flexibility in Research Programs So That an Unexpected Research Result Can Be Incorporated into the Effort

Historically, scientific research has often benefited from serendipitous events. Not all research results can be or should be targeted. When new knowledge is gained, we need to find efficient pathways by which to transfer it to other fields and in turn to gain relevant knowledge from other disciplines.

Better Mechanisms for Taking Advantage of Missions of Opportunity

Currently NASA makes these opportunities available very seldom compared with the frequency at which the opportunities arise, and the selection process is long and cumbersome. NASA should issue mission-of-opportunity Research Announcements more frequently.

Mechanism for Facilitating International Collaborations

The present technology transfer regulations serve to stifle such collaboration. They also affect purely domestic programs.

Education and Public Outreach for the Sun-Earth Connection Program

There are many excellent individual efforts under way, but these could be improved with increased communication, leveraging, and synergy.

2.5 FUTURE PROJECTS

INTRODUCTION

Even after the ongoing missions detailed in Section 2.4 are completed, the research community will need a series of synergistic, interagency missions, projects, and initiatives during the next decade if it is to resolve the outstanding problems outlined under this report's three major themes: magnetospheres as complex, coupled systems; magnetospheres as shields and accelerators; and the creation and annihilation of magnetic fields. These future projects represent the next logical steps toward scientific closure, building on knowledge gained from previous and ongoing projects.

The collective program will comprise a core of basic research, complemented by focused, intermediate-size, principal investigator (PI)-driven missions, rounded out by large, community-wide facilities, missions, and programs. These larger projects will typically require substantial financial investment. In this section, the panel outlines particularly those future projects that demand community consensus and prioritization. The panel recognizes that a healthy science program will always require smaller, PI-driven initiatives as well. Because these smaller initiatives will be competed for through the peer-review process, they require neither discussion nor prioritization here.

The NSF and NASA have separate planning processes for prioritizing major initiatives. Indeed, these planning and prioritization exercises have benefited and continue to benefit from broad community input. Therefore, to the greatest extent possible, the panel takes the results of those exercises as the starting point for its discussion. It notes also that the nascent National Space Weather Program has provided an impetus for the first deliberate, long-range interagency planning of space physics programs. In that vein, some programs, projects, and missions may span several agencies through cooperative funding arrangements. These agencies are the Department of Commerce/NOAA, the Department of Defense/Air Force and Navy, and the Department of Energy.

In the remainder of this section, the panel outlines the projects that will be needed to answer the outstanding questions identified in Section 2.2 in the context of the broader science themes delineated in Section 2.3. The following subsections give thumbnail descriptions of the projects' science objectives along with scenarios for their implementation, vetted at past and ongoing community consensus planning exercises.

The panel emphasizes that the planning for future NASA missions has benefited from substantial community input through the mechanism of internal and external advisory committees; from Office of Space Science NASA Roadmap committees, which develop missions to achieve NASA's strategic science objectives; and from science architecture teams developing targeted science programs such as NASA's Living With a Star program. The panel has reviewed these inputs from the community and ideas that featured prominently in earlier planning documents. Based on this review, the panel sets priorities for basic science in both terrestrial and planetary magnetosphere–solar wind studies, as well as in the targeted science LWS program.

ADDRESSING THE MAJOR THEMES

Before itemizing the specific projects required to meet our science goals, the panel reiterates the major themes that motivate the larger science framework. These themes are the fundamental science foundations upon which the individual projects build. After outlining these themes, the panel then describes each of its recommended future projects, shows how they relate to the earlier science questions, and presents its priorities for addressing these questions.

Magnetospheres as Complex, Coupled Systems

The outstanding questions identified in Section 2.2 require a new type of widely distributed data array that will make it possible to distinguish between spatial and temporal variations. While planetary magnetospheres display as much complexity and coupling as the terrestrial magnetosphere, the need to sample the volume of space under investigation as completely as possible means that complex, coupled processes will have to be studied mainly at the Earth. In order to study the complex, coupled system of the solar wind's interaction with Earth, we need both space-based and ground-based facilities. In space, answers to the outstanding questions require fleets of spacecraft arrayed in constellations and missions that can provide global images of magnetospheric plasmas. Similarly, on the ground, either dense arrays of observing platforms (e.g., magnetometers) or next-generation, synoptic remote-sensing tools are needed. Such projects are the next logical step toward understanding how the magnetosphere reacts as a large-scale complex system, building on previous "anatomical" projects that have systematically dissected the system region by region and process by process. The companion report by the Panel on Atmosphere-Ionosphere-Magnetosphere Interactions also stresses the importance of distributed ground systems for furthering our understanding of the solar wind-magnetosphere-ionosphere system.

Magnetospheres as Shields and Accelerators

Many of the outstanding questions elucidated in Section 2.2 focus on the ongoing mystery of particle acceleration within planetary magnetospheres. This acceleration comes in many different forms and in many different settings through a host of possible processes, some of which may be driven externally and others of which may be driven by an internal energy conversion process. Jupiter, for example, has the most intense radiation belts in the solar system. Closer to home, there exist many compelling and practical consequences of the transient energetic particles trapped within Earth's magnetosphere. On the other hand, magnetospheres, especially Earth's magnetosphere, also act as a shield by blocking the inflow of external energy and harmful solar energetic particles. To this end, missions are needed that will answer the outstanding questions on key boundary shielding regions as well as on particle energization.

Creation and Annihilation of Magnetic Fields

Creation and annihilation of magnetic fields is a theme of critical importance to the other two themes. Magnetic reconnection plays a significant role in the coupling between solar wind and magnetospheric plasmas as well as contributing to the complex dynamics within global magnetospheres. Magnetic reconnection also is an important means by which the magnetospheric shield is compromised and almost certainly plays some role in seeding magnetospheric energetic particles. Even in a magnetosphere such as Jupiter's, where the energization of the plasma and its circulation are derived principally from the rotation of the planet, reconnection is a critical process in governing the plasma circulation. While the creation and annihilation of magnetic fields occurs in microscopic regions, these microscale processes control and are controlled by macroscale processes. Therefore, we require projects that not only probe the fundamental plasma physics of magnetic reconnection (e.g., tight clusters of spacecraft) but also quantify the overall significance of reconnection (e.g., widely distributed constellations of spacecraft and distributed ground systems). It may be some time before we can send clusters of spacecraft to study planetary magnetospheres, but we should take advantage of multi-mission opportunities such as are arising at Mars at present, and such as the Cassini flyby of Jupiter while Galileo was operating.

PROJECT SUMMARIES

Before we can prioritize projects or even indicate what science they could achieve, we must know what these projects do. To this end, the panel summarizes in short descriptive narratives the projects to be prioritized and then presents mission details in Tables 2.5 through 2.8. These projects are in various stages of development. Those that have already had significant community planning by targeted science and technology definition teams are shown in italics in the tables; other missions that have had less formal planning, but enough to be endorsed broadly by the space physics community, are shown in roman (regular) type.

Anticipated and New NASA Solar Wind–Magnetosphere Projects

Pursuit of the themes in Section 2.3 requires multipoint measurements in two different forms. Some of the objectives require closely spaced measurements such as the proposed Magnetosphere Multiscale mission will provide (Figure 2.8). This mission examines fundamental processes at the gyroscale and will reveal how gyroscale processes couple into macroscale processes such as magnetic flux transfer and magnetic reconfiguration. Closely separated spacecraft will also allow us to quantify the nature and significance of turbulence as well as to measure the micro- and macroscale processes responsible for particle acceleration. An analogous mission at low altitudes (Auroral Cluster) would help understand auroral acceleration processes where space-time ambiguities impede scientific closure on outstanding problems.

We also need to capture the dynamics of the global system. Just as ground-based arrays can monitor the consequences of ionospheric plasma dynamics at the feet of the magnetospheric field, constellations of satellites are necessary to capture the sources of these events in the

TABLE 2.5 Anticipated and New NASA Solar Wind-Terrestrial Magnetosphere Basic Science Projects

Primary Science Objective	Measurement Objective	Mission/Facility[a]
Determine microphysics of reconnection, turbulence, and particle acceleration. Determine coupling of microphysics to mesoscale phenomena in Earth's magnetosphere.	Three orthogonal spatial gradients of three-dimensional plasma and energetic particle distribution functions, vector electric and magnetic fields, and electrostatic and electromagnetic waves from ion cyclotron frequency to electron plasma frequency. Four spacecraft with separations from 10 km to 2 R_E and time resolution of 0.75 sec in adjustable orbits that scan the magnetopause and magnetotail.	*Magnetospheric Multiscale (MMS)*—prioritized Solar Terrestrial Probe (STP) mission; study team report completed; phase A study begins late 2003; launch in early 2009
Understand the global dynamics of the coupled solar wind–magnetosphere system by determining the nonlinear dynamics, responses, and connections within Earth's structured magnetosphere, and by revealing the physical processes operating on spatial and temporal scales accessible to global circulation models.	3D plasma distribution functions, energetic particles, and vector magnetic field from 50 to 100 spacecraft between geostationary orbit and ~25 R_E geocentric distance. Science would be enhanced by distributed ground-based system.	*Magnetospheric Constellation (MagCon)*—next solar wind–magnetosphere STP mission as currently prioritized; study team report can be found at <http://stp.gsfc.nasa.gov/missions/mc/mc.htm>; launch in ~2012
Determine global response of inner magnetosphere during magnetic storms and magnetospheric substorms by imaging the proton and electron auroras, the plasmasphere, inner plasma sheet, and ring current with time resolution of 1 minute.	Image ENAs (1 to 500 keV), EUV at 30.4 nm, and FUV at 121.8 nm and 140–190 nm from two spacecraft in crossed Molniya orbits with apogee at 9 R_E while simultaneously making in situ measurements of plasmas and magnetic fields from equatorial orbit with apogee of 10 R_E.	Stereo Magnetospheric Imager (SMI)—in 1997 Roadmap
Define the size, shape, motion, and occurrence rate of the magnetopause boundary phenomena, which regulate the flow of solar wind energy through the magnetopause.	Vector magnetic field and plasma from a network of ~36 spacecraft skimming both sides of the dayside boundary layer over a wide range of latitudes and local times.	Dayside Boundary Constellation (DBC)—in 2000 Roadmap
Determine response of the magnetosphere during magnetic storms and magnetospheric substorms by imaging global magnetospheric total plasma density within 12 R_E of Earth.	Measure Faraday rotation of electromagnetic waves transmitted from Earth to 16 spacecraft in an elliptical equatorial orbit with apogee of 12 R_E.	Magnetospheric Tomography (MagTom)—see <http://sprg.ssl.berkeley.edu/ConstellationClassMissions/ergun.pdf>
Determine how the global topology of the magnetosphere responds to external forcing and instabilities during magnetic storms and magnetospheric substorms through remote sensing.	Image ENAs, EUV at 30.4 nm, and FUV at 121.8 nm and 140–190 nm from either of two spacecraft from fixed position at 50 R_E above north and south poles using solar sails or multiple orbiting spacecraft.	Geospace System Response Imagers (GSRI)—in 2000 Roadmap; also called Polesitter in 1997 Roadmap
Determine the energization processes in the auroral region where the magnetosphere is coupled to the ionosphere.	Particles and fields on a suite of small spacecraft flying in close formation.	Auroral Multiscale (AMS)—in Sun-Earth Connection 2003 Roadmap, which can be found at <http://sec.gsfc.nasa.gov/sec_roadmap.htm>

NOTE: Solar Terrestrial Probes, approximately $400 million, based on estimated cost caps in effect in late 2003.

[a] Projects with significant community planning by targeted science and technology definition teams are shown in italics; those with less formal planning, but enough to be endorsed broadly by the space physics community, are shown in roman (regular) type.

magnetosphere. While these spacecraft can be simply instrumented, many satellites are needed to cover the volume of the magnetosphere in which these events occur. The Magnetospheric Constellation (see Figure 2.9) and the Dayside Boundary Constellation are two such high-priority missions. These missions provide in situ measurements throughout large volumes.

Two complementary approaches addressing this same problem of probing the dynamics of the magnetosphere over a wide volume of space are magnetospheric imaging and tomography. Both the Stereo Magnetospheric Imager and the Geospace System Response Imager (or Polesitter) respond to this need via remote sensing as the next-generation multipoint auroral and neutral

TABLE 2.6 Anticipated and New NASA Solar Wind–Terrestrial Magnetosphere Targeted Science Projects

Primary Science Objective	Measurement Objective	Mission/Facility[a]
Determine the processes that energize ring-current plasmas and produce the radiation belt by establishing global reconfigurations of fields and distribution of particles in the innermost magnetosphere.	Measure three-dimensional particle distributions and fields from a small constellation of spacecraft in nested geosynchronous transfer petal orbits at various local times.	Geospace Probes (GProbes) (also appearing in a larger form as Inner Magnetosphere Constellation)—in 2000 Roadmap
Measure the upstream solar wind and interplanetary magnetic fields and their longitudinal variations associated with coronal mass ejections and interplanetary shocks.	Use three small spacecraft carrying plasma, magnetic field, and energetic-particle instruments in Earth synchronous orbit, close to the Earth-Sun line, but well inside 1 AU, powered by solar sails to gradually vary their mutual separation to investigate three-dimensional structures.	Solar Wind Sentinels—in 1997 Roadmap

NOTE: Living With a Star, approximately $400 million, based on estimated cost caps in effect in late 2003.
[a]See note *a*, Table 2.5.

atom imagers. While the former mission can be done early, technology development is needed for the latter if executed with a solar sail spacecraft.

Radio frequency tomography of density structures of the outer magnetosphere is another attractive approach whose technical readiness lies between that of the two imaging missions. A mature mission concept, Magnetospheric Tomography, would include up to 16 spacecraft (see Figure 2.10). A highly desirable conceptual merger of MagCon/DBC and MagTom approaches would allow for both remote sensing and in situ measurements.

TABLE 2.7 Anticipated and New NASA Solar Wind–Planetary Magnetosphere Projects

Primary Science Objective	Measurement Objective	Mission/Facility[a]
Determine the consequences of the lack of an ionosphere on the response of a magnetosphere to solar variability.	Three-dimensional plasma distribution functions, energetic particles, and vector magnetic field on Mercury orbiter.	Messenger—in 2000 Roadmap, now in development under Discovery program
Determine the relative contributions of planetary rotation and of the interaction with the IMF to jovian magnetospheric dynamics; determine how global electric and magnetic fields regulate magnetospheric processes; identify the particles responsible for the jovian aurora and their source region(s).	Measure particles and fields in situ in the auroral acceleration region, along L shells, and in the conjugate source regions from an elliptical polar orbit with perijove at 1.1 R_J and apojove at 15 to 40 R_J.	Jupiter Polar Orbiter—in 2000 Roadmap
Determine the role that extreme variations in internal magnetic dipole tilt have on the structure and dynamics of a planetary magnetosphere and on its response to solar variability.	Measure plasmas, energetic particles, vector magnetic and electric fields, and plasma waves and image Neptune's aurora from a moderate-inclination eccentric polar orbit.	Neptune Orbiter—in 2000 Roadmap
Determine the role of a strong internal plasma source to a rotating magnetosphere and its impact on magnetospheric dynamics.	Low mass particles and fields experiments in a 6 by 70 R_J equatorial orbit.	Io Electrodynamics (IE)—in 2000 Roadmap as a Frontier Probe
Determine how the upper atmospheres of terrestrial planets are influenced by the solar wind in the absence of a global magnetic field.	Measure neutral species escape rates, isotopic ratios, densities, temperatures, and winds; measure thermal plasma, energetic particles and magnetic and electric fields.	Venus Aeronomy Probe (VAP) and Mars Aeronomy Probe (MAP)—both in the 2000 Roadmap

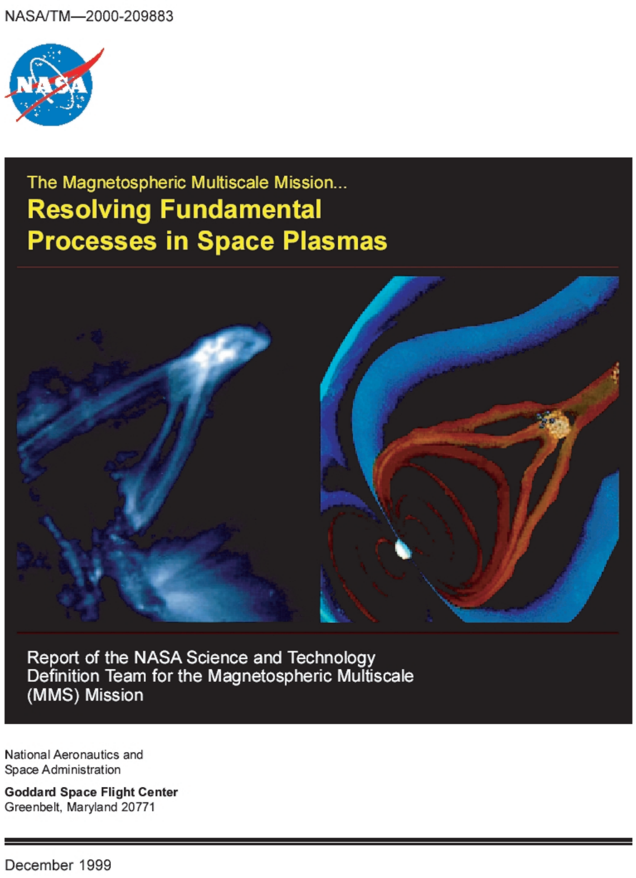
NOTE: See note, Table 2.13. The panel points out that in the solar wind–planetary magnetosphere area greater emphasis should be given to space particles and fields experiments as missions of opportunity on planned national and international planetary missions (e.g., Venus Express, Mercury Orbiter, and the ISAS mission to Venus). The panel also stresses the importance of PI-class planetary missions studying magnetosphere–solar wind interactions through the Discovery and Explorer programs. Future missions to Pluto, Europa, comets, and asteroids could bring much new understanding of the interactions of these bodies with their plasma environments for a small investment in instrumentation.
[a]See note *a*, Table 2.5.

TABLE 2.8 Anticipated New NSF Medium-Size Initiatives

Science Objective	Measurement Objective	Mission/Facility[a]
Determine response of high-latitude ionospheric convection, and hence magnetospheric convection, to varying solar wind input.	Measure vector ionospheric convection velocities throughout polar cap using incoherent-scatter radar located within 10° of magnetic pole. Relocatable for flexible use.	*Advanced Modular Incoherent Scatter Radar (AMISR)*—formerly the Relocatable Atmospheric Observatory
Determine global rapid time variations of magnetospheric currents to obtain density structure and to determine mechanism, time, and location of substorm onset.	Measure X, Y, Z components of surface magnetic field at a cadence of 1 Hz in large-scale arrays covering all latitudes.	Magneto-Seismology Array (MSA)

[a] See note *a*, Table 2.5.

A parallel effort in planetary magnetospheres is also critical to a successful program because these complementary systems allow greater insight into the fundamental processes of magnetospheric physics. For example, Mercury's magnetosphere (to be probed by Messenger), with its absence of a significant ionosphere and its small size, provides such comparisons.



NASA/TM—2000-209883

The Magnetospheric Multiscale Mission...

Resolving Fundamental Processes in Space Plasmas

Report of the NASA Science and Technology Definition Team for the Magnetospheric Multiscale (MMS) Mission

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

December 1999

FIGURE 2.8 Cover of the STDT report detailing the science and mission implementation strategies of the Magnetospheric Multiscale mission. Courtesy of NASA.

The jovian magnetosphere provides the contrast of a rotationally dominated magnetosphere and vast size; Jupiter Polar Orbiter will determine the relative contributions of planetary rotation and of the interaction with the IMF to jovian magnetospheric dynamics (see Figure 2.11). It will also determine how global electric and magnetic fields regulate magnetospheric processes, as well as identify the particles responsible for the jovian aurora and determine their source regions.

Neptune provides a magnetosphere ordered by a dipole field that has a very large tilt angle. The Neptune Orbiter mission will determine the role that extreme variations in the internal dipole tilt have on the structure and dynamics of a planetary magnetosphere and on its response to solar variability.

Io provides an intense source of mass loading to the jovian magnetosphere. The Io Electrodynamics mission will determine the role of a strong internal plasma source in a rotating magnetosphere and its impact on magnetospheric dynamics.

Venus and Mars Aeronomy Probes both show how ionospheres can hold off the flow of the solar wind for weakly magnetized bodies with atmospheres. These missions will determine how the upper atmospheres of terrestrial planets are influenced by the solar wind in the absence of a global magnetic field.

Finally, in the applied science area the panel supports the plans of the LWS program and those of other agencies to monitor and predict space weather. The Geospace Probes will determine the processes that energize ring-current plasmas and produce the radiation belt by establishing the global configuration and evolution of electric and magnetic fields and the distribution of particles of the innermost magnetosphere. Two pos-

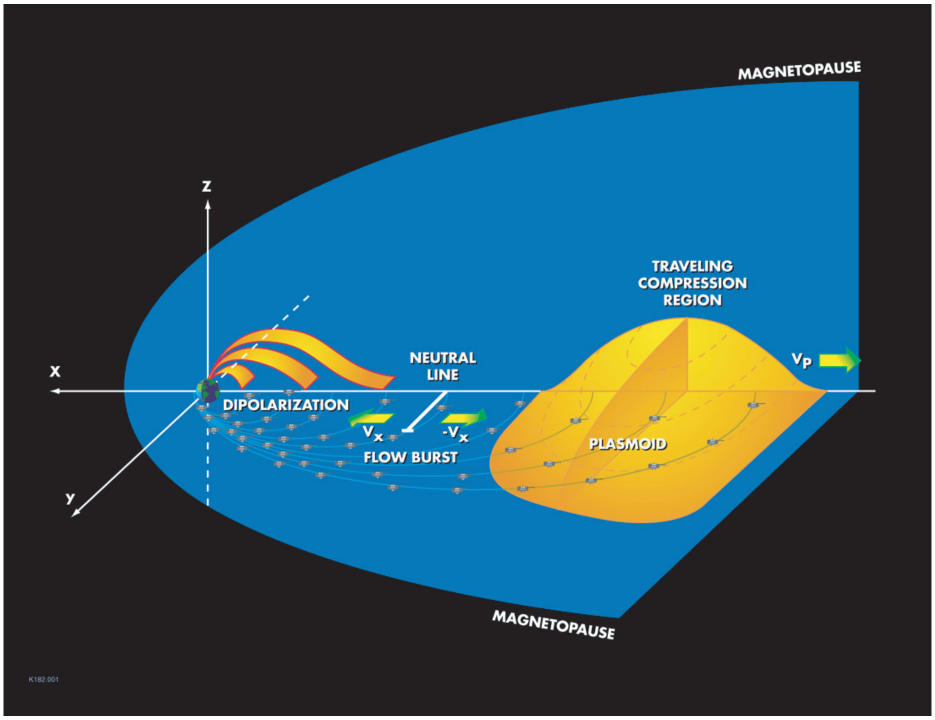


FIGURE 2.9 Artist's conception of the Magnetospheric Constellation capturing the dynamics of the magnetospheric plasma associated with a dipolarization of the near tail and the formation of a plasmoid. Courtesy of NASA.

sible configurations of a Radiation Belt Mapper mission are shown in Figure 2.12.

A key driver of terrestrial space weather is the variability in the solar wind. The Solar Wind Sentinels (SWS) mission will measure the upstream solar wind and interplanetary magnetic fields and their longitudinal variations both for normal conditions and also for conditions associated with coronal mass ejections and interplanetary shocks.

Anticipated New NSF Initiatives

Ground-based facilities provide data on the ultimate fate of the energy and momentum transfer from the solar wind and provide access to regions of the magnetosphere and parameters that are difficult to acquire from space. Our ability to probe the upper atmosphere and ionosphere from the ground with radar and optical instruments has expanded greatly over the last several decades. The most critical region of the magnetosphere for which such measurements are needed is the auroral and polar regions, at best difficult environments in which to work. A most economical way to obtain these data is with NSF's proposed Advanced Modular Incoherent Scatter Radar (AMISR). It would determine the response of the high-latitude ionosphere to the solar wind input by measuring both ionospheric convection and the heating of the upper atmosphere using incoherent scatter radar and optical instrumentation. It could be deployed at different latitudes and longitudes to address a variety of objectives as problems are solved and new ones are identified. While AMISR represents a large investment at a single site, there is also a need for well-dispersed inexpensive instruments. The epitome of such a device has been the magnetometer, which has long been used to measure magnetospheric activity. Recent advances in magnetometry (increased precision and lowered costs) and improved analytical techniques have enabled the magnetometer to make measurements of fundamental importance to magnetospheric physics through magnetoseismology. The attempt to use ULF pulsation data to remotely sense plasmaspheric mass properties has a long history. Since the mass density of the magnetosphere is critical to the speed at which waves travel through the magnetosphere and to the speed at which the magnetosphere can react to stresses, it is critical to

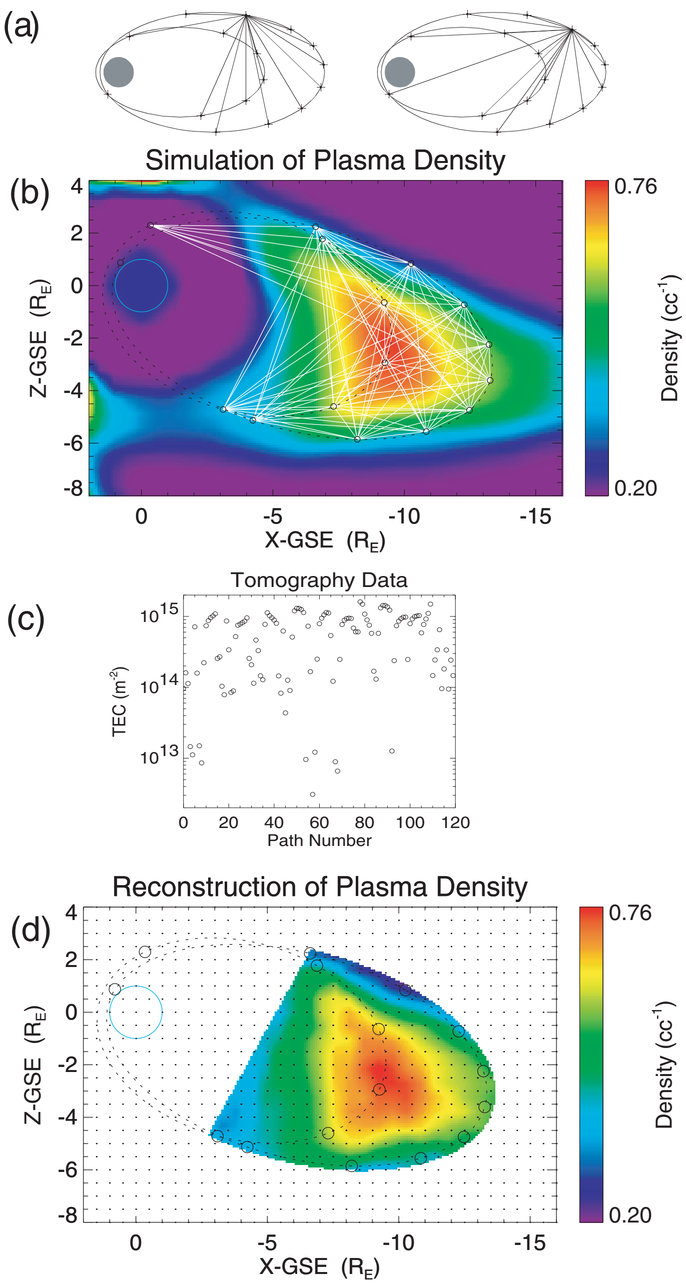


FIGURE 2.10 A simulation of the density maps that the Magnetospheric Tomography mission could make using radio tomography techniques in Earth's magnetotail. Courtesy of R.E. Ergun, University of Colorado at Boulder.

FIGURE 2.11 The Jupiter Polar Orbiter mission will address key questions left unanswered by the Galileo mission. Courtesy of NASA.

know its mass distribution. However, most of the plasma is cold and often invisible to space probes. Fortunately we can now probe the equatorial mass density with local techniques via the measurement of resonant frequencies, much as solar oscillations allow the Sun to be seismically probed. Also, sudden increases in solar wind pressure can be used, in a manner akin to earthquakes, to probe the density structure of the magnetosphere through the inversion of travel times. These two techniques can be used with inexpensive latitudinal chains of modern GPS-synchronized magnetometers. The panel recommends that a Magneto-Seismology Array for both types of density sounding be established.

Finally, the panel notes the utility of irregularities that reflect radio signals; these irregularities occur often enough to provide a useful tool for studying plasma circulation.

SCIENCE TRACEABILITY

In this section, the panel illustrates the ways in which projects are directly responsive to our science questions and themes. Tables 2.9 to 2.11 show the flow from the three science themes to specific science questions and then demonstrate how we will reach closure on these questions with a systematic suite of carefully planned projects. The projects identified by name in the tables are prioritized and detailed in the next section. Projects that address primarily a specific science question are denoted by a P. Those that provide secondary support are denoted by an S. Owing to the broad nature of the themes, several projects are often required to fulfill the overall theme goal. Accordingly, they may take a decade or so to be completed, principally as a result of funding constraints rather than science drivers. Indeed,

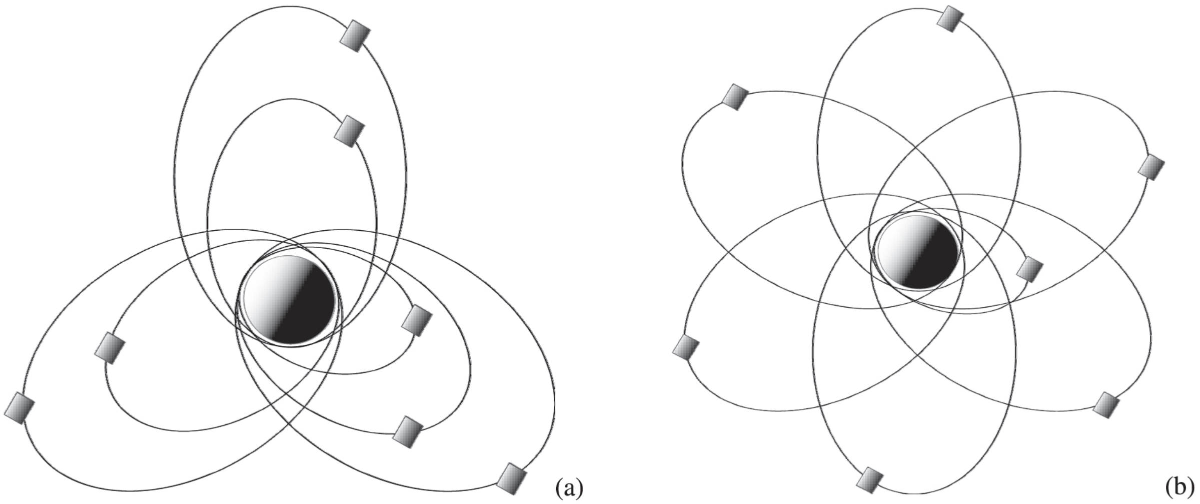


FIGURE 2.12 Two possible Radiation Belt Mapper constellations: Orbit apogees extend just beyond geosynchronous orbit, and perigees are low; orbit planes lie in Earth's equatorial plane in both options. Courtesy of NASA.

the panel stresses that there will be significant additional value if several projects can be done concurrently rather than serially.

PRIORITIZATION: NASA AND NSF

Next, the panel prioritizes the projects listed in Tables 2.9 to 2.11. In assigning priorities, it used the following criteria:

- Importance of the scientific problem,
- Technology readiness,
- Adequacy of the theoretical foundation,
- Likely impact, and
- Appropriate timing of related projects.

The panel has sorted the priorities into several categories, appropriate for the different funding agencies and programs under which such projects would fit most naturally. These prioritized lists follow in Tables 2.12 to 2.14.

The panel also lists here several additional projects that are generally supportive of all its science themes. These missions would manifestly contribute to the science goals of the missions in Tables 2.9 to 2.11, by providing ancillary observations of the solar wind or global auroral imagery. Such missions may be motivated by programmatic issues, but the panel stresses that they would be (and others before them have been) vitally important for achieving the most possible science. The prioritized list of interagency missions follows in Table 2.15.

PRIORITIZATION OF OTHER AGENCY AND INTERAGENCY INITIATIVES

The panel recommends that each agency continue its individual space environment monitoring that serves not only the agency itself but also the space science community at large. Such missions include NOAA's GOES, USAF's DMSP, and DOE's LANL geostationary spacecraft programs. These missions form an invaluable resource for the science community by providing ancillary data that complement other large programs. The panel endorses and encourages the further collection, archiving, and timely dissemination of these data.

Furthermore, the panel recommends conducting the large anticipated missions listed in Table 2.15, all of which may require strong interagency cooperation. While these missions will primarily serve to monitor the space environment, in keeping with their principal objectives, their scientific value is great also. In particular, they may help to provide critically needed solar wind and interplanetary magnetic field monitoring as well as

TABLE 2.9 Magnetospheres as Complex, Coupled Systems

	Project															
Science Question	MMS	GProbes	MagCon	SMI	DBC	SWS	MagTom	GSRI	AMS	JPO	NO	IE	VAP	MAP	RAO	MSA
What is the global structure of magnetosheath fields and plasmas?	P	S	P		P		S									
What are the causes of terrestrial magnetospheric activity, including substorm onset and recovery?	P		P				P		S							P
What is the connection between terrestrial substorms and magnetic storms?		P	S	P			S	P								P
How does the mid and distant tail map to the planet magnetically and dynamically?			P		S		P	S		S						P
What is the role of the terrestrial plasmasphere in modulating ring current and radiation belt particles?		P		P				P								S
What are the scale sizes of geoeffective solar wind structures?					P	P										S
What is the role of turbulence in Earth's magnetosphere and at its boundaries?	P	S	P		P		P		P						P	
Do solar-wind-driven phenomena occur in other magnetospheres?										P	P	S	P	P		
What is the relative importance of various sources for populating the terrestrial plasma sheet?			P				S					P				

NOTE: P denotes projects that address primarily a specific science question; S denotes projects that provide secondary support.

TABLE 2.10 Magnetospheres as Shields and Accelerators

	Project															
Science Question	MMS	GProbes	MagCon	SMI	DBC	SWS	MagTom	GSRI	AMS	JPO	NO	IE	VAP	MAP	RAO	MSA
What are the wave mode transitions in Earth's magnetosheath and how sharp are they?	S		S		P											
How are shocks modified by multiple components?												S	P	P		
What are the origin and destiny of planetary ring current and radiation belt particles?		P	S	P				P		P						S
What is the nature of convective energy transport and geomagnetic activity?		P	P				P								P	P
What is the role of the solar wind in plasma loss at weakly ionized bodies?												P	P	P		
What are the origin and dynamics of Jupiter's aurora?										P						
What is the dynamic evolution of auroral acceleration?								P	P	P					P	S

NOTE: P denotes projects that address primarily a specific science question; S denotes projects that provide secondary support.

global auroral imaging—both tasks are as of yet only assumed but will be essential to many of the basic and targeted science missions in Tables 2.12 to 2.14.

2.6 TECHNOLOGY

INTRODUCTION

Many of the solar wind–magnetosphere interaction science studies and science missions, discussed in earlier chapters, can be achieved only by applying new technology or developing enabling technologies. The areas in which new technologies are required cut across all aspects of space science and space engineering. Even many of the near-term science missions in the current program discussed in Section 2.5 require technologies that are not yet space qualified or qualified for flight. There are five broad areas where investment in new technology is required to meet the science and mission needs identified in this study:

- Propulsion (e.g., solar sails, solar electric, and nuclear propulsion);
- Spacecraft technology (e.g., miniaturization to microsat/nanosat levels, advanced communications, advanced power systems, autonomous operation, mass production, development of a "sciencecraft");
- Science instrumentation (e.g., miniaturization, advanced detector technology, mass production);

TABLE 2.11 Creation and Annihilation of Magnetic Fields

	Project															
Science Question	MMS	GProbes	MagCon	SMI	DBC	SWS	MagTom	GSRI	AMS	JPO	NO	IE	VAP	MAP	RAO	MSA
What is the interplay between micro-, meso-, and macroscale aspects of reconnection?	P		P		P				S	S						
What are the relative roles of component and antiparallel merging?	P		P				P			S						
What is the global reconnection rate on the magnetopause and how does it vary?			S		P					S					S	
Is reconnection patchy or continuous, inherently unstable or quasi-static?	P		P				P		P	S					S	

NOTE: P denotes projects that address primarily a specific science question; S denotes projects that provide secondary support.

- Information architecture (e.g., data synthesis and assimilation, model development, multipoint data visualization); and
- Ground systems and operations (e.g., advanced operations and data handling systems).

All of these are driven by the science requirements and mission objectives spelled out above. In many cases, the science and mission objectives cannot be met using existing technologies. Each of the five areas is discussed in more detail below. For example, the science drivers and mission concept for the Magnetospheric Constellation mission require the use of a large number of very small spacecraft/sciencecraft. The technology that is required to fulfill this mission alone covers a broad range, from new instrument, satellite, thruster control, and mechanical assembly technology to new operations capabilities and data handling and assimilation technologies.

TABLE 2.12 Prioritized NASA-led Missions

Mission	Program
Magnetospheric Multiscale (MMS)	NASA Solar Terrestrial Probes (STP)
Geospace Probes (GProbes)	Living With a Star (LWS)
Magnetospheric Constellation (MagCon)	STP
Stereo Magnetospheric Imager (SMI)	STP
Dayside Boundary Constellation (DBC)	STP
Solar Wind Sentinels (SWS)	LWS
Magnetospheric Tomography (MagTom)	STP
Geospace System Response Imagers (GSRI)	LWS
Auroral Cluster (AC)	STP

NOTE: Approximately $400 million.

PROPULSION TECHNOLOGY

Many current and future missions require development of new propulsion technology, enhancement of existing technology, or the creation of new systems by the marriage of older and newer technologies. For example, new technologies are needed to allow satellites to maintain (non-Keplerian) trajectories that are not possible without continuous thrust being applied. Examples of missions that require such technologies are the Solar Wind Sentinels and the Polesitter missions, which must "thrust" to maintain a proper position relative to the Earth-Sun line and over Earth's polar regions, respectively.

TABLE 2.13 Prioritized NASA Planetary Missions

Mission	Program
Jupiter Polar Orbiter (JPO)	New Frontiers or Discovery
Neptune Orbiter (NO)	New Frontiers
Io Electrodynamics (IE)	New Frontiers or Discovery
Venus Aeronomy Probe (VAP)	Discovery
Mars Aeronomy Probe (MAP)	Discovery or Mars Scout

NOTE: Cost caps in FY 2003 for New Frontiers, Discovery, and Mars Scout were <$700 million, <$350 million, and <$350 million, respectively.

TABLE 2.14 Prioritized NSF Facilities

Mission	Program
Advanced Modular Incoherent Scatter Radar (AMISR)	NSF
Magneto-Seismology Array (MSA)	NSF

Solar sails are a possible technology for doing this. Solar electric propulsion might also achieve the same effect. The limitation of solar electric propulsion is that the satellite must carry fuel for the system, although it would have much less mass than that required by a chemical fuel system. Solar electric propulsion has been spaceflight-qualified by the New Millennium program, but the technology requires further development. Solar sails are an unproved technology that at this point would require significant investment before it is ready to support the missions mentioned. In fact, a privately supported mission may be the first to use a solar sail in space.[3]

Deep space and outer planet missions could use solar sails or a marriage between nuclear power and electric propulsion. Missions that must go very close to or very far from the Sun could take advantage of such propulsion. However, there has recently been significant reluctance to use nuclear power on satellites and there has been no attempt to develop nuclear electric propulsion. The very deep space missions have no choice but to use nuclear power for the instruments and avionics. Extending nuclear power to the propulsion system would seem to be a natural evolutionary step. The combination of nuclear power and electric propulsion would enable deep space missions, such as a mission to the heliopause, to be completed on much shorter time scales than otherwise possible. Some relatively high power nuclear systems have been flown in the past so some technology exists but they are high-mass systems that impose a heavy-lift launch vehicle requirement on missions that use them. The recently announced initiative by NASA to develop nuclear power systems for space is clearly a step in the right direction.

In addition to the above, it may be possible to combine heat engines with electric propulsion for long-duration, near-Sun missions, once the satellite arrives near the Sun (having been launched by conventional rockets). However, no such technology is being considered at this time.

At this point, solar sail technology is being studied by NASA because it is capable of achieving large velocity changes and requires no fuel. It also meets many mission requirements.[4] Many studies of solar sail use have been completed or are in progress. Significant investment is required to develop prototypes and obtain flight experience in time to support science missions. Ultimately, one would like to have a power sail that would provide not only the propulsion but also the electric power necessary for satellite operation.

Another area where propulsion technology needs development is the support of constellation-class missions that require numerous micro/nanosatellites to be placed or scattered throughout a region to obtain the required spatial coverage. Most current concepts require that each satellite have a thruster to obtain its final position, with the bulk of the energy provided by a mother or dispenser ship. While multiple satellites (<10) were dispersed from single launchers in the past (i.e., Iridium and Globalstar), those missions did not have the kind of satellite distribution requirements that are envisioned by a mission such as the Magnetotail Constellation (MagCon) mission. At present, the MagCon dispenser ship is a concept only. Significant effort is required to develop dispenser ship technology by the end of the present decade, when MagCon is scheduled to fly. Similar propulsion difficulties exist for other Living With a Star missions, such as the Radiation Belt Mappers, which envision micro/nanosatellites in multiple "petal" orbits. Such orbits are not easily obtained from a single chemical booster. Missions like this may require a dispenser ship, possibly with solar electric propulsion, to distribute the constellation of satellites properly.

---

[3] Mark Alpert, Sailing on sunlight, *Scientific American,* 285, 1, 2001.

[4] NASA, OSS. 2000. *Strategic Plan.* NASA, Washington, D.C.

TABLE 2.15 Prioritized List of Anticipated Interagency Projects (<$200 million)

Science Objective	Measurement Objective	Mission/Facility	Lead Program
Monitor the upstream solar wind and interplanetary magnetic fields as well as continuously observe the emissions from the daylit hemisphere of Earth for global change studies.	Remote sensing and in situ plasma and fields instrumentation on spacecraft at forward Lagrange point.	Triana—spacecraft essentially complete and awaiting a launch to L1 point	NASA
Measure the far upstream solar wind and interplanetary magnetic fields associated with coronal mass ejections and interplanetary shocks.	Measure solar wind plasmas and fields well upstream of the L1 point using solar sail technology to provide earlier warning for approaching geoeffective solar wind structures.	Geostorms—several concept studies completed	NOAA/NASA
Monitor global auroral zone emissions to quantify global energy inputs from magnetosphere to ionosphere and to complement constellation-class missions.	Remote sense multiple wavelength auroral emissions relevant to both electron and proton precipitation.	Auroral Monitor	NOAA/NASA

So far, the focus of constellation missions has been to modify existing technologies to come up with a dispenser ship. However, the large number of satellites involved means significant development. Test flights of any such ship would be required before it can be committed to carry a large constellation of satellites.

Another technical challenge for constellation-class missions is the ability to track a large number of satellites and return their measurements to the ground. Early, limited experience with operating small numbers of spacecraft in a cluster suggests that the operational aspects of a constellation mission may be very different from the operational aspects of traditional science missions.

Finally, there is a need for new propulsion technology to support upper atmosphere and lower ionosphere probing or "dipper" missions. Such missions require relatively high-thrust, low-mass propulsion systems. Current systems are too low in thrust or too high in mass. They can support only very limited duration missions.

SPACECRAFT TECHNOLOGY

The major spacecraft technology advances required to meet identified science goals are in micro/nanosatellite development, autonomous and robust satellite operation, and long life. The need to reduce mass and power extends across all space science missions. Technologies that do so enable or enhance the missions. For example, reduction of satellite mass reduces propulsion requirements, which itself enables some missions. Similarly, technology that makes satellites robust and more autonomous reduces the need for operational support.

The panel has identified several science missions that require multiple distributed satellites. For example, the Geospace Mappers (Ionosphere and Radiation Belt Mappers), Dayside Boundary Constellation, Auroral Cluster, and Magnetospheric Tomography missions all require multiple (4 to 16) satellites with a range of capabilities, and the MagCon mission requires the largest number (~60 to 100). All require the development of new multisatellite technology capabilities.

The MagCon mission is an example of a science mission that is driving nanosatellite technology development. It requires multiple simultaneous in situ observations throughout a specific region of space. This drives technology on many fronts. The main challenge is to minimize all satellite subsystems from the mass, size, power, and cost perspectives. Thus, each satellite subsystem function must be examined with an eye to reducing it, combining it with other functions, or eliminating it if possible. In addition, to generate such a fleet of nanosatellites on a normal development schedule requires that they be mass produced and tested much as is done today for limited runs of specialty automobiles. For example, the Iridium satellites were produced using assembly-line techniques. ICO, which recently acquired a majority stake in bankrupt rival Globalstar, is an example of a company having a multisatellite program. However, ICO and Globalstar produced relatively large and expensive satellite systems. Furthermore, both companies were well capitalized.

MagCon requires the development of small and cheap platforms. This challenge may require abandoning the usual concept of building a bus to accommodate science instruments developed and fabricated by multiple institutions. Instead, the instruments and satellites would be developed as a system to be produced by one fabricator. This would be a significant departure from the usual process and would require much new process technology. The phase B portion of such development takes on a greater importance, because there is no room for fixing things downstream in phases C and D of such programs. Considerable investment in new management and design concepts and technologies is required now if large numbers of nanosatellites for constellation missions are to be successfully implemented. Any new management structure must have the instrument PIs as an integral part. The instrument PIs have the ultimate responsibility for sensor design and must work closely with the satellite (or sciencecraft) fabricator to implement the design and support the development of the data handling and analysis processes.

To reduce the mass and power requirements of satellite subsystems means looking to new components and technologies such as microelectromechanical systems (MEMS) gyros and Sun sensors for attitude determination and microthrusters (thrusters on a chip) for attitude control, where necessary. It also means developing much denser, more capable, and lower-power satellite data processing and control subsystems. Technologies are being developed in the commercial world that could be adapted to space systems. Much effort is currently being put into developing low-power, high-density, and high-speed application-specific integrated circuits (ASICs). These run the gamut of analog ASICs, digital ASICs, and hybrid analog/digital ASICs, and many of them are finding their way into science instruments and satellite systems. Again, investment in these technologies is required to successfully achieve significant reductions in satellite mass and power resource requirements. It should be noted that such technology would enable a range of space science and engineering missions cutting across the whole space science enterprise.

Ground station and personnel costs are a significant part of a mission's total cost and extend over long periods. Reduction of cost and simplification of operations requires robust, long-lived satellites, preferably autonomous to a great degree. Autonomy is primarily a system and software robustness issue. Development of such systems is a long-term endeavor and requires continued investment to advance the technology even as the spacecraft and instrument capabilities grow. The autonomous operation of satellites relies on technology that could benefit many missions.

For deep-space missions, new communications technology is a pressing need. Inflatable antenna structures may be an enabling technology for such missions insofar as they reduce the ground or NASA Deep Space Network (DSN) system resources required to support the missions. Such antennas could provide the signal gain needed to allow use of very-low-power transmitters for near-real-time measurements in support of the LWS program. The high-gain signals could allow the LWS program to use relatively cheap and small dedicated ground receivers instead of DSN to support this component of the program. It must also be recognized that some expenditure of technology development funds for the space-based portion of the communication link can alleviate growing pressure (oversubscription) on the ground station. The use of Ka or even optical communications can greatly increase the bandwidth, reducing contact times, but often these systems are dropped because of their initially high development cost.

For near-Earth missions, cell-phone-like communication links using commercial near-Earth satellite constellations (such as Iridium or ICO) to retrieve data and support satellite operations would be an enabling technology. Investment is required to solve the problems imposed by limited reception range and Doppler shifts in signal frequencies between the moving platforms. However, commercial communications systems could be a cheaper, more flexible way to support and operate new science missions such as the Ionosphere Mappers. Satellite-to-satellite communications technology would be enabling for multisatellite or constellation missions. NASA's Space Technology 5 program has this as one of its technology development goals.

Greater use of GPS signals to obtain near-Earth satellite ephemeris on board is an enhancing technology that could reduce the need for tracking and ranging to obtain ephemeris data. Such technology supports multiple near-Earth missions, even beyond the GPS orbital altitude. GPS could also provide the very precise timing signals needed to synchronize simultaneous measurements from constellation spacecraft and to perform interferometry. Again, some investment is required before such capabilities become routine and reliable enough to become a common element of science mission design.

There are several other satellite technologies in development that should be sustained or accelerated to meet the needs of new science missions. While these technologies are not required, each offers capabilities that can be mission enhancing. They include the development of lithium-ion batteries and charging systems; active thermal control using electrosensitive surface materials and coatings; miniature star trackers; MEMS gyros and Sun sensors for attitude determination; and new, deployable structures that unfold and become rigid support elements. Each of these is under development in industry and provides a capability that can be used by several science missions. Development progress should be monitored and supported.

SCIENCE INSTRUMENTATION TECHNOLOGY

Science instrumentation struggles with the constraints imposed by the basic physics of measurements and the need to obtain statistically meaningful data. This difficulty limits the ability to easily miniaturize all sensors. However, there are several technologies in development that could reduce instrument resource requirements. For example, the expanded use of ASICs is dramatically reducing instrument complexity and power while increasing capability. Qualified high-density ASICs lead to more reliable instruments purely because there are fewer parts. Many are in development and promise to reduce power requirements by orders of magnitude while increasing performance by similar orders of magnitude. These developments need to be identified and supported so the technology is spaceflight-qualified in time to meet science mission needs. Generally, the ASICs, being application specific, are developed for a particular type of processing, such as fast Fourier transforms for wave data and time-of-flight systems for particle sensors. However, such needs cut across all space science missions, and an ASIC developed for an interplanetary wave instrument can be applied in a wave instrument elsewhere, for example.

The advancement of sensor and detector technologies also is a strong driver. The physics of the devices that make the measurement, be it of electromagnetic energy, photons, or particles, becomes the limiting factor for the science. Advances in sensor technology are critical to all science missions. Examples of such sensor development are the δ-doped charge-coupled device (CCD) arrays, both as normal CCDs for imaging systems and as pixelated detectors for miniature low-energy particle detection systems (such as solar wind spectrometers). While these hold much promise for increasing sensor capability and reducing sensor complexity and size, especially for particle detector systems, they require significant investment and effort to make them as accessible as off-the-shelf detector elements like microchannel plate detectors, for example.

Uses for MEMS technology in sensors are evolving. For example, MEMS devices could be used to create adaptive sensor geometries that allow an instrument to respond to current conditions, such as highly variable particle or light fluxes, by modifying the effective gathering power of the system. This could reduce the number of sensors required to make a measurement under conditions where there is a large dynamic range of possible input. Such capability is important for missions of discovery or exploration where the dynamic range of measurement parameters that will be encountered is unknown or where best estimates exceed the dynamic range of current sensors. At present, multiple sensors are often flown to cover the full measurement range. Missions to planetary magnetospheres and to the heliopause and missions near the Sun (Solar Probe) are the types of missions where such capability could mean measurement success instead of failure.

Smart processor technology can be enabling for some science measurements. One development that has shown promise is the use of processor technology based on field-programmable gate array (FPGA). This is a fast and adaptive technology wherein program logic is executed in FPGA hardware. The adaptability comes from the ability to reprogram the FPGA hardware and obtain processing speeds that are not possible from a software-driven system. Work has been done to develop compilers to convert programs into FPGA circuit designs. This technology could have application in spacecraft systems where fast processing and programmability are necessities, such as some deep space missions. At present, these systems are somewhat bulky and power hungry. However the concept is good and the technology is advancing. Continued support is required to maximize the capability of such systems.

In many cases, significant investment will be required to continue the evolution of sensor technology. The basic technology exists but must be refined and adapted to the kinds of measurements that need to be made. Time-of-flight sensor technology for plasma and energetic particle composition measurements is one such development. The technique is well known but difficult to implement. Development of a time-of-flight electronics in ASIC form is a goal for the space plasma community, which would bring this powerful technique to missions that do not have sufficient mass and power resources to use it now. As electronic capabilities and material technologies evolve, the science output from missions increases. This is true for optical, magnetic field, and plasma wave measurements also. There must be in place a process to identify the promising technologies and to support their migration into sensor systems. This means there must exist a cross-enterprise sensor technology development program that provides the support to do the migration and requires that the results be made accessible to all instrument builders.

INFORMATION ARCHITECTURE TECHNOLOGY

As the density and complexity of information increase, science missions are driven by the need to handle and ingest the resultant data. New technologies and concepts are required to make massive amounts of data accessible to science and scientists. The science and engineering community, with support from NASA, needs to continuously evolve this process. There is no clear-cut separation between data processing, data compression, data retrieval, data storage, and data distribution and assimilation onboard the spacecraft and like processing on the ground. Different missions will require that different parts of the process be done in different places. For example, a deep space or outer planet mission may send back mostly highly processed data to reduce bandwidth requirements. In future missions there may be a need to assimilate the observational data into models onboard the spacecraft and telemeter the results to the ground. The different technologies used must be migrated to the appropriate platforms.

The constellation-type missions, in particular, have a recognized need to assimilate the data into global models. This requires development of software and systems that can take the data inputs in real time independent of their quality and completeness. The concept is to use adaptive physical models (like adaptive MHD codes) to provide connectivity between the independent in situ observations and to generate a complete moving picture of the dynamic system being studied. This is not a current capability except for single-point interplanetary data inputs to MHD codes. What the constellation science teams envision is to dynamically modify the MHD codes to best represent the totality of the observations. Data from constellation missions impose constraints on the modeling codes. An intensive effort is required to construct codes with appropriate data assimilation technologies.

Model development itself is driven by a need for new technology, in some cases a new computational technology. In other cases it is the development of new software technology and numerical techniques. Model development is a cross-enterprise issue, and theoretical and modeling missions should be considered to be as important as experimental missions.

TECHNOLOGY FOR GROUND SYSTEMS AND OPERATIONS

As noted above, the cost of ground systems and operations can be very large, especially for long-duration missions. Technologies that reduce and simplify the systems free resources that can be redirected into the science component of the missions. Just as autonomous satellites can reduce the burden on ground and operations systems, autonomous ground systems can reduce the manpower required for mission support. With several upcoming missions being multisatellite or constellation missions, it is imperative that new operation concepts and technology be developed and implemented. Examples of possible technologies can be found in commercial communications satellite organizations. Many of these companies run large numbers of satellites with relatively small crews. In fact, these organizations are ahead of NASA in the area of autonomous operations of systems of spacecraft. Their technologies should be studied and emulated where appropriate; otherwise they should be used as a starting point for the development of the operations and ground systems that will be required by MagCon and other multisatellite science missions.

A separate issue is the retrieval and handling of the science data from modern missions. Again, in the past this has been a manpower-intensive and costly effort throughout the mission. Intensive support is required to find and implement architectures and technologies able to handle the massive increase in data that new missions will generate. New technologies are required to make such massive amounts of data easily accessible by all scientists. The science and engineering community as a whole needs to continuously evolve this process. The different technologies must be migrated to the appropriate platforms.

RECOMMENDATIONS AND PRIORITIES

As noted above, there needs to be a significant focus on developing new instrument, satellite, propulsion, operations, data assimilation, and processing technologies. The top priority in each of these areas is addressed below:

- *Science measurements.* A space-science enterprise-wide instrument development program is needed that is separate from SR&T budgets. This issue clearly needs to be addressed if quality measurements are to be made on the smaller micro/nanosatellites being envisioned for future multisatellite and constellation missions. In addition, new management and team structures must be generated for developing highly integrated, micro/nano, science-craft-type spacecraft. This must be done on a time scale that meets the development needs and implementation of multisatellite missions.

- *Propulsion.* The top priority is to push the development of solar sail technology to support missions that require non-Keplerian orbits in the interplanetary medium to make the necessary science measurements at appropriate locations. NASA has a good start in this area but needs to move forward with hardware development and spaceflight demonstrations so the technology is flight qualified in this decade.
- *Satellite technology.* The push should be to develop fabrication, integration, and testing technology for micro/nanosatellites that will enable constellation mission science. This requires bringing together new material and electronics technologies, new management techniques and structures, and new integration and testing processes. A way of integrating the development of spacecraft and science instruments is needed. Test-bed-type development programs that take multiple micro/nanosatellite systems from concept to flight should be considered as one way of developing and testing the new processes. This work needs to be done soon if it is to support missions that are identified in NASA's Strategic Plan.
- *Operations and data handling.* The focus should be on implementing evolving technologies that reduce manpower and costs for all missions. In particular, attention should be given to the complete data chain, from the operations required to get the science data to the manner in which the data are brought to Earth, assimilated into models, and ultimately presented to scientists. This must be started now if the needed capabilities are to mature on a time scale that meets planned schedules for complex new interplanetary observatories and multisatellite constellation missions.

2.7 SOLAR WIND–MAGNETOSPHERE INTERACTIONS: POLICY ISSUES

INTRODUCTION

The need to observe coupled dynamics of the large magnetospheric system in response to solar wind conditions has several implications. First, the difficulty of adequately specifying the state of the dynamic system means that the number of observing platforms should be as large as possible. This goal would be best achieved by interagency coordination to maximize the opportunities for instrumentation on non-NASA platforms. Second, because ground-based observations provide distributed knowledge of convection and boundaries that cannot be achieved in space, ground-based capabilities must not only be sustained but enhanced, and coordination between space- and ground-based observations needs to be exploited worldwide to the fullest extent possible. Third, it is now firmly established that knowledge of solar wind conditions is critical to developing further understanding of the magnetosphere-ionosphere system, so that sustained, continuous monitoring of solar wind input is essential. Fourth, global physics-based computer simulation codes are an essential component of the research program because they will play a central role in maximizing the information that can be extracted from the observations and unifying disparate data sets within a comprehensive framework. These implications in turn have specific ramifications for the policies that should be pursued to achieve the key science objectives identified above.

INTERAGENCY COORDINATION

Because the resources required to make the necessary observations exceed those available from any one agency and because the societal impact of the science return is relevant to a variety of agencies and interests, efficient coordination between agencies is a preeminent policy concern. In fact, one could argue that the coordinated system is much more valuable than the sum of its components.

NOAA: Transitioning New Operational Observing Platforms and Models

The National Oceanic and Atmospheric Administration has two roles to play. First, the transitioning of space instrument platforms from basic science research to operational systems needs to be anticipated and implemented in a timely fashion. Since space-based science platforms are almost exclusively the substance of NASA programs, these transitions will require coordination with NASA. Typically, this is done by taking the key instrumentation and data reduction techniques developed under NASA research programs and implementing them under NOAA. Observations that are or should be planned for transition include the following:

- *Interplanetary magnetic field and solar wind observations analogous to those provided in real time from* *the ACE spacecraft at the L1 point*. The necessity for continuous IMF/solar wind observations from L1 has been abundantly demonstrated scientifically and operationally. Nearly all predictive models of magnetospheric response depend principally on IMF/solar wind inputs. Since the ACE spacecraft is operating beyond its design life, it would be prudent to implement a new L1 platform for this purpose in the very near future (<2 years). Steps should be taken to continue such L1 in situ monitoring as an operational system.

- *Solar coronal observations*. The dramatic advance in our appreciation and understanding of the causative link between coronal dynamics, coronal mass ejections, and high-speed streams, in particular, and major geomagnetic disturbances made possible by results from the instrumentation on YOHKOH and SOHO has motivated both the Solar Dynamics Observatory of the LWS program and the STEREO mission. There is now little doubt that observations of this class will play a central role in operational space forecasting, and steps should be taken to coordinate operations with NASA in the short term (next 5 years) and to deploy a line of operational monitors in the medium term (5–8 years). The GOES SXI instrument is an important first step.
- *Auroral imaging*. The advances in understanding magnetospheric dynamics, particularly nightside/magnetotail processes made possible with global auroral imaging, demonstrate the value of these observations for monitoring intrinsic magnetospheric dynamics as well as energy transport to the ionosphere. New results from the IMAGE mission promise to increase our understanding of the physical correlates of these observations, thereby improving their operational value. It is already clear that real-time auroral imaging will prove operationally valuable, and plans should be laid for NOAA to provide global auroral imaging on an operational platform in the long term (8–10 years).

The second role for NOAA concerns the transitioning of models from science research tools to operational resources. NOAA, NSF, DOD, and NASA all have theory and modeling efforts in magnetospheric physics. The Space Environment Center of NOAA supports a small effort to transition science models to operational use. Discussions with both agency and scientific community personnel indicate that this transition effort is undersupported. They note that there appear to be many more models available in the community that could be useful to space weather than are being actively converted to operational status.

A parallel concern exists in the research community, which has a separate need for standard models. In the next decade, the research community also will require access to realistic global models, including MHD simulations. The Coordinated Community Modeling Center at GSFC[5] is providing the first community access to such global models, in part in coordination with the NSF's GEM program. The challenge of providing sophisticated models to the community is a significant one and will require an advisory structure to prioritize and ensure maximum efficiency in coordination between agencies and to minimize duplication of effort. The panel recommends that these transitioning efforts be supported aggressively to meet the science and applications objectives of the space environment community.

DOD-DOE: Coordinated Planning for Launch and Flight Opportunities and Access to Relevant Data Sets

The Department of Defense and the Department of Energy conduct operational flight and observation programs that are directly relevant to the science objectives of magnetospheric–solar wind interactions. Ensuring appropriate use of these resources is an extraordinarily cost-effective means of achieving several of the observational goals described above. Launch opportunities will be available, particularly for sending smaller payloads, ~300 kg, into geosynchronous transfer and low Earth orbits on DOD vehicles. These are key regions for magnetospheric dynamics, particularly radiation belt dynamics. Historically, launches of opportunity have proven problematic in practice because of cost concerns associated with launch schedules. (The cost growth and resulting cancellation of IMEX were due in part to problems of this sort.) Mechanisms for accommodating NASA payloads on DOD launch schedules without raising NASA mission costs via prolonged launch delays need to be studied. In addition, agreements between NASA and DOD regarding launch opportunities need to be formalized so that the availability of these opportunities and the mutual commitment to support the programs that use them do not hinge on agency personnel remaining in key positions. Discussions with DOD representatives indicate that launch opportunities will continue to be available, but at present arrangements to use these opportunities are made only on an ad hoc or informal basis.

---

[5] See <http://ccmc.gsfc.nasa.gov>.

Current, future, and planned DOD and DOE platforms obtain or will obtain measurements for operational purposes that are directly relevant to magnetosphere–solar wind interaction science objectives. These measurements include data from DMSP and NPOESS particle and fields detectors and from auroral imagers, particle data from DOD and GPS satellites, and total electron content from ground-based GPS receivers. These data sets provide valuable augmentation of space-based observations and need to be exploited to the fullest extent possible. Support for preliminary processing and archiving of these data for retrospective scientific analyses is essential if the scientific community is to make meaningful use of these assets. NASA and NSF are urged to coordinate with DOD and DOE to facilitate preliminary processing and archiving activities so that these resources are leveraged to the fullest extent possible.

NSF: Central Role for Ground-Space Coordinated Observations

Global magnetospheric dynamics are reflected in ground processes. Our understanding of the specific correlation between ground observables and magnetospheric configuration and dynamics has dramatically increased in recent years, in part through the efforts of NSF's GEM campaign. We now understand the relationship between convection distribution and the underlying magnetopause reconnection geometry. We also appreciate how to identify time variations in reconnection in ground observations. We now know how to relate auroral spectra to the precipitating source population and critical boundaries in the magnetosphere, both in the magnetotail and on the dayside.

Ground-based observations provide distributed measurements of quantities that are difficult or impossible to measure with comparable distribution in space. These measurements include radar and ground magnetometer observations of convection; multispectral and high time and spatial resolution auroral imagery, meridian scanning photometers, and ULF pulsations. The ground observations therefore provide a means of monitoring the enormous system with remarkable efficiency. Historically, NSF assumed the central role in both establishing the ground-based observatories and coordinating these observations with measurements from space.

Maintaining and expanding these ground observation assets is critical for two reasons. First, ground observation assets will prove even more valuable as new space measurements are made and as models become able to assimilate these data. Recent advances in identifying ground signatures with specific magnetosphere–solar wind interaction phenomena make the observations even more valuable because they provide quantitative contextual and distributed information that is key to specifying the system and unavailable any other way. Second, some impediments remain to establishing an unambiguous link between phenomena that can be observed from the ground and magnetospheric–solar wind dynamics. These impediments include our incomplete understanding of ionospheric conductivities and the difficulty of specifying the net result of auroral processes that couple the high-altitude magnetosphere to the ionosphere. Only by comparing extensive ground observations with space-based observations will it be possible to further improve the power of ground-based observations.

COORDINATION BETWEEN PROGRAMS AND DIVISIONS WITHIN AGENCIES: NSF AND NASA

Because magnetospheric physics is one of a number of priorities in NSF and NASA space science programs, and because responsibility for it is split between NASA and NSF, it is important to recognize and eliminate unnecessary compartmentalization. The panel encourages cooperation and coordination between agencies and between programs within each agency. Several areas in which coordination is desirable are discussed next.

Comparative Magnetospheres and Planetary Exploration

Comparative magnetospheres remains a vital proving ground for theories of magnetospheric dynamics, because different systems present configurations and conditions not found at Earth. Our understanding of magnetosphere–solar wind physics will be seriously deficient unless these extraterrestrial systems are explored in ways that allow us to test our theories of their dynamical behavior. Solar system exploration therefore needs to provide avenues for observations of other magnetospheres.

In the past, major solar system missions could accommodate planetary geology and atmospheric and magnetospheric science payloads. This has not proven to be true under the Discovery program, whose missions are more highly focused. Instruments whose purpose is to further the understanding of comparative magnetospheres have less appeal in the Discovery mission environment than instruments that provide new information on a particular solar system body. The Solar Terrestrial

Probe missions and, to a greater degree, the Living With a Star missions tend to focus on the Sun-Earth Connection rather than on comparative magnetospheres. The panel encourages the Solar System Exploration and Sun-Earth Connection programs to coordinate their programs in the upper atmospheres and magnetospheres of the planets and to develop missions that address the outstanding problems in these areas.

Ties Across Organizational Boundaries

As the science of the solar wind, magnetospheres, and ionospheres matures and a new emphasis on serving the needs of space weather emerges, it is natural that new programmatic structures are being adopted within and between funding agencies. As these new structures are adopted, however, it must be recognized that the physical systems whose study is being overseen are closely linked and do not respect administrative boundaries. These structures may be reasonable and based on general distinctions between disciplines, but are nonetheless artificial constructions. Science disciplines must be allowed the freedom to explore the linkages between these physical systems. As new organizational structures are adopted, the ties between subdisciplines must not be lost, and research that spans administratively different areas must not be allowed to fall through the cracks. Often, cross-disciplinary research is not given priority by either discipline and therefore languishes. Planetary magnetospheres and solar wind interactions are an example of disciplines where such a lacuna occurs. A formal mechanism to fairly evaluate and support cross-disciplinary research should be adopted. Broadening the categories within NASA's SR&T program to allow magnetospheric and ionospheric research to be considered together is commendable in this regard.[6] Similar coordination between NSF's magnetospheric program in its Division of Atmospheric Sciences and its planetary magnetospheres and atmospheres research in its Division of Astronomical Sciences would also be most welcome. Steps appropriate to each case and agency need to be taken to ensure healthy cross-disciplinary research in other areas, including comparative magnetospheres, solar wind-magnetospheric physics, and the transitioning of research to application tools.

OPPORTUNITIES FOR SPACE MEASUREMENTS IN ENTITIES OTHER THAN NASA'S OFFICE OF SPACE SCIENCE

NASA's Office of Space Science provides the most regular opportunities to gain access to space through its major missions and principal-investigator-led missions, but there are other opportunities to have instrumentation carried into space. One of these is payloads attached to the International Space Station (ISS); another is launches by DOD or its foreign partners. In this section the panel discusses issues related to these two opportunities.

Space Station Attached Payloads

The knowledge gained in studies of the interaction of the solar wind with the magnetosphere and the ensuing understanding of the entry of solar energetic particles into the magnetosphere is particularly beneficial to the ISS and its occupants.[7] On the other hand, the ISS is not a natural or optimum platform for observing magnetosphere–solar wind interactions. It provides at best a limited opportunity for space physics research, owing to its orbit and facility configuration constraints. Furthermore, the additional qualification and safety issues pertaining to flight aboard a crewed vehicle add significantly to the cost of development, further diluting research resources. For these reasons, ISS is not a preferred platform for conducting magnetospheric physics research.

The panel emphasizes that continued progress in magnetosphere–solar wind interactions is of importance to ISS. The space environment plays a significant role in constraining ISS operations, as it does in constraining all space-based technology assets. Thus, continued basic research on the science of Earth's space environment is a high priority for ISS even when space physics instruments cannot be attached to the ISS per se.

Missions of Opportunity

The ability to optimize the return on launch opportunities by funding individual researchers to build instruments for opportunities on non-NASA missions is an excellent concept. Nevertheless, as presently executed, it is not achieving its full potential.

---

[6] See NRC, 2000, "Interim Assessment of Research and Data Analysis in NASA's Office of Space Science," letter report, Sept. 22.

[7] See NRC, 2000, *Radiation and the International Space Station: Recommendations to Reduce Risk,* National Academy Press, Washington, D.C.

To achieve the greatest science return, adjustments need to be made in the mission of opportunity (MOO) program to increase the frequency of these opportunities. The panel strongly endorses a change that would separate MOO from the SMEX and MIDEX announcements of opportunity, thereby allowing more frequent consideration and implementation of MOO proposals. To accomplish this change, the cost cap for MOO, including attached payloads on the ISS, should be approximately halved, from $35 million to approximately $15 million. This is one mechanism whereby launches of opportunity with DOD could be more effectively leveraged for science. (The panel recommends semiannual considerations to provide a better match with the frequency of such opportunities and with the development times of the missions.)

SCIENCE IN THE STRUCTURE OF PROJECT MANAGEMENT

The principal investigator model for missions has proven highly successful in terms of science return on the investment. One important reason for this success is that science issues are given the same weight as spacecraft and mission design issues. Strategic missions such as Solar Terrestrial Probes and Living with a Star missions could benefit from emulating some of the management structure of these missions. A position of science manager, equal in importance to the project manager, should be established for future strategic missions. To ensure the highest quality leadership, this position should be selected competitively. The panel believes that a science consortium lead by a competitively selected PI would be another way to infuse science into the management process.

INTERNATIONAL COOPERATION

Historically, research in space science, especially in solar wind–magnetosphere interactions, has had a strong international element. This international element arises first from the need for globally situated, ground-based measurements and then from the immensity of the task, which requires a cooperative effort to obtain the critical mass for its successful outcome. Recently, barriers have arisen to meaningful international cooperation.

ITAR and Export Controls

The International Traffic in Arms Regulations govern the export of both information and equipment that might be used by foreign entities against the United States. All space-associated investigations are now included under these regulations, which—as implemented by the State Department—have placed substantial burdens on the nation's space science community. These burdens are manifested in two ways. Space physics missions have always been conducted in close collaboration with our international colleagues in Europe and Asia, primarily Japan, and in Canada. The ITAR restrictions have made it extremely difficult to continue working with these colleagues on U.S. missions like STEREO, in which international contributions to the science payload are major elements of the design. Even rudimentary essential information concerning mission design concepts and spacecraft design plans has been subject to control, making it extremely difficult, if not impossible, to involve our foreign colleagues in making fully informed scientific judgements. It is simply impossible to properly design and build a scientific instrument without free access to relevant data on the spacecraft and mission design. The problem is even more acute in cases where instrument subsystems are provided by our foreign partners. One cannot team effectively if the instrument designs to which the team members contribute are sequestered.

The latter point suggests the second debilitating effect that these new restrictions are having on the nation's scientific community. The tremendous return to the United States from participation in foreign missions is illustrated by the SOHO (ESA), GEOTAIL (ISAS), and CLUSTER (ESA) missions, which were implemented by foreign agencies in Europe and Japan with significant NASA instrumentation, operations, and science participation. Now, however, the burdensome impact on foreign collaborating agencies has jeopardized opportunities to participate in foreign missions in the future. It is even harder to build an instrument jointly with our foreign collaborators. Clearly, the U.S. science community would not be on an equal footing with its international colleagues had it not been able to join them in these missions.

The ITAR situation is serious. Research scientists have been subjected to criminal charges and penalties. Consequently, some universities have refused to allow their researchers to accept grants and contracts with restrictive ITAR clauses. The inability to share information among partners in a mission could lead to mistakes and mission failures.

An amended ITAR rule was published on March 29, 2002, which applies only to university-based space research. The rule attempts to clarify the regulations and to remove obstacles to the conduct of university-based fundamental research in space. However, there remain a number of serious practical problems with the new rule, including continued restrictions on which students and staff at a university can have access to information and who in partner nations can gain access. The universities still may find the regulations too restrictive and ban their staff from entering into such programs. Moreover, the revised statutes do not address the equally serious problem—namely, that U.S. universities cannot work with the U.S. space industry without being subject to ITAR regulations. Here the restrictions are even greater than the restrictions on foreign collaborations.

Information Security

Present federal policies require all personnel having access to NASA spacecraft and science payload command systems to have background security checks. This is enforced by ensuring that contracts with universities are consistent with NPG 2810. This regulation requires that any individual having access to a spacecraft or its subsystems (such as science payloads) above a certain value, including the computers used to command science payloads, must be so screened. The universities are not generally convinced that they can require this of employees, especially those already hired. NASA's ruling means that university computer systems managers, project managers, and certain technicians and programmers must submit to background checks as part of their institution's contractual agreement with NASA on flight projects. However, university mission participants typically have no access to spacecraft system commands or controls. Firewalls are generally placed between the external workstations from which commands are sent to the science payload and the mission operations center that sends them. The investigation teams historically assume responsibility for the correctness of the commands sent to their instruments on board the spacecraft, and this has not presented a security problem in the past. For a few low-cost missions, some academic institutions have assumed full responsibility for operations and commanding. These missions present an information technology security conundrum, for they have been extremely successful.

MODELING, THEORY, AND DATA ASSIMILATION

Modeling and theory need to be integrated into ongoing research. Because the terrestrial magnetosphere's reconfiguration time scale is tens of minutes, far shorter than satellite orbit periods of hours to tens of hours, data sampling in Earth's magnetosphere will always be sparse and incomplete. For this reason, theoretical models and global simulations play a crucial role by forming a framework of understanding and context for the observations. The modeling and observations need to be wedded closely via data assimilation in global models. This will ensure that the models are properly constrained by the observations and can provide a suitable basis for extrapolating the observations to characterize the state and dynamics of the whole system.

Because of the central role that theory and global models play, support for them needs to be robust and sustained. Global simulation codes require teams of researchers, each with specialized expertise in the underlying physics, in numerical techniques, in visualization, and in user interfaces. To attract and maintain qualified researchers for efforts of this scope, the efforts cannot be supported by small (<\$100,000) 3-year research grants but must be supported by larger grants (>\$300,000 per year) for longer durations (5 years). It is also critical that more than one code be developed and used, because different techniques can sometimes lead to different behavior in the simulations, and comparisons between different codes are essential to identify consistent behavior potentially reflecting the real behavior of the system.

Theory, simulation, and modeling will also become increasingly instrumental in the planning and implementation of future missions. Understanding the character of the measurements required and the degree of improved understanding afforded by them and assessing the numbers and locations of observations to most efficiently achieve definitive results will require detailed analysis with models. Mission definition and design will therefore need to draw on the modeling resources of the community. Reliance on models will continue throughout each phase of future missions, including data analysis and assimilation. The dependence of future mission success on modeling underscores the need for sustained and substantial support for this effort.

The increasingly integral role played by models in data analysis implies that community access to models is another aspect of the theory, modeling, and simulation work that needs to be supported. As discussed in the section "NOAA: Transitioning New Operational Observing Platforms and Models," modeling is an area that is very appropriate for coordination with NOAA, which needs operational models. Furthermore, even in the arena of pure scientific inquiry, such coordination and community availability are important. Under NSF the GEM program has made initial strides in this direction, but making state-of-the-art models available to the community remains a challenging task that requires resources as well. Attempting to achieve this objective by requiring modeling teams to make their models available is manifestly the wrong approach, since these teams are already hard-pressed to develop robust models. Rather, the effort to convert research models into community models is a separate task, which requires support for developing interfaces and computational architecture. In many ways the community model is an intermediate step in the conversion of models to operational use. Support for this task should therefore derive not only from basic science, which is its primary purpose, but also from those agencies with an interest in developing more robust, physics-based models for prediction and forecasting.

Support for theory and modeling is therefore a natural area for interagency coordination. The benefits of modeling extend across all areas of interest, from basic science to prediction and forecasting, to mission development and planning. All of the relevant agencies—NSF, NASA, NOAA, and DOD—have a vested interest in maintaining a strong theory and modeling effort, and they should find ways to coordinate their activities to ensure that support is provided in a coherent way that addresses the concerns described above.

TECHNOLOGY DEVELOPMENT

The technological challenges for future solar-terrestrial missions are substantial and will require an effort distinct from SR&T, including an SEC program similar to the Planetary Instrument Definition and Development Program of the planetary community. The primary challenge for future magnetospheric missions will be meeting the need for constellation-class observations. For a new generation of spacecraft, the task is to design and develop a spacecraft architecture that can realize dramatic economies of scale even in limited production runs (tens of units). How low the ultimate cost per unit can go is not known, but the cost of the Iridium satellites, which were quite large, was ultimately reduced almost to $5 million. There are no fundamental technological reasons why a smaller platform could not be designed to cost less than this, but the task faces significant systems engineering, management, integration, and testing problems. It is not insurmountable, however, and is the type of ambitious but achievable goal that should be a focus for NASA or DOD. The New Millennium Program has been successful in developing spacecraft technologies and it would seem most appropriate for it to focus some of its technological investments on enabling constellation missions.

A comparable development effort is required to gain the ability to deliver tens of calibrated scientific instruments. Similar challenges of system engineering, manpower management, integration, and testing activities confront instrument builders contemplating the delivery of large quantities of instruments. Again, although the task is not an easy one, it does not appear to be impossible, and an instrument incubator program would provide a mechanism to fund the long-lead-time development of instrument technologies for this purpose.

For both the spacecraft and instrumentation development efforts, proper consideration must be given to an inherent feature of constellation-class missions—namely, that the large number of spacecraft and measurement points mitigates risk concerns and relieves the demands on instrument performance. The risk to the mission posed by the failure of a single spacecraft unit is extremely low, because the science return from, say, 45 satellites is nearly the same as that from 50. Because the redundancy is built into the constellation concept itself, one can accept single-string concepts in the spacecraft design.

In a similar way, the science return is enhanced primarily by the large number of distributed measurements rather than by the high precision of the measurements, so that the requirements for instrument performance relative to that demanded for single-satellite missions should be critically examined. Experience with non-science-grade instrumentation strongly suggests that individual instruments performing at a much lower level can yield dramatic scientific advances when deployed in constellations.

Finally, innovative and commercial solutions to spacecraft communications should be encouraged to reduce mission operations costs. Requiring the use of already overloaded systems such as the Deep Space Network for satellite tracking and communications for constellation missions is patently unworkable because of the enormous operating costs that such an approach necessarily entails. Innovative, automated communications approaches exist for Earth-orbiting satellites; such approaches were used very successfully for missions such as Freja and FAST and are being applied for other programs. These low-cost approaches to satellite communications and tracking need to be expanded aggressively to support constellation missions.

DATA ANALYSIS, DISSEMINATION, AND ARCHIVING

The analysis, dissemination, and archiving of data acquired from NASA and non-NASA missions as well as from ground observatories and networks are of paramount importance to successfully achieving the science advances described above. Given that the interrelated data sets to be acquired will be complex and more difficult to analyze than any acquired previously, the resources devoted to their analysis will need to be more substantial than those for earlier missions. The manpower that needs to be brought to bear will be correspondingly greater, and the best way of mobilizing this expertise will be to ensure that the data are available community-wide.

Data dissemination is therefore a key element of future research that advances in information technology have made much easier than in the past. The experience with missions such as ACE, SOHO, and IMAGE demonstrate that electronic dissemination of data works extremely well and facilitates community involvement in their analysis. There is no reason this success cannot carry over into the next decade with equal or greater success.

Given that the missions envisioned in the coming decade will not be superseded or repeated in the foreseeable future, the preservation of their data for subsequent analysis is critically important. The standardization system developed for the ISTP data exemplifies the level of commonality that will be needed for these new data sets. The standardization should be extended to ground data sets as well, so that their community use can be equally widespread. Standardization is also crucial for preservation of the data sets. While it is expected that the distributed data systems associated with different investigators and investigations will be maintained for some period of time after the prime mission or observation campaign, a centralized repository for the data will also be required and needs to be supported. It is almost certain that the number of basic issues that these data can be used to resolve will not be exhausted in the normal mission or observation lifetime of the spacecraft or the facilities used to obtain the data.[8]

EXTENDED MISSIONS

It is widely recognized that extended missions can provide a high science return for modest additional investment, and they are strongly encouraged. The panel endorses the practice of giving priority to those candidates for extension that most clearly support new research missions and strengthen or expand the science achieved. However, the costs of mission operations and data acquisition could be reduced considerably if tracking and communications for extended missions could be transferred to commercial or academic institutions at the discretion of the mission PI or project management. The use of this option is consistent with the philosophy of extended missions since their prime mission objectives would already have been achieved, fulfilling their intended charter. If the cost of extending missions could be significantly reduced and the pressure on mission operations and data analysis resources relieved to allow more simultaneous operations, a broader array of productive observatories could be maintained for magnetosphere–solar wind interaction science.

ADDITIONAL READING

A strategy for the conduct of space physics research has been set down in a number of reports by the NRC's Space Studies Board and its predecessor, the Space Science Board. These reports include the following:

Space Science Board, National Research Council. 1985. *An Implementation Plan for Priorities in Solar-System Space Physics*. National Academy Press, Washington, D.C.

Space Science Board, National Research Council. 1983. *The Role of Theory in Space Science*. National Academy Press, Washington, D.C.

Space Science Board, National Research Council. 1980. *Solar-System Space Physics in the 1980's: A Research Strategy*. National Academy of Sciences, Washington, D.C.

Space Studies Board, National Research Council. 1995. *A Science Strategy for Space Physics*. National Academy Press, Washington, D.C.

Space Studies Board and Board on Atmospheric Sciences and Climate, National Research Council. 1991. *Assessment of Programs in Solar and Space Physics—1991*. National Academy Press, Washington, D.C.

The research in this field is summarized in both textbooks and conference proceedings, including the following:

---

[8] See NRC, 2002, *Assessment of the Usefulness and Availability of NASA's Earth and Space Science Mission Data,* National Academy Press, Washington, D.C., pp. 41–44.

M.G. Kivelson and C.T. Russell (eds.). 1995. *Introduction to Space Physics*. Cambridge University Press, New York.

A. Nishida, D.N. Baker, and S.W.H. Cowley (eds.). 1998. *New Perspectives on the Earth's Magnetotail*. Monograph 105. American Geophysical Union, Washington, D.C.

B. Hultqvist, M. Oieroset, G. Paschmann, and R. Treumann (eds.). 1999. *Magnetospheric Plasma Sources and Losses*. Kluwer Academic Publishers, Dordrecht.

S.I. Ohtani, R. Fujii, M. Hesse, and R.L. Lysak. 2000. *Magnetospheric Current Systems*. Monograph 118. American Geophysical Union, Washington, D.C.

P. Song, H.J. Singer, and G.L. Siscoe (eds.). 2001. *Space Weather*. Monograph 125. American Geophysical Union, Washington, D.C.

Note added in proof: New Horizons, the first Pluto probe, has been selected as the first mission in NASA's New Frontiers program and is now in development. The probe, which will arrive at Pluto in 2015, carries solar wind plasma and energetic particle detectors in addition to its suite of remote sensing instruments and a dust experiment. In addition to its reconnaissance of the Pluto-Charon system, the probe is expected to encounter one or more Kuiper Belt objects.

3

Report of the Panel on Atmosphere-Ionosphere-Magnetosphere Interactions

SUMMARY

Earth is the single most interesting object in the universe to its inhabitants, the only place where we can be certain that a suitable environment for life exists. Furthermore, its complex systems are close enough to study in the sort of detail we will never obtain elsewhere. Earth and its sister planets are embedded in the outer atmosphere of the Sun. This outer atmosphere is continually being explosively reconfigured. During these explosive events, Earth is engulfed in intense high-frequency radiation, vast clouds of energetic particles, and fast plasma flows with entrained solar magnetic fields. Even though only a small fraction (generally <10 percent) of this energy penetrates into geospace, the effects are dramatic.

Space science programs to date have given us a detailed understanding of the average behavior of the component parts of geospace, in effect providing us with climatologies upon which to base educated guesses about the dynamic behavior of the global system. To go beyond this and understand the coupling processes and feedback that define the instantaneous response of the global system is much more difficult. The atmosphere-ionosphere-magnetosphere (A-I-M) system occupies an immense volume of space. At the same time, processes on scales from micro to macro impact the global system response.

GOALS AND OBJECTIVES

The overarching goals are as follows:

1. To understand how Earth's atmosphere couples to its ionosphere and its magnetosphere and to the atmosphere of the Sun and
2. To attain a predictive capability for those processes in the A-I-M system that affect human ability to live on the surface of Earth as well as in space.

Researchers currently have a tantalizing glimpse of the physical processes controlling the behavior of some of the individual elements in geospace. Some of the crosscutting science issues are these:

- The instantaneous global system response of the A-I-M system to the dynamic forcing of the solar atmosphere,
- The role of micro- and mesoscale processes in controlling the global-scale A-I-M system,
- The degree to which the dynamic coupling between the geophysical regions controls and impacts the active state of the A-I-M system,
- The physical processes that may be responsible for the solar forcing of climate change,
- The origin of the multi-MeV electrons in the outer magnetosphere and the cause of the pronounced fluctuations in their intensity, and
- The balance between internal and external forcing in the generation of plasma turbulence at low latitudes.

These critical science issues thread the artificial boundaries between the disciplines. The maturity of the A-I-M disciplines leads to a close connection between A-I-M science and applications for the benefit of society. The application of space physics and aeronomy to societal needs is now referred to as space weather. The space weather phenomena that most directly affect life and society include radiation exposure extending from space down to commercial airline altitudes, communications and navigation errors and outages, changes in the upper atmosphere that affect satellite drag and orbital decay, radiation effects on satellite electronics and solar panels, and power outages on the ground due to geomagnetically induced currents (GICs), to name a few.

STRATEGY AND REQUIREMENTS

The next decade may revolutionize our understanding of the dynamical behavior of the A-I-M system in response to driving from both the solar wind and the lower atmosphere. A carefully orchestrated collaboration between agencies with interest in space weather and space science research is required, since no one agency has the resources to provide the global view. Furthermore, new ground-based and space-based observing programs are required that make use of innovative technologies to achieve a simultaneous global view, highly resolved in space and time. Clusters of satellites flying in close formation can resolve dynamical response and separate spatial from temporal variations. New data storage and handling technologies are necessary to manage the shear volume of data generated, the multisatellite correlations, the mapping between in situ observations and images, searches across distributed databases, and other essential functions that will be necessary in the next decade to achieve an understanding of the entire system.

The systems view requires enhanced efforts to develop global theoretical models of the Sun-Earth system, including the simultaneous development of new software technologies for efficient use of parallel computing environments and adaptive grid technologies to address the large range in spatial and temporal scales characteristic of the global system structure and response. However, the A-I-M system is not simply multiscale, but it also requires inclusion of additional physical processes of ionized and neutral gases made up of individual particles. Data assimilation technologies are crucial for integrating new observations into research and operational models of the space environment. The problems associated with the transition of research models and data sets to operations must be specifically addressed in the planning and implementation of research programs aimed at improving space weather forecasting and specification.

The National Science Foundation's (NSF's) highly successful Solar, Heliospheric, and Interplanetary Environment (SHINE) program, its Coupling, Energetics, and Dynamics of Atmosphere Regions (CEDAR) program, and its Geosphere Environment Modeling (GEM) program, and the recent coordination of these groups into Sun-to-Earth analysis campaigns, highlight the need to focus this broad range of expertise on issues involved in coupling between the Sun, solar wind, magnetosphere, and ionosphere/atmosphere regions. To this end, NSF recently funded the Science and Technology Center for Integrated Space Weather Modeling. NSF's information technology initiatives should be utilized as much as possible to develop important collaboration technologies in support of such major community analysis efforts.

The investigation of planetary A-I-M systems reveals details of value to understanding the terrestrial system. Future planetary missions should regularly be outfitted to carry out at least a baseline set of observations of the upper atmosphere, the ionosphere, and the magnetosphere. In addition, theoretical studies linking our understanding of the terrestrial environment with other planetary environments are an effective way of bringing extensive knowledge of plasma and atmospheric processes in the terrestrial environment to bear on the interpretation of planetary phenomena.

While the National Oceanic and Atmospheric Administration (NOAA) and the Department of Defense (DOD) have pursued space environment forecasting for many years, their connection to the science community was facilitated by the inception of the National Space Weather Program (NSWP) in 1995 and NASA's new Living With a Star (LWS) program. The NSWP is a multiagency endeavor to understand the physical processes, from the Sun to Earth, that result in space weather and to transition scientific advances into operational applications. NASA's new LWS program represents an important opportunity to provide measurements and develop models that will clarify the relationship between sources of space weather and their impact.

Enhancements and innovations in infrastructure, data management and assimilation, instrumentation, computational models, software technologies, and methods for transitioning research to operations are essential to support the future exploration of geospace.

RECOMMENDATIONS

In the next decade, NASA should give highest priority to multispacecraft missions such as Magnetospheric Multiscale (MMS), Geospace Electrodynamics Constellation (GEC), Magnetospheric Constellation (MagCon), and Living With a Star's geospace missions, which take advantage of adjustable orbit capability and the advancing technology of small spacecraft. Missions that involve large numbers of simply instrumented spacecraft are needed to develop a global view of the system and should be encouraged. NSF, for its part, should support extensive ground-based arrays of instrumentation to give a global, time-dependent view of this system. Ground- and space-based programs should be coordinated—as, for example, is being done in the Thermosphere-Ionosphere-Mesosphere Energetics and Dynamics (TIMED)/CEDAR program—to take advantage of the complementary nature of the two distinct viewpoints. NASA, NSF, DOD, and other agencies should encourage the development of theories and models that support the goal of understanding the A-I-M system from a dynamic point of view. Furthermore, these agencies should work toward the development of data analysis techniques, using modern information technology, that assimilate multipoint data into a three-dimensional, dynamic picture of this complex system. Funding for the NASA Supporting Research and Technology (SR&T) program should be doubled to raise the proposal success rate from 20 percent to the level found in other agencies. Solar Terrestrial Probe (STP) flight programs should have their own targeted postlaunch theory, modeling, and data analysis support.

Major NSF Initiative

Simultaneous, multicomponent, ground-based observations of the A-I-M system are needed in order to specify the many interconnecting dynamic and thermodynamic variables. As our understanding of the complexity of the A-I-M system grows, so does the requirement to capture observations of its multiple facets. The proposed Advanced Modular Incoherent Scatter Radar (AMISR) will provide the opportunity for coordinated radar-optical studies of the aurora and coordinated investigations of the lower thermosphere and mesosphere, a region not well accessed by spacecraft. Initial location at Poker Flat, Alaska, will allow coordination of radar with in situ rocket measurements of auroral processes. Subsequent transfer to the deep polar cap will enable studies of polar cap convection and mapping of processes deeper in the geomagnetic tail.

1. The National Science Foundation should extend its major observatory component by proceeding as quickly as possible with Advanced Modular Incoherent Scatter Radar (AMISR) and by developing one or more lidar-centered major facilities. Further, the NSF should begin an aggressive program to field hundreds of small automated instrument clusters to allow mapping the state of the global system.

Ground-based sensors have played a pivotal role in our understanding of A-I-M science and must continue to do so in the coming decade and beyond. Anchored by a state-of-the-art phased-array scientific radar, the $60 million AMISR is a crucial element for A-I-M. A distributed array of instrument clusters would provide the high temporal and spatial resolution observations needed to drive the assimilative models, which the panel hopes will parallel the weather forecasting models we now have for the lower atmosphere. Much of the necessary infrastructure for such a project has already been demonstrated in the prototype Suominet, a nationwide network of simple Global Positioning System (GPS)/meteorology stations linked by the Internet. The proposed program would add miniaturized instruments, such as all-sky imagers, Fabry-Perot interferometers, very-low-frequency (VLF) receivers, passive radars, magnetometers, and ionosondes in addition to powerful GPS-based systems in a flexible and expandable network coupled to fast real-time processing, display, and data distribution capabilities. Instrument clusters would be sited at universities and high schools, providing a rich hands-on environment for students and training with instruments and analysis for the next generation of space scientists. Data and reduced products from the distributed network would be distributed freely and openly over the Internet. An overall cost of $100 million over the 10-year planning period is indicated. Estimated costs range from $50,000 to $150,000 per station depending on instruments to be deployed. Adequate funding would be included for the development and implementation of data transfer, analysis, and distribution tools and facilities. Such a system would push the state of the art in information technology as well as instrument development and miniaturization.

Extending the present radar-centered upper atmospheric observatories to include one or more lidar-centered facilities is crucial if we are to understand the boundary between the lower and upper atmosphere. Fortunately, a number of military and nonmilitary large-aperture telescopes may become available for transition to lidar-based science in the next few years. Highest priority would be given to a facility at the same geographic latitude as one of the existing radar sites.

NASA Orbital Programs

The Explorer Program has since the beginning of the space age provided opportunities for studying the geospace environment just as the Discovery Program now provides opportunities in planetary science. The continued opportunities for University-Class Explorer (UNEX), Small Explorer (SMEX), and Medium-Class Explorer (MIDEX) missions, practically defined in terms of their funding caps of $14 million, $90 million, and $180 million, respectively, allow the community the greatest creativity in developing new concepts and a faster response time to new developments in both science and technology. These missions also provide a crucial training ground for graduate students, managers, and engineers. Imager for Magnetopause-to-Aurora Global Exploration (IMAGE), launched in March 2000, is an example of a highly successful MIDEX mission; it was preceded by the first two ongoing SMEX missions, Solar Anomalous and Magnetospheric Particle Explorer (SAMPEX) and Fast Auroral Snapshot Explorer (FAST), launched in 1992 and 1996, which have provided enormous scientific return for the investment. The Aeronomy of Ice in the Mesosphere (AIM) SMEX was recently selected for launch in 2006. The UNEX program, after the great success of the Student Nitric Oxide Explorer (SNOE), launched in February 1998, has effectively been cancelled. This least expensive component of the Explorer program plays a role similar to that of the sounding rocket program, with higher risk accompanying lower cost and a great increase in the number of flight opportunities. An increase in funding to $20 million per mission with one launch per year would make this program viable with modest resources.

2. The SMEX and MIDEX programs should be vigorously maintained and the UNEX program should quickly be revitalized.

The STP line of missions defined in the NASA Sun-Earth Connection (SEC) Roadmap (strategic planning for 2000 to 2025) has the potential to form the backbone of A-I-M research in the next decade. The missions that are part of the current program include TIMED, launched in February 2002, Solar-B, the Solar Terrestrial Relations Observatory (STEREO), MMS, GEC, and MagCon. After TIMED, launched in February 2002, the next A-I-M/STP mission, MMS, is in the process of instrument selection for a 2009 launch. The STP cadence, with one A-I-M mission per decade (TIMED was significantly delayed), has fallen behind the NASA SEC Roadmap projections.

3. The panel heartily endorses the STP line of missions and strongly encourages an increase in the launch cadence, with GEC and MagCon proceeding in parallel.

The A-I-M research community has very successfully utilized the infrastructure developed within the International Solar-Terrestrial Physics (ISTP) program. The integration of the data from spacecraft and ground-based programs beyond those funded by the ISTP itself—such as those of NOAA, LANL, and the DOD—have contributed substantially to our understanding of the global system.

Comparisons between the Sun-Earth system and other Sun-planet or stellar-planet systems provide important insights into the underlying physical and chemical processes that govern A-I-M interactions. Improved understanding of A-I-M coupling phenomena such as planetary and terrestrial auroras would benefit from such an approach.

4. The Sun-Earth Connection program partnership with the NASA Solar System Exploration program should be revitalized. A dedicated planetary aeronomy mission should be pursued vigorously, and the Discovery Program should remain open to A-I-M-related missions.

NASA Suborbital Program

The NASA Suborbital program has produced outstanding science throughout its lifetime. Many phenomena have been discovered using rockets, rockoons, and balloons, and many outstanding problems brought to closure, particularly when space-based facilities are teamed with ground-based facilities. These phenomena include the auroral acceleration mechanism, plasma bubbles at the magnetic equator, the charged nature of polar mesospheric clouds, and monoenergetic auroral beams. This program continues to generate cutting-edge science with new instruments and data rates that are more than an order of magnitude greater than typical satellite data rates. Both unique altitude ranges and very specific geophysical conditions are accessible only to sounding rockets and balloons, particularly in the campaign mode. Many current satellite experimenters were trained in the Suborbital program, and high-risk instrument development can occur only in such an environment. To accomplish significant training, it is necessary that a graduate student remain in a project from start to finish and that some risk be acceptable; both are very difficult in satellite projects. The high scientific return, coupled with training of future generations of space-based experimenters, makes this program highly cost-effective.

The sounding rocket budget has been level-funded for over a decade, and many principal investigators (PIs) are discouraged about the poor proposal success rate as well as the low number of launch opportunities. The sounding rocket program was commercialized in 2000; in this changeover, approximately 50 civil service positions were lost and the cost of running the program increased. Approved campaigns were delayed by up to a year, and it is not yet clear whether the launch rate will ever return to precommercialized levels. Effectively, commercialization has meant a significant decline in funding for the sounding rocket program. An additional concern is that, as currently structured—i.e., with a fixed, 3-year cycle for all phases of a sounding rocket project—funding is not easily extended to allow graduate students to complete their thesis work, because it is generally thought that such work should fall under the SR&T program, already oversubscribed. The rocket program has a rich history of scientific and educational benefit and provides low-cost access to space for university and other researchers. Further erosion of this program will result in fewer and fewer young scientists with experience in building flight hardware and will ultimately adversely affect the much more expensive satellite programs.

5. The Suborbital program should be revitalized and its funding should be reinstated to an inflation-adjusted value matching the funding in the early 1980s. To further ensure the vibrancy of the Suborbital program, an independent scientific and technical panel should be

formed to study how it might be changed to better serve the community and the country.

Societal Impact Program

The practical impact on society of variations in the A-I-M system falls into two broad categories: the well-established effects of space weather variations on technology and the less clear yet tantalizing influence of solar variability on climate. The societal impacts of space weather are broad—communications, navigation, human radiation hazards, power distribution, and satellite operations are all affected. Space weather is of international concern, and other nations are pursuing parallel activities, which could be leveraged through collaboration. The role of solar variability in climate change remains an enigma, but it is now at least being recognized as important to our understanding of the natural—as opposed to anthropogenic—sources of climate variability.

6. The study of solar variability—both of its short-term effects on the space radiation environment, communications, navigation, and power distribution and of its effect on climate and the upper atmosphere—should be intensified by both modeling and observation efforts.

NASA's Living With a Star program should be implemented, with increased resources for the geospace component. Missions such as the National Polar-orbiting Operational Environmental Satellite Systems (NPOESS) and the Solar Radiation and Climate Experiment (SORCE) are needed to provide vital data to the science community for monitoring long-term solar irradiance. NPOESS should be developed to provide ionosphere and upper atmosphere observations to fill gaps in measurements needed to understand the A-I-M system. An L1 monitor should be a permanent facility that provides the solar wind measurements crucial to determining the response of the A-I-M system to its external driver, and the NSWP should be strengthened and used as a template for interagency cooperation. International participation in such large-scope programs as LWS and NSWP is essential.

7. The NOAA, DOE/LANL, and DOD operational spacecraft programs should be sustained, and DOD launch opportunities should be utilized for specialized missions such as geostationary airglow imagers, auroral oval imagers, and neutral/ionized medium sensors.

NASA's new Living With a Star program can, over the next decade, provide substantial new resources to address these goals. It is crucial that there be overlap between the geospace and solar mission components of LWS for the system to be studied synergistically, that resources be adequate for the geospace component, and that theory, modeling, and a comprehensive data system, which will replace the ISTP infrastructure, be defined at the outset, as called for in the Science Architecture Team (SAT) report findings. NSWP, a multiagency endeavor established in 1995, addresses the potentially great societal impact of physical processes from the Sun to Earth that affect the near-Earth environment in ways as diverse as terrestrial weather. The program specifically addresses the need to transition scientific research into operations and to assist users affected by the space environment. Such multiagency cooperation is essential for progress in predicting the response of the near-Earth space environment to short-term solar variability.

The interagency cooperation established in the NSWP is outstanding and is a model for extracting the maximum benefit from scientific and technical programs. It has also been effective at bringing together different scientific disciplines and the scientific and operations communities. Interagency cooperation has worked well in the AFOSR/NSF Maui Mesosphere and Lower Thermosphere Program, and it has been key to the success of the NOAA GOES and POES programs of meteorological satellites with space environment monitoring capabilities. International multiagency cooperation has been very successful for the ISTP program, which involves U.S., European, Japanese, and Russian space agencies. Global studies require such international cooperation. The panel recognizes that much more science can be extracted by careful coordination of ground- and space-based programs.

Maximizing Scientific Return

Funding for NASA's Supporting Research and Technology program, including guest investigator studies and focused theory, modeling, and data assimilation efforts, is essential for maximizing the scientific return from large investments in spacecraft hardware.

Supporting Research and Technology

While spacecraft hardware projects are concentrated at relatively few institutions, the NASA SR&T program is the primary vehicle by which independent investigations can be undertaken by the broader community. Likewise, NSF helps individual investigators to carry out targeted research through its Division of Atmospheric Sciences (ATM) base programs—SHINE, CEDAR, and GEM. Such individual PI-driven initiatives are the most inclusive, with data analysis as well as theoretical efforts and laboratory studies, and often lead to the highest science return per dollar spent. The funding for such program elements falls far short of the scientific opportunities, with the current success rate for submitted NASA SR&T proposals being 10 to 20 percent. Furthermore, limited available SR&T funds have been used for guest investigator participation in underfunded STP-class flight programs. Without adequate MO&DA funding for NASA orbital and suborbital programs, the SR&T budget intended for targeted research on focused scientific questions has been utilized to support broader data analysis objectives.

8. The funding for the SR&T program should be increased, and STP-class flight programs should have their own targeted postlaunch data analysis support.

9. A new small grants program should be established within NSF that is dedicated to comparative atmospheres, ionospheres, and magnetospheres (C-A-I-M).

A new C-A-I-M grants program at NSF would allow the techniques (modeling, ground-, and space-based observations, and in situ measurements) that have so successfully been applied to A-I-M processes at Earth to be used to understand A-I-M processes at other planets. Such a comparative approach would improve our understanding of these processes throughout the solar system, including at Earth. Currently, a modest $2 million planetary science program at NSF covers all of solar system science (except for solar and terrestrial studies), with only a small fraction going to planetary A-I-M research.

Theory, Modeling, and Data Assimilation

Theory and modeling provide the framework for interpreting, understanding, and visualizing diverse measurements at disparate locations in the A-I-M system. There is now a pressing need to develop and utilize data assimilation techniques not only for operational use in specifying and forecasting the space environment but also to provide the tools to tackle key science questions. The modest level of support from the NSF base programs (CEDAR, GEM, SHINE) and NASA SR&T has been inadequate to build comprehensive, systems-level models. Rather, individual pieces have been built, and first stages of model integration achieved with funding from such programs as NASA's ISTP program and its Sun-Earth Connections Theory Program (SECTP), the AFOSR MURI program, NSF Science and Technology Center programs, and the multiagency support to such efforts as the Community Coordinated Modeling Center. Such programs enable the development of theory and modeling infrastructure, including models to address the dynamic coupling between neighboring geophysical regions. Their value to the research community is clearly their provision of longer-term funding, which has been essential to developing a comprehensive program outside the purview of SR&T.

10. The development and utilization of data assimilation techniques should be enhanced to optimize model and data resources. The panel endorses support for theory and model development at the level of the NASA Sun-Earth Connections Theory Program, the AFOSR/ONR MURI program, NSF Science and Technology Center programs, and the multiagency support to such efforts as the Community Coordinated Modeling Center (CCMC). Support should be enhanced for large-scope, integrative modeling that applies to the coupling of neighboring geophysical regions and physical processes, which are explicit in one model and implicit on the larger scale.

The preceding science recommendations can be grouped into three cost categories and prioritized (see Table 3.1). Equal weight is given to STP and LWS lines, as indicated by funding level. Small programs are ranked by resource allocation, while the Advanced Modular Incoherent Scatter Radar is the highest priority moderate initiative at lower cost than others.

3.1 INTRODUCTION

Earth, unique in the universe as the only object known to support life, follows an orbit in the outer atmosphere of the Sun—an outer atmosphere that is continually being explosively reconfigured. During these events, Earth is engulfed in intense high-frequency radiation, vast clouds of energetic particles, and fast plasma flows with entrained solar magnetic fields. Even though only a small fraction (generally <10 percent) of this energy penetrates into geospace, its effects are dramatic.

TABLE 3.1 Panel's Recommended Priorities for New Initiatives

Initiatives in Geospace		Recommended 10-Year Funding (million $)
Major		
Solar Terrestrial Probes (2)		800
Living With a Star		500
Discovery (1)		350
	Subtotal	1,650
Moderate		
Advanced Modular Incoherent Scatter Radar and Lidar Facilities		92
Explorer Program (assume 3 missions in the 10 years will be devoted to AIM)		300
L1 Monitor (excluding tracking)		50
Small Instrument Distributed Ground-Based Network		100
	Subtotal	542
Small		
Suborbital program		300
NSF Supporting Research and Technology		200
National Space Weather Program		50
NSF SHINE, CEDAR, GEM, C-A-I-M (new)		135
Theory		138
Living With a Star (geospace)		60
Sun-Earth Connection Theory Program (geospace)		18
DOD MURI (ionosphere)		20
NSF STC (geospace)		20
HPCC (geospace)		20
NOAA, DOE/LANL, and DOD science for the A-I-M community		50
	Subtotal	848
Total		3,065

To date, space science programs have provided a detailed understanding of the average behavior of the component parts of geospace, in effect providing climatologies upon which to base educated guesses about the dynamic behavior of the global system. To go beyond this and understand the coupling processes and feedback that define the instantaneous response of the global system is much more difficult. The A-I-M system occupies an immense volume of space. At the same time, processes on scale sizes from micro to macro impact the global system response.

The ISTP program is the most ambitious program to date to explore the A-I-M system. ISTP samples the huge volume of the A-I-M system by simultaneous measurements from a handful of satellites. Despite the sparse coverage, analysis of data from ISTP satellites has allowed scientists to begin to glimpse the rich variety of coupling and feedback processes that define the global response of the geospace environment to solar wind disturbances. The first experiments with innovative imaging technologies that view large regions of geospace in snapshots (e.g., from the IMAGE spacecraft) have already provided insights into the instantaneous response, unattainable by past missions. The first attempts to achieve the high spatial and temporal resolution needed to survey the microscale controls of the global system (e.g., from the FAST spacecraft) have revealed new details about acceleration processes and electrodynamic coupling. With these new missions, we are replacing our steady-state view of geospace regions with a dynamical view. But we are far from understanding the complex coupling processes and interplay between components that dictate the integrated global system response.

It is clear that the A-I-M system actively responds to the solar wind and that components of this system may be preconditioned or may interact in ways that redistribute solar wind energy throughout the system, actively limiting the entry of solar wind energy into geospace during extreme events. A few examples are given in the next pages to illustrate the complexity of this interaction and the challenges that lie ahead.

Life on this planet is protected from the high-energy radiation and dangerous particle clouds in interplanetary space because Earth has its own magnetic field and is surrounded by an absorbing atmosphere. Earth's magnetic field presents a northward-directed magnetic field barrier to the oncoming solar wind in the ecliptic plane (Figure 3.1). This barrier can be breached, however, if southward-directed solar magnetic fields impact it and merge or reconnect with Earth's magnetic field. Fortunately, strongly southward-directed magnetic fields are not a persistent feature of the quiet interplanetary medium. They are mainly confined to structures generated in explosive solar events and in high-speed plasma streams.

The passage of southward interplanetary magnetic field (IMF) structures by Earth pumps energy into the near-space environment. The tightly coupled nature of the A-I-M system is clearly revealed by its response to interplanetary magnetic clouds (IMCs), which have strong and long-lived southward magnetic fields and drive the most intense magnetic storms. Intense convection is produced, which brings particles from the magnetotail storage region (called the plasma sheet) deep into the inner magnetosphere on open drift paths, ener-

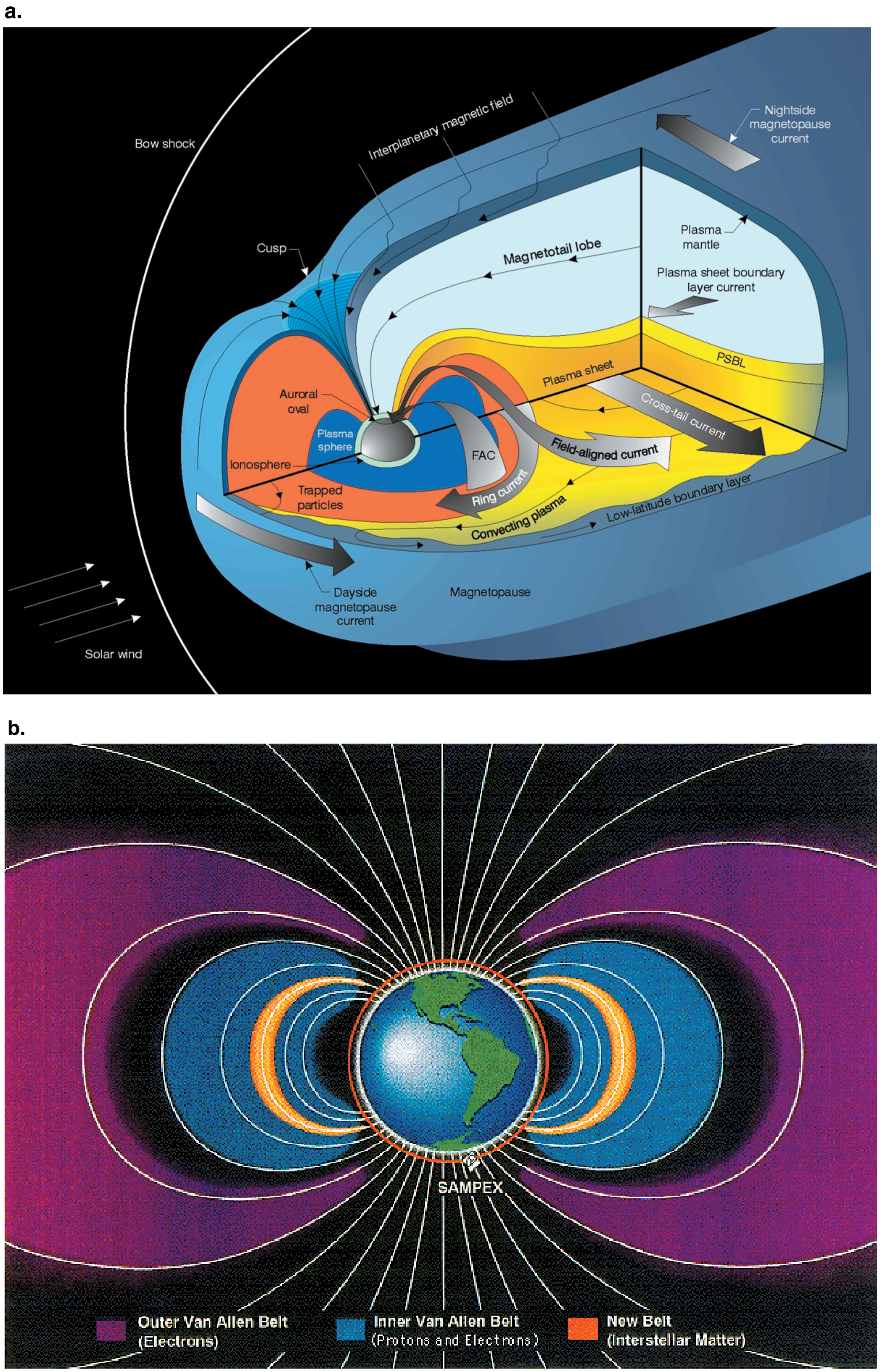


FIGURE 3.1 (a) A schematic of the magnetosphere showing major particle populations and current systems; (b) close-up of radiation belts (trapped particles), including inner and outer zone populations and trapped anomalous cosmic rays (interstellar matter).

gizing them to form the storm-time ring current, shown schematically in Figure 3.1. Under extreme conditions, the strong current produced by these particles cannot close upon itself in the equatorial plane (to form the ring of current that its name implies) but is forced to close through the subauroral and midlatitude ionosphere. This closure produces a strong electric field in the ionosphere called a polarization jet. The electric fields in the polarization jet map upward along the field lines to the magnetosphere, producing a penetration electric field in the inner magnetosphere, further changing the drift paths of the ring-current ions. The plasmasphere (corotating plasma of ionospheric origin) responds strongly to the penetration electric fields and enhanced convection. Long plumes of plasmaspheric material snake out to the dayside magnetopause, draining thermal plasma into the dayside boundary layers. This plasma may later become a source for the plasma sheet (Figure 3.1). Along drift paths mapped into the ionosphere, storm-enhanced ionization moves toward the dayside polar cap, where geomagnetic field lines connect to the interplanetary magnetic field. Steep plasma density gradients at the edges of ionospheric density patches form in the changing convection pattern, playing havoc with technologies like the Global Positioning System, and ionospheric irregularities form, disrupting communication systems.

There are indications that the ionosphere may, in turn, have a major impact on the dynamics of the magnetosphere. Solar wind dynamic pressure variations trigger ionospheric outflows from the vicinity of the polar cap. These ionospheric "mass ejections" begin well before the storm maximum, preconditioning the tail plasmas with heavy ionospheric ions energized by the solar wind interaction. Enhanced convection during magnetic storms stresses the magnetotail, producing dramatic reconfigurations of the basic structure of the magnetotail, called substorms. Auroral currents associated with substorms also produce an outflow of ionospheric ions directly into the magnetotail. Substorms may actually sever the outer plasma sheet from the magnetotail, producing a major loss of plasma and energy. Dipolarization of the magnetic field during substorms, which reduces the stretching of magnetotail field lines, generates intense electric fields, which accelerate the storm-enhanced plasma sheet. Since this accelerated plasma drifts earthward under conditions of strong convection to form the ring current, there is a clear connection between magnetotail dynamics and magnetic storm effects in the near-Earth region of the magnetosphere. Variations in plasma sheet density play an important role in modulating magnetic storm intensity and substorm processes and represent another means by which the A-I-M system internally modulates the geo-effectiveness of solar wind disturbances. The interplay between removal of plasma sheet material and refilling of the plasma sheet from the solar wind and the ionosphere to supply the ring current during storms is not understood. Even the basic mechanisms for refilling the plasma sheet (Figure 3.1) have not yet been determined.

Earth's outer magnetosphere is often populated to a surprising degree by relativistic electrons, which pose a radiation hazard to space-based systems. The origin of the multi-MeV electrons in the outer zone is not known. They are generally correlated with increased substorm activity driven by high-speed solar wind and favorable coupling to southward interplanetary magnetic field, both of which have semiannual and solar-cycle variations. Enhancements occur with relatively regular 27-day periodicity during the declining phase of the 11-year sunspot cycle and are well associated with high-speed solar wind stream structures. Flux variations at the solar activity maximum are dominated by coronal mass ejections (CMEs) (see the report of the Panel on Solar-Heliospheric Physics), which launch magnetic clouds toward Earth, producing geomagnetic storms. The mechanism(s) by which magnetospheric particles are accelerated to relativistic energies are unknown at present, although a number of interactions with plasma wave modes are promising candidates. The impact of solar wind shock structures on Earth's magnetosphere has been shown to generate induction-electric-field pulses that rapidly accelerate electrons and protons, generating an entire new radiation belt in a matter of minutes. An interesting coupling between the ring current and radiation belts results wherein magnetic fields generated by the ring current cause scattering and loss of radiation belt particles. In addition, waves near the ion gyrofrequency generated by ring-current ions are believed to contribute to electron precipitation losses in the dusk sector. Dramatic losses from the electron radiation belts (Figure 3.1b) also result from interaction of energetic electrons with lightning-generated waves, called whistlers. Lightning-induced electron precipitation events exemplify direct coupling of tropospheric weather systems with the radiation belts and the ionospheric regions overlying thunderstorms.

Much of the energy and momentum entering Earth's environment eventually finds its way into the upper atmosphere. The in situ absorption of solar EUV radiation not only protects life on Earth, but drives large day-night temperature and tidal wind variations in the upper atmosphere, which vary dramatically with the solar cycle.

During geomagnetic storms the additional energy and momentum input from the solar wind is initially focused at the poles but is so intense that ultimate effects are worldwide. The upper atmospheric regions underlying the polar cap and the adjacent auroral zone are not only heated but are also subject to momentum transfer from the solar wind. The ionosphere is the intermediary in this momentum transfer and acts like an ion engine driving the wind in huge twin vortices in both hemispheres. Uninhibited by plasma boundaries, this wind permeates the globe, generating its own electric fields by dynamo action. This additional source of electrification can modify and even reverse the polarity of the ambient low-latitude electric field. It thus can turn on and off the ionospheric processes responsible for much of the space weather that adversely affects communication and navigational systems. Impulsive high-latitude energy inputs driven by substorm reconfigurations of the goemagnetic tail generate tsunami-like atmospheric waves, which propagate at near-supersonic speeds across the upper boundary of the atmosphere, disturbing the ionosphere along their path.

The neutral atmosphere is an active player in the coupled system. It is heated at high latitudes by auroral currents and particle precipitation associated with the substorm current systems, expanding to higher altitudes and undergoing changes in composition. The temperature and composition changes move in great sloshing waves from the conjugate auroral regions, meet at the equator, and pass through, modifying atmospheric conditions as they travel. The changing neutral atmosphere modifies the ionospheric plasma, creating traveling disturbances that have yet to be successfully modeled. These variations alter the conductivity of the ionospheric plasma and thus actively modulate the current flow between the ionosphere and magnetosphere.

Even without the energy and momentum input from the solar wind the ionosphere exhibits a rich variety of weather. In fact, the highest disturbance level of transionospheric or subionospheric communication channels and signal propagation occurs at low latitudes. Here the nighttime ionosphere is supported against gravity by electrical currents. But this equilibrium is unstable, and the slightest disturbance is enough to set off huge convective upwellings, much like thunderstorms, which rise to heights more than 1,000 km and last for hours. This phenomenon has been intensely studied by ground-based radars and more recently by rockets and satellites. Some of the earliest successes in computer simulations occurred when researchers applied codes developed for nuclear weapons effects to this explosive natural phenomenon. At midlatitudes we have been greatly surprised by the pictures of the ionosphere that are now available and that reveal much more structure than previously thought. A highly remarkable coupling between atmospheric waves and ionospheric electric fields seems to be occurring, which we simply do not understand.

Ground-based observations indicate that the upper/middle ionosphere is driven from below by both mechanical and electrodynamic inputs from the massive lower atmosphere. Gravity waves launched from thunderstorms and orographic features on Earth's surface grow as they propagate upward and break in the mesosphere, driving the large-scale flow and mixing of constituents. Lightning activity in tropospheric thunderstorms imposes huge transient electric fields on the middle atmosphere, leading to intense heating and ionization, manifested by transient luminous displays known as sprites and elves. These same electric fields can drive avalanche acceleration of relativistic electron beams, which produce gamma-ray emissions during their upward traverse and may escape into the radiation belts. Tropospheric thunderstorm activity may influence the electrical and chemical properties of the middle and upper atmosphere by means of these newly discovered phenomena and by upward conduction currents.

The coldest temperatures on Earth, achieved at the summer mesopause, are due to dynamical processes. There is much to learn about the dynamics and coupling between the upper/middle and lower atmospheric regions and even more about the possible impact on weather and climate. Noctilucent clouds (Figure 3.2), observed at 82 km altitude at 50 to 60 degrees latitude when temperatures drop below 140 K, are occurring more frequently, and sightings are moving to lower latitudes. These are part of a growing body of information that indicates the upper atmosphere has cooled over the past 20–50 years—a cooling that is thought to be associated with a warming trend at lower altitudes, possibly due to anthropogenic influences. Variations in the solar constant with solar activity seem too small to serve as the basis for such changes. Answers may lie in the solar cycle modulation of cloud nucleation through cosmic ray intensity, the effects of energetic particle precipitation on ozone chemistry, and solar cycle changes in the global electric circuit. A significantly improved understanding of the links between the upper atmosphere and weather and climate is needed. NASA's TIMED spacecraft mission, launched in February 2002, is providing an exploratory look at the response of the mesosphere and lower thermosphere to solar and magnetospheric inputs from above and atmospheric inputs from below.

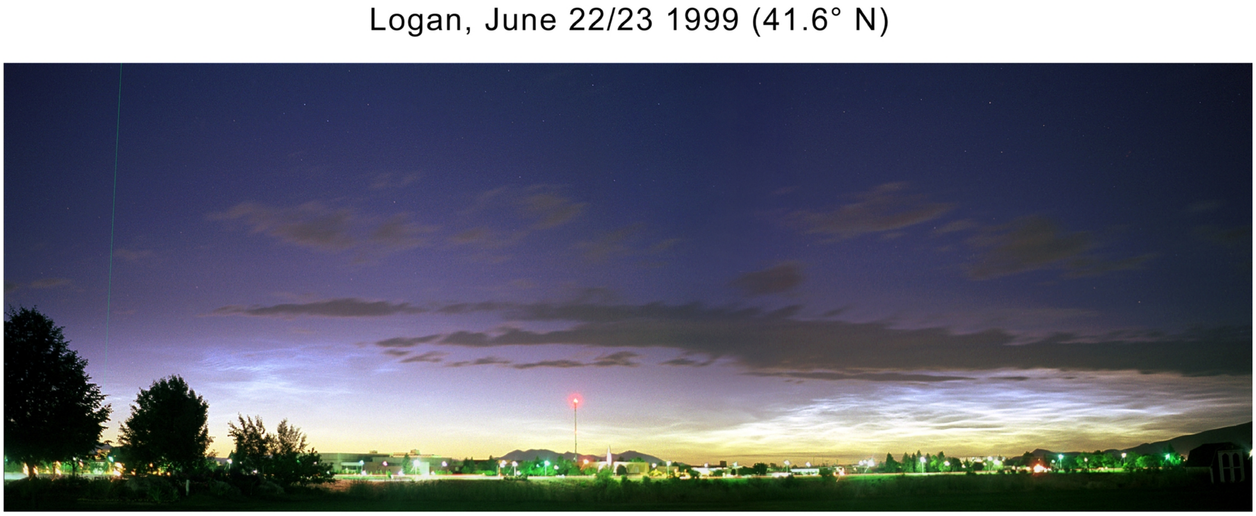


FIGURE 3.2 Image of noctilucent clouds (NLC) observed at midlatitudes from Logan, Utah (41.6 N, 111.8 W), recorded at 10:30 p.m. on June 22 and 23, 1999. The NLC consisted of two bright patches containing diffuse and billow-type wave structures that appeared bluish white in the twilit sky (the dark silhouettes are due to tropospheric clouds). Faint NLC are also evident toward the image center. NLC have been observed at high latitudes (>55 degrees) for more than 100 years, with recent evidence for migration to midlatitudes suggesting climate change. Courtesy of M.J. Taylor, Utah State University.

The AIM Explorer, scheduled for launch in 2006, will investigate the causes of high-altitude noctilucent clouds and serve as a baseline for further study of long-term changes in the upper atmosphere.

Finally, the solar system is a natural laboratory for comparing planetary A-I-M systems. Comparison of the Sun-Earth system to other Sun-planet systems provides insights into the broad atmospheric, ionospheric, and magnetospheric responses under a variety of conditions. The physical and chemical processes that control the atmosphere and plasma environments of the planets and their satellites are essentially the same throughout the solar system, but they are manifested in very different ways due to differing heliocentric distances, planetary sizes, atmospheric composition, and intrinsic magnetic fields. Studies of the A-I-M systems on other planets are significant in a number of ways. First, the interpretation of plasma processes on other planets is often aided by knowledge of similar processes on Earth. Second, processes that have counterparts on Earth are often modified in interesting ways in the extreme environments of other planets (such as possible substorms on Mercury in the absence of an ionospheric source of plasma and substorms on Jupiter to replace magnetic flux drifting out under the influence of outflowing torus plasma). Third, by studying a wide variety of solar wind–magnetosphere environments, a range of scenarios is built up for use in interpreting conditions on extrasolar planets. Within this category falls the study of plasma processes with no direct analogue at Earth. Lastly, the study of basic plasma physics processes within these planetary environments supplies important information on processes operating within astrophysical plasmas throughout the universe.

A-I-M science themes for the coming decade are organized not by isolated regions in geospace but by crosscutting themes that describe global system behavior. These include the Earth as a particle accelerator; the global electrical environment; a new view of the neutral atmosphere; micro- and mesoscale control of the behavior of the global system; the response of the global system to forcing, exemplified by magnetic storms and substorms; the possible impact of the A-I-M system on global climate; and comparative aspects of planetary A-I-M systems.

The maturity of the A-I-M discipline leads to a close connection between A-I-M science and applications for the benefit of society. The application of space physics and aeronomy to societal needs is now referred to as space weather. Space weather phenomena that most directly affect life and society include radiation exposure extending from space down to commercial airline altitudes; communications and navigation errors and outages; changes in the upper atmosphere that affect satellite drag and orbital decay; radiation effects on satellite electronics and solar panels; and power outages on the ground due to geomagnetically induced currents (GICs), to name a few. High-frequency satellite-based navigation systems, of which the Global Positioning System is an example, rely on propagating signals through the ionosphere from ground to satellite. Any outage or degradation due to space weather is a threat to life as well as to military advantage. Ionospheric irregularities, therefore, are just as important for navigation applications as they are for satellite communication. Precise positioning can also be compromised by the total electron content (TEC) of the ionosphere. The highly dynamic nature of the system introduces an uncertainty for single-frequency GPS users (for instance, commercial airlines). HF is the communication channel most susceptible by far to space weather, with absorption refraction, multipath, and scintillation all adding to the problems. Many military systems depend on HF and UHF communications systems, as do commercial airlines. Systems that use HF waves that refract from the bottom of the ionosphere are vulnerable to space weather effects.

During large geomagnetic storms, rapidly time-varying electric fields and currents, often co-located with the aurora, intensify and expand to lower latitudes, sometimes producing power outages that affect transmission lines through GICs. With improved forecasting, power companies can take protective measures to reduce or eliminate effects on their systems. Astronauts are subject to considerable radiation even in low Earth orbit. When the International Space Station orbit was changed to accommodate Russian launches, it entered a more dangerous portion of the radiation belt. Even polar airplane flights have high radiation levels. Likewise, instruments and entire satellites are placed at risk because of spacecraft charging by energetic electrons, loss of computerized control systems, and severe magnetic torques. In addition, satellites in low Earth orbit are significantly affected by atmospheric drag. During a magnetic storm the atmosphere is heated and expands. After one such event, NORAD lost track of many space objects and had to reacquire them. Our knowledge of the timing of such events and their prediction could greatly assist in keeping track of all the space debris in orbit.

3.2 SCIENCE THEMES AND OPPORTUNITIES FOR THE COMING DECADE

EARTH AS A PARTICLE ACCELERATOR

High-energy particles, photons, and radio waves permeate the cosmos and provide glimpses of particle acceleration processes throughout the universe. Remarkably, Earth is also capable of generating powerful beams of this sort. A classic example arose when researchers using Compton Gamma Ray Explorer data found that thunderstorms near the surface of Earth were the source of the intensely energetic photons. The advantage of near-space regions of Earth as a cosmic particle accelerator and the source of photons is that we are free to delve deeply into the origins and effects of these processes. Detailed study is possible and we are not forced simply to blend theory with the small amounts of information we can obtain from distant sources. In the next sections the panel discusses a few of the more spectacular acceleration mechanisms that have been found in the near-space regions of Earth and that we hope to understand by the end of this decade.

Auroral Physics

Earth's aurora was one of the earliest motivations for the study of space plasmas, and it remains one of the most fundamental areas of space research (Figure 3.3). With in situ measurements provided by rockets and satellites, great progress has been made in understanding these stunningly beautiful light displays. Nonetheless, the physics of auroral processes is among the most complex in all of space physics, spanning the entire range from macroscale to microscale.

The broad outlines of auroral physics are generally understood. Auroral electrons are accelerated out of Earth's plasma sheet to energies of between 1 and 10 keV, and when they strike the atmosphere, they excite atmospheric atoms, which then emit the characteristic greenish blue and reddish purple colors of the aurora. The source of energy for this process appears to come from the magnetotail and is closely related to disturbed geomagnetic conditions and substorms. The details of exactly how the particle acceleration occurs remains a fundamental question of auroral physics. It is generally agreed that quasi-static electric fields parallel to Earth's

FIGURE 3.3 The aurora viewed from Alaska. Courtesy of Jan Curtis.

magnetic field are the dominant form of acceleration. Just how these fields are set up and maintained is not known, but it is closely linked to the flow of current upward along the magnetic field.

Auroral electrons can also be accelerated by wave processes such as the Alfvén wave, a transverse perturbation of the background magnetic field by an induced electric field, which causes plasma to oscillate as if tied or frozen in to the magnetic field. These waves are thought to be launched from the equatorial tail region and to propagate toward the ionosphere. Alfvén waves are observed in both the auroral zone and at altitudes well above the auroral acceleration region. Theoretical work also suggests a link between the Alfvén waves and the quasi-static fields, in which the waves, by reflecting off the ionosphere, can evolve. Alfvén waves are also thought to occur as part of global, field line resonances, in which a standing wave is set up along a field line from one ionosphere to the other. In all of these Alfvén wave processes, the perpendicular extent of the wave is a key parameter, but one that has not been measured.

Where an aurora occurs, a variety of associated physical phenomena occur along the magnetic field line that threads the ionospheric end where the light emission occurs. Just as electrons are accelerated downward, ions are also accelerated upward to the same energies. Ions are also observed to be heated, predominantly in the perpendicular direction. Owing to the magnetic mirror force, which reflects charged particles away from a stronger magnetic field region, these heated ions acquire parallel as well as perpendicular energy as they move upward. These field-aligned particle flows are unstable to a variety of plasma waves that are emitted from the auroral regions.

In the past decade, two satellite missions, Freja and FAST, along with several sounding rockets, have elucidated many details of this picture of the auroral acceleration region. These spacecraft exploited high data rates and state-of-the-art instrumentation. Perhaps the most striking result of these missions is the understanding that not only is the downward acceleration of electrons important, but also the upward acceleration of electrons plays a regular role in auroral processes. FAST has made clear just how ubiquitous this upward acceleration of electrons is and that field-aligned current is fundamental to the process. A variety of waves are also associated with this downward current region. This has brought a symmetry to the auroral acceleration process, in which electrons are accelerated both upward and downward, with the current that they carry closing through the ionosphere.

Measurements from FAST have also shown that within the region where electrons are accelerated, the dense background ionospheric plasma is frequently evacuated. The mechanism by which these cavities can be created is not yet known. Other discoveries include the association of ion heating with a variety of low-frequency waves, the observation of solitary wave structures in both upward and downward current regions, and the correlation of high-frequency waves with electron bunching.

Despite this list of accomplishments, significant questions remain: How is the quasi-static potential drop that accelerates the electrons distributed along the magnetic field? How is this potential drop set up and maintained? What are the perpendicular scales of the Alfvén waves that are observed? How are auroral density cavities generated? What role do solitary structures play in auroral physics? Most of these are not new questions, but rather are questions that are not tractable with single spacecraft and existing theories.

Several key developments are needed: (1) the recommended ground-based instrumentation, including the AMISR, (2) a multispacecraft mission for auroral science, and (3) the creation of sophisticated, self-consistent plasma models of the auroral region. On the measurement side, a dedicated, three- or four-spacecraft mission that allows determination of the evolution of temporal phenomena and spatial distribution (both along and across field lines) of auroral features is needed to make progress on issues such as the structure of electric fields in both quasi-static and wave cases. On the theoretical side, it is time to develop models that can handle an entire auroral field line; the effects of precipitating electrons, protons, and heavier ions; and the physics of the cold, dense ionospheric plasma, using nonideal magnetohydrodynamic and two-fluid simulations evolving to include two- and three-dimensional particle effects.

Ring Current

The ring current, consisting of ions ranging in energy from tens to hundreds of keV, embodies sufficient plasma pressure to reduce the magnetic field locally and measurably at equatorial locations on Earth's surface, especially during geomagnetic storm periods of enhanced convection, ultimately driven by the solar wind. The contribution of radiation belt particles (1 keV to many MeV) to plasma pressure is negligible, which allows them to be treated as test particles in dynamic models, neglecting the fields they produce. During geomagnetic storms, the ring current can consist primarily of oxygen of ionospheric origin, posing the question, How can eV ions be so quickly accelerated to 100 keV energy on the time scale of enhanced magnetospheric convection (a few hours)? Recent measurements from the Polar and Cluster spacecraft suggest that strong solar wind pressure pulses, which frequently accompany the onset of a geomagnetic storm, play an important role in reconfiguring the dayside cusp region, providing ionospheric plasma access to the dayside region, where plasma is swept antisunward on reconnected magnetic field lines, increasing the oxygen content of the plasma sheet in the tail (see Figure 3.1). Enhanced convection then brings the oxygen-enriched plasma earthward to form the storm-time ring current. Global measurements of ring current development from the IMAGE satellite, combined with plasma sheet composition measurements of the source population from the planned MMS mission, as well as cusp and nightside auroral zone outflow measurements, will make it possible to trace the energization steps and determine, with solar wind input measurements, which storm triggers are most geoeffective at producing an oxygen-enriched ring current. Recovery rates of the storm-time ring current, which affects the buildup of relativistic electrons inside geosynchronous orbit, are species-dependent.

Radiation Belts

The radiation belts are divided into an inner zone, with peak flux around L = 2 (approximate equatorial distance in units of Earth radii), which comprises predominantly protons produced by cosmic ray interaction with the atmosphere, and a dynamic outer zone of electrons whose flux is determined by solar wind control of magnetospheric processes. Figure 3.4 shows the electron variability over most of a solar cycle (bottom panel), as seen by the low-altitude polar-orbiting SAMPEX satellite. The most intense fluxes are seen during the declining phase of the solar maximum period (1993–1995) and are associated with intervals of recurring high-speed solar wind stream interaction with Earth's magnetosphere. These high-speed streams map back to solar coronal hole access to the ecliptic plane. Approaching the solar maximum (2000–2001), the flux enhancements are more sporadic and can be identified with CME-driven geomagnetic storms. The solar minimum was quite evident in mid-1996. The black line trace in the bottom panel is a plot of the parameter that characterizes a buildup of the ring current, the average horizontal component of Earth's magnetic field at the equator. One sees a strong correlation between flux enhancements and this measure of geomagnetic storms, strongest when it is most negative, since the ring current opposes and reduces the magnetic field due to Earth's dipole. Radiation belt electron fluxes build up at lower L values during stronger geomagnetic storms, posing a greater threat to constellations of spacecraft such as the GPS and to the International Space Station during those times. Note that the worst is yet to come in the current solar cycle (23), in terms of both relativistic electron and inner zone

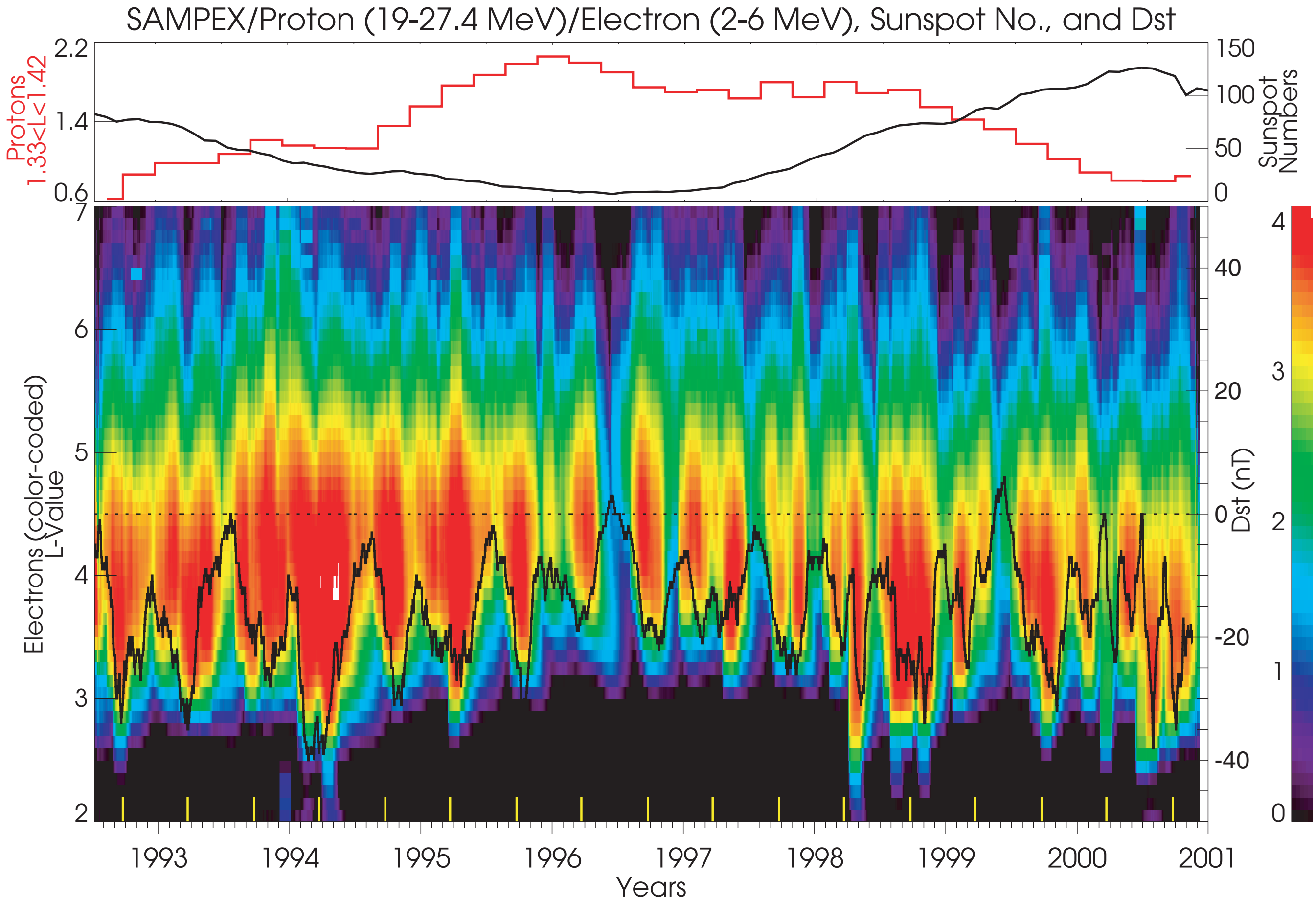


FIGURE 3.4 Thirty-day averaged MeV proton (top) and electron (bottom) fluxes from July 1992 to January 2001, as measured by the SAMPEX satellite in low-altitude polar orbit. Solar cycle variation is evident, with electron fluxes greater during the declining phase of the solar cycle, penetrating to lowest L-value (equatorial radial distance) during large storms characterized by negative Dst geomagnetic activity index. Inner zone proton fluxes maximize at solar minimum (1996), when cosmic rays have greater access to the inner heliosphere. SOURCE: X. Li, D.N. Baker, S.G. Kanekal, M. Looper, and M. Temerin, 2001, Long term measurements of radiation belts by SAMPEX and their variations, *Geophysical Research Letters* 28(20): 3827.

proton exposure. The latter anticorrelates with the solar cycle, since fluctuations in the solar wind are weaker at solar minimum, allowing cosmic rays greater access to the inner heliosphere and Earth's upper atmosphere. These protons pose the greatest threat to the electronics of spacecraft passing through the South Atlantic Anomaly (SAA) of Earth's magnetic field and are the reason that the Chandra satellite, designed for X-ray and gamma-ray astronomy measurements, has instruments that are not operated during SAA encounters. The planned operation of the satellite was modified on orbit once these effects were encountered. A better understanding of Earth's energetic particle environment will improve our ability to design future spacecraft missions for both research and applications.

Terrestrial Gamma-Ray Flashes

The observation of terrestrial gamma-ray flashes above thunderstorms constitutes one of the most unexpected discoveries of the Compton Gamma-Ray Observatory (CGRO) mission. The gamma-ray photon energy extends above 1 MeV, indicating bremsstrahlung radiation from >1 MeV electrons, consistent with C.T.R. Wilson's predictions of upward beams of electrons, now called "runaways," accelerated by thundercloud fields. This discovery provides possibly the first observational evidence for relativistic runaway breakdown, a fundamental new plasma acceleration process never before observed in nature or produced in the laboratory, which can proceed at much lower electric field levels than conventional runaway breakdown. Runaway acceleration may well be an important particle acceleration/radiation process for astrophysical plasmas. Computer models show that intense transient thundercloud electric fields impose a total potential drop between 20 and 80 km altitude of ~100 MeV, producing highly nonlinear runaway avalanche formation of relativistic electron beams over large spatial regions, with gamma-ray flashes emitted between 30 and 80 km altitude. The intense upward-driven relativistic electron beams eventually escape the collisional atmosphere, enter the radiation belts, and may become trapped as a result of scattering due to plasma wave interactions.

EARTH'S ELECTRIC FIELD

Earth's magnetic field has been studied for centuries and determines the geometry in which space plasma physics takes place. Its dynamics, however, are controlled by Earth's electric field. Techniques for measuring the electric field are only several decades old and we are just beginning to explore its complexity. Earth is embedded in the Sun's atmosphere, a gigantic magnetohydrodynamic generator that powers the entire magnetosphere from high above the planet. These interplanetary electric fields map down the magnetic field lines, particularly when the interplanetary magnetic field has a southward component, and even reach the stratosphere. From below, the atmosphere acts as a hydrodynamic generator, creating electric fields ranging from planetary scales and tidal periods down to turbulent eddies and polarized structures smaller than a kilometer. The associated electric potentials generated map upward for vast distances along the highly conducting magnetic field lines. Many of these electric field sources are poorly understood, yet they are a major source of such phenomena as particle energization and the turbulence in the ionosphere that can create communication and navigation problems on transionospheric propagation paths. A goal for this coming decade is to increase our understanding of the global electric field and its dynamic and irregular behavior to match our understanding of the magnetic field.

Magnetospheric Electric Fields

The complexity of the magnetospheric electric field has become apparent over the last decade. Throughout much of the magnetosphere, the basic picture we have of the electric field is one in which the field is considered to be electrostatic. Since the rarefied plasma in the magnetosphere can be considered to be collision free, the electric field parallel to the ambient geomagnetic field can be shorted out by free charge carriers to first approximation. This implies that the field lines are equipotentials and that the electric field will therefore map along field lines, scaling inversely with the area of a flux tube or cross section subtended by field lines. The significance of this mapping lies in the fact that the electric field is connected to the flow velocity of the plasma. The high conductivity in the collisionless plasma implies that the electric field vanishes in the reference frame that moves with the plasma. Thus, in any other reference frame (such as that of Earth), the electric field can be given by the cross product of the background magnetic field and the plasma velocity. In this way, the electric field is directly related to plasma convection in the magnetosphere. This situation is referred to as the frozen-in condition, since magnetic field lines can be treated as if they move with the plasma convection in this picture, a simplification that breaks down, for example, along auroral field lines. Since the solar wind flows past the magnetosphere, this gives an electric field directed in the dawn-to-dusk direction, which leads to a potential of the order of 100 kV across the magnetopause. Some fraction of this potential drop penetrates into the magnetosphere and maps along magnetic field lines, causing the convection of the magnetosphere and the potential drop across the polar cap. The magnitude of the potential drop is a good measure of the strength of the convection electric field and of the transfer of energy into the magnetosphere.

Across the polar cap, this potential maps to the region of antisunward flow, or equivalently, duskward-directed electric field. At lower latitudes, the electric field (and flow) is reversed, resulting in sunward flow. The region of flow shear is one in which there are field-aligned currents, active aurora, and enhanced ionospheric conductivity. All of these attest to the dynamic importance of electric fields in the magnetosphere. Indeed, the laminar character of the flow over the polar cap is rarely as smooth as this picture would suggest. Dependencies on interplanetary magnetic field orientation, solar wind dynamics, and magnetospheric activity usually produce a situation that is much more complex. Although the study of electric fields in the polar region has seen many successes in understanding the basic configurations of the polar cap potential structure, the understanding in terms of fundamental theoretical models remains sketchy at best. The key to further understanding is the increasingly global coverage of polar cap electric field measurements from radars and satellites. As these measurements give a more complete understanding of the polar cap configuration, the future holds promise for developing a deeper knowledge, including predictive capability based on external inputs.

It is now becoming clear that the large-scale earthward convection in the equatorial plasma sheet cannot be characterized simply by mapping the high-latitude ionospheric convection potential along field lines to the equatorial plane (as described above). Rapid magnetic field changes and large inductive electric fields in the midtail plasma sheet decouple motions in this region from those at the ionospheric ends of the flux tube. The long-held view of laminar earthward convection in the plasma sheet is being replaced by a view of disordered small-scale slow flows (which contribute little to earthward convection) and transient high-speed flow bursts (called bursty bulk flows, or BBFs) that transport an estimated 80 percent of the earthward-directed magnetic flux, energy, and mass. BBFs, which are azimuthally narrow but extended in length downtail, appear to be the normal mode of convective transport in the midtail under all conditions. These flow bursts have been seen as close to Earth as 10 Earth radii (R_E). Our detailed understanding of the magnetotail potential patterns and how they evolve in time is still quite primitive. This is largely because electric fields at the equator are much weaker than those at the polar cap (the same cross-field potential is spread over a much larger area) and they encompass a much larger volume of space; a low-altitude, polar-orbiting satellite can traverse the polar cap in a matter of minutes and measure electric fields of several to hundreds of millivolts per meter. In the equatorial plane, even a satellite relatively near Earth—for example, one in geosynchronous orbit—takes several hours to cover a significant fraction of the magnetosphere.

In the near-Earth plasma sheet, the large-scale convection pattern is again well represented by a cross-tail electric field proportional to the solar wind motional electric field, which increases in magnitude with increasing magnetic activity. The penetration of the cross-tail electric field into the inner magnetosphere is controlled by plasma processes. Currents flowing near the inner edge of the plasma sheet cause the dusk side to charge positive and the dawn side to charge negative. The resulting dusk-to-dawn electric field opposes the convection electric field and acts to shield the inner magnetosphere. "Undershielding" of the convection field can result if the cross-tail potential increases abruptly. This lasts until the plasma sheet particles drift sufficiently far earthward to adjust the shielding to an equilibrium level. The converse of this produces "overshielding." We are left with a picture of a simple cross-tail convection electric field smoothly decreasing in magnitude with decreasing radial distance due to shielding effects. This simple convection electric field pattern is modified by Earth's rotation. Dense ionospheric plasma near Earth is dragged along as the result of electron-neutral collisions and co-rotates with Earth. This creates an electric field that extends to a few Earth radii and drives corotation of the entire inner magnetosphere. This combination of sunward and co-rotating convection controls the drift of magnetospheric particles around Earth. Recent results from the Combined Release and Radiation Effects satellite indicate that the large-scale electric field during disturbed times looks very different from the simple models described here. The observed convection during disturbed periods is enhanced in keeping with these simple models but is spatially structured and often reaches maximum values deep in the inner magnetosphere at the location of the ring current.

The low-altitude counterpart of these enhancements produces polarization jets. These jets result when the field-aligned portion of the partial ring-current loop closes in the subauroral ionosphere through a steep conductivity gradient. Intense electric fields are produced that map up field lines to the inner magnetosphere. These strong and structured electric fields alter the drift paths of the ring-current ions that originally created them and produce structuring of the plasmasphere.

New instrumentation utilizing improved double probes and other techniques such as the electron-beam-drift technique are now producing improved measurements of low-amplitude equatorial fields. It is clear, however, that multiple spacecraft observations are essential to developing a better understanding of the dynamics of the equatorial electric field. Some of these observations will come from the Cluster mission as the satellite separations are increased. Magnetospheric Multiscale will provide multiple point measurements to fill in the picture. In coordination with these multisatellite observations, auroral imaging is unparalleled in its ability to (1) provide an unrestricted field of view, spatial resolution, sensitivity, and sunlit imaging capability needed to map transient and localized auroral features such as poleward boundary intensifications (the auroral signature of BBFs) and substorm initiation; (2) separate spatial from temporal variability; and (3) investigate dynamical behavior in the global context. These new measurements will lead to fundamental discoveries about the equatorial electric field—although we have a crude model of the basic electric field configuration, the structure of the dynamics and evolution will only be understood as we begin to get sufficiently global coverage.

Mesospheric and Ionospheric Effects of Lightning-Driven Electric Fields

At any given time, as many as 2,000 thunderstorms are active over the surface of the Earth, providing a global lightning rate of ~100 per second. Lightning discharges radiate intense electromagnetic pulses of >20 GW peak power and produce transient quasi-static electric fields of up to ~1 kV per meter at mesospheric altitudes, with total potential drop between 20 and 80 km altitude of ~100 MeV. These fields can heat, accelerate, and precipitate electrons in the lower ionosphere, mesosphere, and the radiation belts.

Dramatic recent experimental evidence of strong electrodynamic coupling between tropospheric lightning discharges (<15 km altitude) and the mesosphere/lower ionosphere (30 to 100 km) include spectacular luminous optical emissions known as red sprites and elves (Figure 3.5), as well as rapid ionization and conductivity changes. Visual accounts of glows in clear air above thunderstorms have appeared in the literature since the 19th century, the most vivid accounts being those by air transport pilots. The possibility of "upward" lightning or "lightning to the ionosphere" was seriously considered long before what are now known as sprites were first documented during the past decade. The discovery of these elusive lights high in the sky has captured the imagination of the scientific community and the public, with over a thousand articles having appeared in newspapers and popular magazines worldwide. These new findings raise fundamental questions about the nature of the electrodynamic coupling between thunderclouds and the upper atmosphere.

The new observations have been interpreted using several new interaction and coupling mechanisms, including the heating of the ambient electrons by lightning electromagnetic pulses or by large quasi-electrostatic thundercloud fields and by runaway electron processes. Runaway acceleration and ambient heating processes may influence reaction rates and lead to the production of different types of ions, potentially affecting mesospheric chemistry and dynamics. The mechanisms underlying these spectacular phenomena are just now being uncovered, and their effects on a global scale will need to be assessed during the coming decade. At the very least these luminous phenomena provide a window of measurability for the electrodynamics of an A-I-M region not accessible for in situ measurements.

The current global electrical circuit models, which consider the fair weather current to be driven by quasi-static thunderstorm electric fields, may need to be substantially revised to account for the sporadic and highly nonlinear component of charge transfer that can be seen in electrical breakdown channels extending from the lower atmosphere to the ionosphere.

Electrodynamic Coupling

One of the most important roles of Earth's electric field is its ability to couple the different regions of the magnetosphere electrodynamically. It is common to speak of the electric field "mapping" along auroral field lines. Although this mapping of the electric field along field lines is a convenient way of discussing magnetospheric convection, it is somewhat misleading in a time-dependent situation, since the change in plasma velocity must be associated with forces that accelerate or decelerate the plasma. It is equivalent to say that the

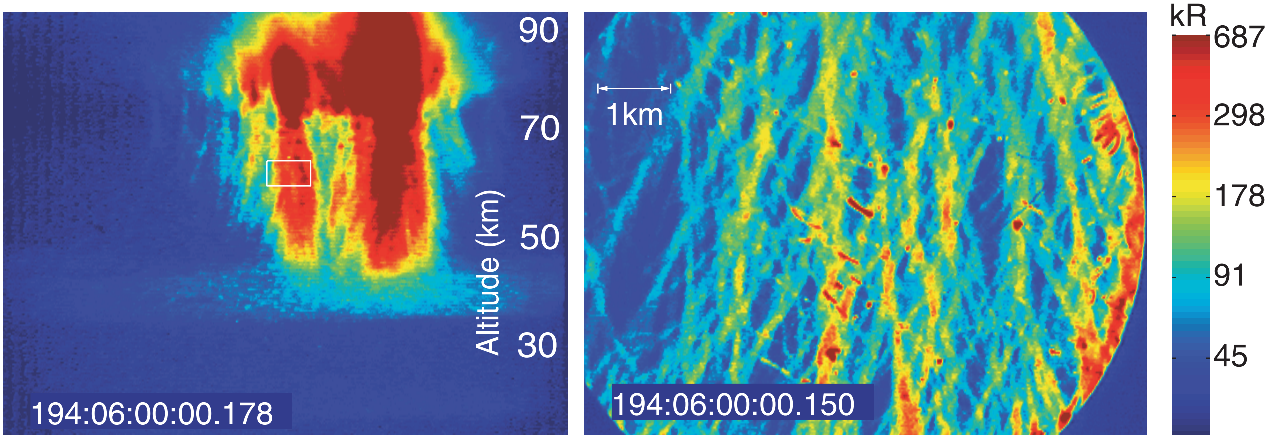


FIGURE 3.5 Sprite event imaged July 13, 1998, from Langmuir Laboratory in Socorro, New Mexico, during a storm over northern Mexico, 491 km away. Left-hand image was taken with a 50-mm lens-intensified CCD camera, wide field-of-view (FOV) system, while narrow FOV camera (right-hand image) zooms into the sprite with a 16-in. Newtonian refracting telescope. The FOV of the telescope in relation to the wide FOV is outlined with a white box on the left-hand panel. This sprite was a huge event associated with a large cloud-to-ground lightning flash of peak current 128.5 kA. It extends from 40 to 90 km altitude and is ~50 km wide. In the wide FOV the sprite appears almost homogeneous, and little structure is evident. The telescopic image reveals that the sprite consists of numerous densely packed glowing filaments. Courtesy of Elizabeth Gerken, STAR Laboratory, Stanford University.

motion of plasma causes a shear stress on magnetic field lines tied to the collisionless plasma by almost perfect conductivity, and that the resulting force is responsible for the acceleration of the plasma. In a time-dependent situation, this stress propagates along magnetic field lines as an Alfvén wave. Because this wave carries changes in the field-aligned current, this electrodynamic coupling of the magnetosphere induces a current system that permeates the magnetosphere. The study of these waves, in the ultralow-frequency (ULF) range between 1 mHz and about 5 Hz, is an important aspect of magnetospheric physics in its own right since these waves transmit momentum and energy throughout the magnetosphere.

When field-aligned currents encounter the inherently conductive ionosphere, they flow across geomagnetic field lines. Pedersen currents, which flow in the direction of the ionospheric electric field, provide an energy sink for the magnetospheric plasma. Since the neutral atmosphere has a much higher mass density than the plasma in the ionosphere, the transfer of momentum to the atmosphere provides a drag force on plasma convection in the magnetosphere, much like trying to pull a heavy rock along the ground with a rope. If the convection is strong and long lasting, such as during geomagnetic storms, neutral winds can be accelerated in the direction of convecting ions owing to collisions between the ions and neutrals.

Mapping of electric fields along field lines is violated when parallel electric fields develop in the auroral acceleration region. In addition to accelerating auroral particles, these parallel electric fields decouple the magnetospheric convection from the atmospheric drag, allowing convection to proceed with reduced dissipation. New data from satellites such as FAST and Polar have provided new details of these processes, providing strong constraints on theories of this region. These new observations, combined with new theoretical models and numerical simulations of increasing levels of sophistication, are leading to a period of intense research activity in the physics of auroral acceleration processes.

Ionospheric conductivity can be modified by precipitating electrons that produce increased ionization in the ionosphere. This enhances the conductivities of the ionosphere and modifies the current systems that set up the parallel electric fields in the first place. Such interactions can give rise to a positive feedback that leads to narrow, very intense current structures. Although auroral arcs have been measured to extend horizontally less than 100 m, most theoretical scenarios are unable to produce such narrow structures. This ionospheric feedback instability is a promising mechanism for creating such narrow arcs. Indeed, recent observations suggest that the formation of narrow arcs is more likely in cases where the background ionospheric conductivity is low, i.e., under non-sunlit conditions. This is consistent with a feedback instability which is stabilized by high conductivity.

Because of the electrodynamic coupling along magnetic field lines in the magnetosphere, the auroral ionosphere acts as a screen on which magnetospheric dynamics can be projected. Many observational techniques—such as radar measurements, measurements from ground magnetometer arrays, and measurements from sounding rockets and low-Earth orbiting satellites—take advantage of the fact that ionospheric convection can be seen as an image of magnetospheric dynamics. An understanding of the coupling between the ionosphere and magnetosphere as well as the breakdown of such coupling is crucial for the interpretation of such measurements. Moreover, these processes are essential for understanding auroral formation and the flow of plasma through the magnetosphere, and the study of this coupling should be considered a high priority in the next decade.

Earth's Atmosphere as a Generator

The upper atmosphere is in constant motion. Just as the ocean surges in response to gravitational forces, solar heating forces the air to flow across the top of the atmosphere from day to night. This is particularly important in the thermosphere, where the x rays and EUV from the Sun are absorbed on the dayside, raising the temperature to well over 1000 K. In this height range these in situ tides reach velocities upwards of 200 m/s (720 km/hr), the highest wind speeds in the atmosphere. In fact these winds would flow even faster if the ionospheric plasma did not act as an electromagnetic break. The wind tries to force an ionized particle across the magnetic field, but the particle is tied by gyration about the magnetic field and so hardly moves and must be struck again and again by the neutral gas to be transported. This ion drag effect can also be described by the electromagnetic force density **J** x **B** since the wind induces a current when it tries to carry the plasma across the field lines, just as occurs in an electrical generator. If the thermosphere was unbounded, no electrical potential would exist in Earth's frame of reference; instead, electric fields are generated virtually everywhere the wind blows.

Electric fields and currents induced by tides were first discovered when it was found that the magnetic field on the surface of Earth varied regularly every day. We now know that propagating tides also surge up from the lower atmosphere, growing dramatically as they enter the tenuous upper regions of the atmosphere. Absorption of solar photons by stratospheric ozone and tropospheric water vapor is the source of the dominant, solar-driven diurnal and semidiurnal tides. During the past decade, measurements by the Upper Atmospheric Research Satellite (UARS) revealed the seasonal and interannual variability of these global waves along with the existence of other tides that do not propagate with the apparent motion of the Sun. Such so-called nonmigrating tides may generate significant longitudinal variability and will modulate the low-latitude winds that generate the electric field. We need to characterize the sources of these tides fully during the next decade.

Tides are only the tip of the iceberg as far as atmospheric winds are concerned, since planetary waves also occur at lower frequencies, while internal gravity waves are very common at higher frequencies. Gravity waves also grow with altitude and become huge in the region collocated with the ionospheric plasma. Currently of great interest is the likelihood that gravity waves induce electric fields in much the same was as tides. If they do, as recent rocket data suggest, they would have a considerable effect on ionospheric physics, space weather, and the magnetosphere. The magnetic field lines act as such good conductors that the electric field maps throughout the magnetosphere. For low levels of magnetic activity, Earth's wind field, and the associated electric field it produces, dominates the electrical structure of the inner magnetosphere.

VOLATILE WEATHER IN THE UPPER ATMOSPHERE

Compared with some other regions of the space environment, many of the processes controlling the basic thermal and dynamic structure of the ionosphere-thermosphere system are reasonably well understood. However, fundamental science questions remain, including the role of ionosphere-thermosphere processes in magnetosphere-ionosphere coupling, the cause of equatorial and midlatitude ionospheric irregularities, and upward coupling of energy and the thermal structure.

Equatorial Convective Storms

At Earth's magnetic equator, plasma in the F layer of the ionosphere, a region 200–500 km above Earth's surface, is supported against gravity by the electromagnetic force. The situation is analogous to one in which a light fluid supports a more massive fluid above it, an unstable equilibrium. The result on many nights is a massive overturning of the plasma, which can fill more than a billion cubic kilometers with highly turbulent ionized gas or plasma. When this overturning occurs, the low-density regions surge upward at high velocity due to the large electric fields that are generated in the plasma. The primary process operates at large scales, but the large velocities involved lead to structure over a range of scale sizes. In situ rocket spectra, coupled with radar observations, show that structure occurs from hundreds of kilometers down to 10 cm, over seven orders of magnitude. The plasma instability analogue of a heavy fluid on top of light fluid has been shown to account for the large-scale upwelling of plasma as the ionosphere attempts to smooth out the sharp nighttime density gradient. Some of the very first computer simulations of space plasmas were built to provide nonlinear analysis of this phenomenon.

Satellites were a valuable tool as well, giving the sort of worldwide coverage only they can provide, and we now have a deep understanding of the causes and seasonal/geographic control of this phenomenon. In the coming decade, we will begin the next step, moving into the predictive era for this important space weather problem. Crucial to success is fully understanding the equatorial electric field since it will be the single most important parameter in any prediction scheme. The Communications/Navigation Outage Forecast System (C/NOFS) satellite will be dedicated to such measurements along with other diagnostic parameters. Its goal is to use the electric field data to predict scintillations or flickering of satellite signals, which can disrupt communications. In the future we envision numerous micro- or even nanosatellites dedicated to this endeavor.

Surprises in the Midlatitude Ionosphere

The midlatitude ionosphere has generally been thought to be relatively quiescent. Compared with the stormy equatorial zone and the high-latitude zone, this is certainly true. But this is not to say that no structure exists or that space weather effects are unimportant. A major breakthrough in midlatitude studies occurred when charge-coupled device-based all-sky cameras became capable of imaging vast portions of the nighttime sky. We found that the midlatitude ionosphere indeed had weather and that we simply did not understand its origin. An example is presented here showing the complexity and beauty of the ionization clouds that form and move rapidly across the sky (Figure 3.6). Here the photos are taken in the red-line emission of atomic oxygen. This excitation occurs during a two-step process that removes oxygen ions from the ionosphere. The reaction rate depends on both the altitude of the layer and its plasma content and, when combined with other emissions or GPS data, can be used to describe the height and content of the layer. Such data may prove invaluable as we enter an era when prediction of space weather becomes feasible. The inset shows several measurements of the total electron content between the ground and GPS satellites. Huge dropouts are seen when the line of sight passes through a dark portion of the image.

For now we are still in the discovery and characterization phase for most ionospheric processes, but the pace from discovery to understanding should quicken now that two- and three-dimensional imaging and tomography can be applied. From a space weather standpoint, one of the worst difficulties involves the effect of the total plasma content between the satellite and the ground or airborne GPS user. This plasma content creates time delays, which translate into navigational errors. Ionospheric models are not even close to predicting features such as those in the Figure 3.6 image, in which huge gradients in plasma content exist. But we are getting close to the era when mesoscale fair-weather conditions can be modeled. A sizable effort has begun to calibrate simple sensors capable of providing data input analogous to the millions of data points assimilated daily by meteorological models. The global automated instrument cluster outlined in Recommendation 1 extends this concept to include all-sky imagers, Fabry-Perot interferometers, VLF receivers, passive radars, magnetometers, and ionosondes, in addition to powerful GPS-based systems in a flexible and expandable network coupled to fast real-time processing, display, and data distribution capabilities. A second thrust is to validate the models using more sophisticated observations, such as those made possible by incoherent scatter radars.

Challenges of High-Latitude Ionospheric Science

Despite 70 years of ionospheric study, it is still impossible to specify ionospheric composition and density accurately at high latitudes. Unpredicted variability occurs on spatial scales of meters to hundreds of kilometers and time scales of seconds to hours. This variability appears to be due to poor knowledge of sources, instabilities, and transport. Solar ionization by the extreme

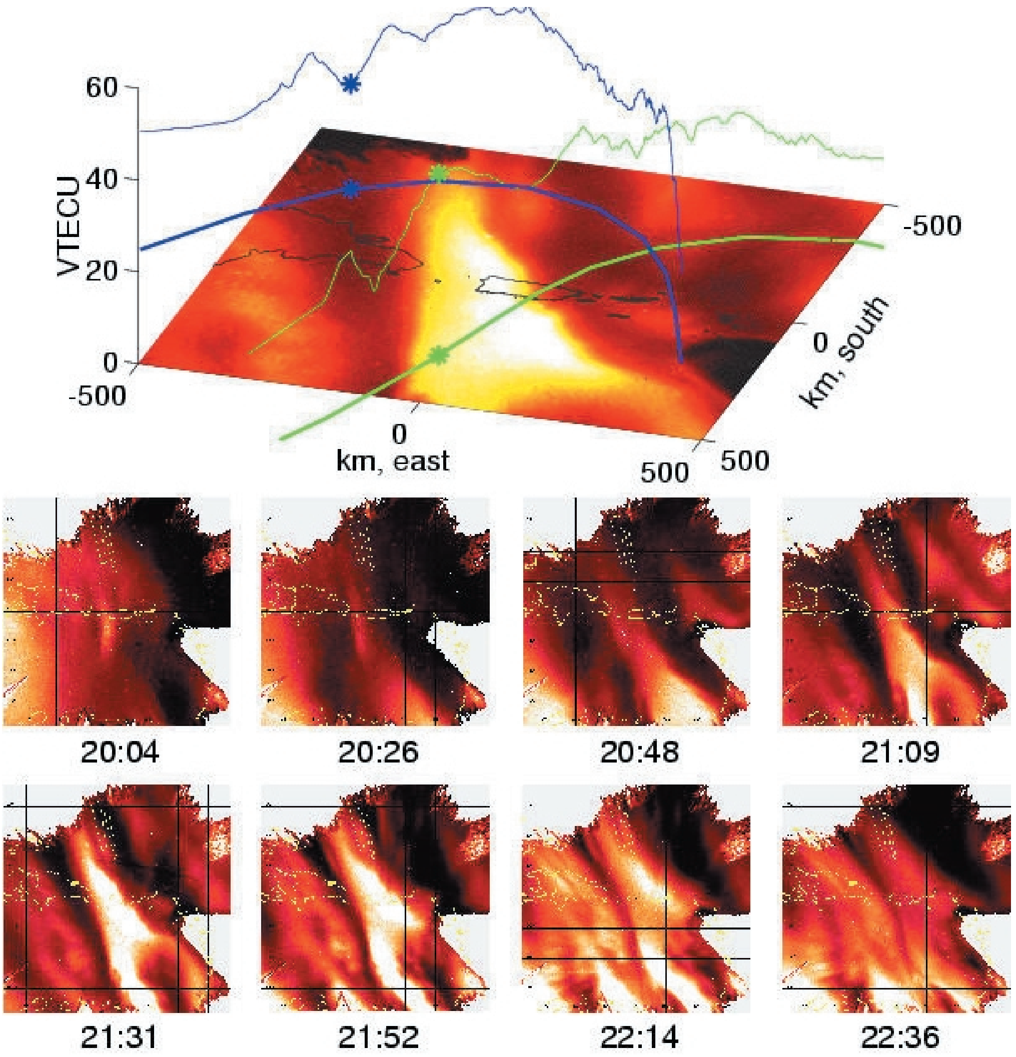


FIGURE 3.6 Observations of 630.0-nm nightglow emission made with the Cornell All-Sky Imager (CASI) at Arecibo Observatory on February 17, 1998. Each image has been projected onto a map of Earth assuming an emission height of 275 km. Superposed on the top image is the total electron content (TEC) mapped to vertical as measured by satellite-receiver pairs in the GPS constellation. Two satellite-receiver pairs are shown (satellite 10 viewed from St. Croix, Virgin Islands, blue line, and satellite 24 viewed from Isabela, Puerto Rico, green line). The ground tracks shown are for the 350-km pierce point; the satellites move from west to east. The asterisks indicate the location of the satellite at the time of the image. The increase in the 630.0-nm emission is seen to be colocated with an increase in TEC (as shown by satellite 24), while a decrease in intensity is seen to be colocated with a decrease in TEC (as shown by satellite 10). SOURCE: Unpublished figure courtesy of Jon Makela, Cornell University.

and far ultraviolet is well known except for variations during solar flares. However, particle sources of ionization are coupled in from the magnetosphere and known only from individual measurements or statistical studies. Instabilities develop where there are sharp spatial structures in the ionosphere that lead to the breakup of large-scale features and the creation of smaller-scale irregularities. Convection transport of large- and small-scale features superimpose unexpected variability on the specification predicted by ionospheric models.

Ionospheric convection due to electric fields perpendicular to the background magnetic field is well understood in principle, but its application in ionospheric models fails to match reality because the sources of variability of E cannot be prescribed adequately. The electric field is partly imposed externally, by coupling to the magnetosphere and solar wind, and partly internally, through development of polarization charges due to plasma density gradients perpendicular to **B**. This problem is expected to be solved as coupled ionosphere-magnetosphere models become mature.

Large-scale (1,000-km) ionospheric patches are features of the dark high-latitude ionosphere that persist and travel across the polar cap in a few hours. The source may be the breakup of the tongue of high ionization entering the polar cap from the sunlit dayside. Ionospheric models demonstrate that externally applied electric fields that are the result of structure in the solar wind cause the convection pattern to have periods of rapid change, when flow is broken up, causing patches of disconnected plasma. Although such a mechanism is plausible, the distribution of patches cannot be reliably predicted without specification of the solar wind source and its coupling through the magnetosphere. Once the patch has formed, it degrades by instabilities that create meter-scale irregularities in its wake. These irregularities are the source of ionospheric scintillations (scattering of radio signals) that impede communications and navigation.

The ionosphere is embedded in the neutral atmosphere, which acts as a source of ionospheric mass though ionization, thermodynamic modulation, and dynamic friction through collisional coupling and dynamic drag. This strong coupling of the ionospheric and neutral media is well represented in current models such as those at NCAR and NOAA. In developing an adequate model of the high-latitude ionosphere, the main thrust should be to extend these codes to couple with the magnetosphere and solar wind.

We are entering the era when the remaining scientific problems in ionospheric physics will be picked off one by one and solved. An important tool for this effort is the Advanced Modular Incoherent Scatter Radar and lidar facility. AMISR will be a state-of-the-art facility with a phased-array incoherent scatter radar (ISR) as the centerpiece. This highly versatile instrument will be ringed by less expensive complementary systems, typically optical in nature. The science plan is to target unsolved problems in aeronomy by placing the AMISR in appropriate geographic locations for 3 to 5 years. The first science goal is to understand the coupling of the neutral atmosphere to the high-speed current-carrying plasma in the auroral region. This coupling involves momentum transfer from the plasma to the neutrals, heating due to currents, composition changes of the thermosphere, and particle impact ionization associated with the aurora, to name a just a few aspects. The first AMISR site will be in the Fairbanks, Alaska, area to take advantage of already existing instrumentation and the Poker Flat Research Range. Subsequent sites will be decided upon with community input from a scientific advisory panel. Candidate locations include deep in the polar cap, which has never been studied using the ISR technique, and the off-equatorial zone to study effects of equatorial ionospheric upwelling and downflow along magnetic field lines to adjacent latitudes, which can severely affect communications.

One Atmosphere: Upward Coupling of Energy

One of the major unresolved issues in a quantitative description of the A-I-M system is the extent of upward coupling of energy and momentum from the lower atmosphere. The thermal structure of the mesosphere and lower thermosphere provides one of the most vivid displays of the direct dynamical coupling between the upper and lower atmosphere (Figure 3.7). The inverse temperature gradient with low temperatures in the summer polar mesosphere is due to a gravity-wave drag force that supports strong meridional circulation from the summer pole to the winter pole, accompanied by upwelling and downwelling that in turn result in cooling and warming, respectively. The drag is thought to arise from the breaking up and dissipation of small-scale gravity waves carrying momentum and energy from sources in the lower atmosphere. It is realized by modelers that waves have important effects in Earth's upper atmosphere, and that in the absence of measurements they introduce representative parameters to replace the wave sources. Important effects of these waves must be parameterized in current models of Earth's upper atmosphere. The existing schemes incorporate a number of tunable parameters that allow prescribing a realistic set of tropospheric sources. Little is known about longitude variability or about the processes involved with small-scale mixing at the base of the thermosphere (turbopause) that lead to global variations. There is an urgent need to better characterize the sources and evolution of gravity waves on the global scale. Owing to the paucity of ground-based observing sites, which inherently precludes measurements over the oceans, this can be accomplished only by satellite-borne experiments.

Solar atmospheric tides that originate in the lower atmosphere are particularly strong and ubiquitous features of the mesosphere and lower thermosphere. Upper Atmospheric Research Satellite observations reveal that

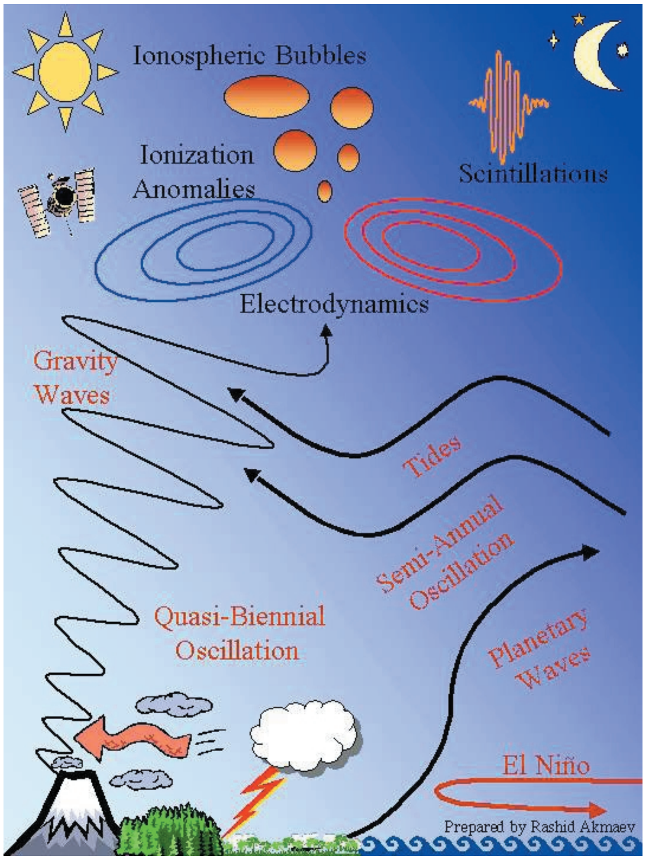


FIGURE 3.7 Illustration of the multitude of dynamical and physical processes throughout Earth's atmosphere and ionosphere, including gravity waves, tides, planetary waves, and the quasi-biennial and semiannual oscillations. Courtesy of Rashid Akmaev, CIRES, University of Colorado, Boulder.

diurnal tidal amplitudes of the meridional wind exceed 80 to 100 m/s at low latitudes at equinox, with a strong semiannual variation. Very strong variations from the day-to-day to interannual time scales of both the diurnal and semidiurnal tides are clearly seen in the UARS data and are supported by observations from ground-based optical and radar facilities. These perspectives are valuable but limited. Any satellite provides a global picture of a tide at only two local times on a given day. Continuous ground-based observations are needed, but the existing facilities are too few to provide a complete geographic perspective on these inherently global-scale waves. A complete characterization of atmospheric tides and their important effects in the upper atmosphere can only be accomplished by correlative analysis and assimilation of both ground-based and satellite-borne measurements. Further, tidal diagnostics from a satellite constellation are needed to fully resolve the spatial and temporal variability of atmospheric tides.

In the equatorial, low-latitude, and mid-latitude ionosphere, a number of investigations have interpreted oscillations in plasma density and electrodynamics as a manifestation of planetary waves in the neutral atmosphere with 2-, 5-, 10-, and 16-day periods. Several interpretations have been suggested, including modulation of the turbopause, excitation of normal modes in the upper atmosphere, and the modulation of other upward-propagating waves including tides and gravity waves. Recent studies also show that between 15 and 25 percent of observed ionospheric electron density variations cannot be ascribed to space weather or geomagnetic activity but are probably attributable to waves of tropospheric origin. The wave drivers can only be resolved with additional satellite measurements that are analyzed in concert with observations from ground-based networks.

Comparatively little is known about the interactions between these various waves and the impact they have on the upper atmosphere of Earth. There is an urgent need to develop comprehensive observation and modeling programs to untangle the complex web of interacting processes that occur over scales from kilometers to thousands of kilometers. It is very likely that many of the outstanding questions about the thermal and dynamic structure of the upper atmosphere will turn out to be linked to the labyrinth of global wave coupling processes.

Plasma-Atmosphere Interactions

Collisional interactions between ionized and neutral particles are a dominant process in the upper atmosphere. At high latitudes, the imposed magnetospheric electric fields set the plasma into motion through collisions with the dense neutral medium and can drive strong neutral winds. The atmospheric resistance to ion motion gives rise to dissipation in the form of frictional heating, raising neutral and ion temperatures, driving wind surges and upward ion outflow, and changing the global neutral composition and circulation. In regions of large electric fields, anomalous electron collisions due to plasma instabilities can change conventional conductivities, and variability or fluctuation in the electric field can induce uncertainties in heating rates by a factor of 2.

At midlatitudes the collisional interaction between plasma and the atmosphere is reversed. In the presence of inclined magnetic fields, neutral winds, driven by global pressure gradients, force plasma up and down in the direction of the magnetic field. During space weather events the wind surges, propagating from high latitudes, and circulation changes can drive winds that are many times their normal quiet-time values, which can cause a substantial redistribution of plasma. Winds can push ions to regions of altered neutral composition, either increasing or decreasing plasma densities. In addition, the circulation changes drive upwelling at high latitudes, causing neutral composition changes that can be transported by the background and storm-induced winds and further modify the plasma structure at all latitudes.

Differing mobilities of electrons and ions in the presence of magnetic and electric fields also give rise to field-aligned currents and vertical electric fields. Tidal waves propagating from the lower atmosphere grow to large amplitudes in the lower thermosphere before they dissipate and, through collisional interaction with the plasma, drive a global current system. During geomagnetic disturbances both E- (~100 km) and F-region winds can be disrupted and decrease or reverse normal quiet-day electrodynamics. At low latitudes, with the near-horizontal magnetic field, the neutral wind-driven electric fields dominate the plasma structure, giving rise to the equatorial ionization crests on both sides of the magnetic equator and anomalous wind and tempertaure features. The redistribution of plasma sets up conditions conducive to the generation of equatorial convective storms. We are only just beginning to understand the complexity of plasma atmosphere coupling during solar-wind-driven storms, when the neutral wind fields and electrodynamics are disturbed.

The Geospace Electrodynamic Connections mission, currently scheduled for launch in 2009, will make systematic multipoint measurements and when combined with ground-based observations will delineate and bring to closure our understanding of the key roles the ionosphere and thermosphere play in the Sun-Earth connection. This cluster of dipping spacecraft will resolve spatial and temporal variations and energy transfer in the transition region between the magnetosphere and the ionosphere and thermosphere. GEC will identify the relative importance of key processes, including energetic inputs, current systems, ionosphere-thermosphere coupling, and thermospheric waves on varied spatial and temporal scales.

Although many of the physical processes in plasma-atmosphere interactions are understood, the aspect that is still poorly characterized is the global-scale dynamic response to solar events. There is a pressing need to understand the interactions between global-scale neutral dynamics, composition changes, plasma density and conductivity, and electrodynamics, particularly during geomagnetic storms, when variability is at its greatest. Previous single-satellite missions and ground-based observations have provided a valuable but limited perspective. Remote sensing, such as with the GUVI instrument on the recently launched TIMED satellite, has the potential to follow global-scale neutral composition, but simultaneous in situ observations are required to interpret and relate the images to the actual conditions present in the atmosphere. Pairs of satellites can begin to address the important scale sizes in the plasma density and neutral dynamics, but ultimately, multispacecraft missions will be required to track the global-scale dynamic response. The technology that will enable such nanosat constellations in low-Earth orbit is coming soon and must be a priority in future missions.

Photochemistry and Radiative Transfer

The ionosphere is created and maintained by energetic solar photon irradiance in the extreme ultraviolet and x-ray regions of the spectrum that ionizes a small fraction of the neutral atmosphere. Absorption of solar EUV radiation at other wavelengths also heats the highest part of the neutral atmosphere. At longer ultraviolet wavelengths, solar radiation penetrates progressively deeper into the atmosphere, and its main effect is photodissociation of molecular gases. The deposition of this energy drives a complex cycle of photochemical response that interacts strongly with atmospheric transport. The thermosphere becomes mostly atomic above ~200 km, while the lower thermosphere from ~100 to 200 km acts as a crucible for photochemical energy deposition and dynamical interchange of atoms and molecules.

Solar cycle variation of the solar spectrum increases with increasing energy, from roughly a few percent in the near- to midultraviolet to about double in the extreme ultraviolet and by an order of magnitude or more in the x-ray region. Instruments on UARS provided accurate measurements of the ultraviolet spectrum adjacent to visible wavelenths throughout the 1990s. The extreme ultraviolet and x-ray regions were measured during the 1970s by the Atmospheric Explorer satellites. After a prolonged hiatus, the Solar Extreme Ultraviolet Experiment (SEE) instrument on TIMED began making new measurements in December 2001. Several problems in thermospheric photochemistry remain outstanding, leading to a long-standing suspicion of problems with the older short-wavelength data. New observations of solar soft x rays (from about 1 to 20 nm) by a combination of rocket measurements and data from the solar x-ray photometer on the Student Nitrous Oxide Explorer (SNOE), a UNEX-class satellite, have found fluxes four or five times higher than the Atmospheric Explorer-era estimates at all levels of activity. SEE is also observing these higher fluxes. Applied to photochemical and general circulation models, the ionospheric and minor species problems now appear to be solved. This leads to a general revision of thermosphere/ionosphere variability, with larger solar cycle variation expected. Other unresolved aeronomical issues, including discrepancies between measurements and model predictions, must be explored in the context of solar irradiance measurements and variability. It must be remembered that although the solar EUV flux and the upper thermosphere response have been measured in the past, comprehensive, simultaneous measurements of both the source and the atmospheric response are still lacking. Since solar EUV is the dominant and highly variable energy and ionization source in the thermosphere and ionosphere, it is imperative that the relationship between EUV flux and the state of the ionosphere-thermosphere system be assessed quantitatively.

The mesopause is particularly amenable to remote-sensing observation of airglow emission layers. Measurements of both airglow and auroral emissions by ground-based CEDAR instrumentation have contributed significantly to our understanding of this region of the atmosphere. The TIMED mission will bring even greater advances. TIMED is remotely measuring winds, temperature, composition, and cooling rates throughout the mesosphere and lower thermosphere as well as auroral energy inputs, and it also carries solar EUV and x-ray instrumentation to quantify the primary sources of energy. Our understanding of chemical-dynamical coupling should vastly improve with input from TIMED measurements. But there are fundamental and unresolved issues that TIMED cannot decipher, including the characterization of gravity waves, which are the underlying drivers of the basic state of the mesopause region and the inherent spatial and temporal tidal variability of the upper atmosphere.

Particle Effects in the Middle Atmosphere

The particle flux from the Sun and the magnetosphere represents a large source of energy and ionization for the lower thermosphere and ionosphere. The energy flux, which varies by two orders of magnitude, is deposited by particles with energies that range from hundreds of eV to several hundred MeV. The intensity, spectrum, and distribution of the precipitation are functions of solar and geomagnetic activity. The mean values also show longer-term dependence on the solar cycle. The particle-induced ionization leads to dissociation of tightly bound N_2 molecules and to the formation of the reactive nitrogen compounds NO and NO_2. These compounds, when transported to lower altitudes, participate in a catalytic cycle leading to destruction of ozone. Continuing research seeks to understand the impact of magnetospheric and solar energetic particles on the chemistry and electrodynamics of the middle atmosphere.

The production of NO and NO_2 due to the precipitation of energetic particles is determined by the deposition of energy from particle penetration. The particle energy is absorbed by the atmosphere through collisions with, and scattering by, atmospheric constituents, which result in either dissociation or ionization of the atmospheric constituents. It takes an energy deposition of approximately 35 eV to produce each electron pair in the atmosphere. In addition to the energy imparted directly to the atmosphere during collisions, a certain amount of energy is converted to x rays by the bremsstrahlung process. This is associated with the rapid deceleration of energetic electrons during their penetration into the atmosphere, providing additional ionization down to altitudes of 20 km.

The ion and neutral odd-nitrogen chemistry initiated by energetic particles produces secondary electrons, which in turn ionize and dissociate the major atmospheric species. In the sunlit atmosphere, the photodissociation of NO is important; however, in fall, winter, and spring in the polar region, the photolytic reaction is negligible and the resulting NO lifetime is sufficient for NO to be transported to the mesosphere and stratosphere. There have been several studies of these effects suggesting that precipitating electrons could lead to the formation of oxides of nitrogen and hydrogen. These could affect ozone as low as 60 km, and the ion formation rates due to these electrons could dominate ion formation rates of other processes down to approximately 40 km. Studies of ion formation rates using electron measurements from geostationary orbit have shown that the fluxes of energetic electrons and hence the ion formation rates are strongly modulated by solar activity. Simulations suggest that precipitating electrons could significantly affect the budgets of both odd nitrogen compounds and ozone within the stratosphere. This coupling mechanism between solar activity, solar wind structures, the energetic particle populations within the magnetosphere, and the chemical state of the middle atmosphere represents an important linkage that must be understood to assess the natural variations in the global chemical state of the middle atmosphere.

MICRO- AND MESOSCALE CONTROL OF GLOBAL PROCESSES

The global behavior of the thermosphere, ionosphere, and magnetosphere is critically dependent on small-scale processes in some key locations. Although it is not intuitively obvious that processes on small scales can have global consequences, there are some that are essential to the existence of large-scale structures as we observe them. It is now understood, for example, that the thermal structure of the mesosphere and its circulation patterns are much more dependent on energy and momentum carried by small-scale waves than on global thermal sources such as solar ultraviolet radiation. Similarly, kinetic effects due to turbulence and heating are often more effective than classical diffusion in determining the diffusion rate. More dramatically, some structures and boundaries would not exist at all without small-scale processes. The magnetopause is a thin shell at the nose of the magnetosphere that has spatial scales comparable to the ion gyroradius. Auroral arcs would be diffuse and unable to create the observed dramatic, localized effects without small-scale acceleration regions that form spontaneously a few thousand kilometers above Earth and create the kilometer-scale structures. The importance of small-scale processes is well understood, but their simulation in global-scale models is made difficult because they are sub-grid-scale and must be parameterized. Inclusion of properly formulated physical representations of these sub-grid-scale effects is an urgent research requirement.

Transfer of Momentum and Energy by Internal Atmosphere Waves

Earth's dense and dynamic troposphere is a source of very strong internal waves, which carry energy and momentum upward. To conserve energy in an exponentially decreasing atmosphere, these waves grow in amplitude and eventually break, depositing their energy and momentum in the mesosphere. A remarkable circulation pattern is set up, leading to temperatures in the summer polar regions, in full 24-hour sunlight conditions, as low as 100 K. This is the coldest fluid in the inner solar system and results in the highest clouds on Earth (see Figure 3.2). These clouds seem to have become more frequent over the last 100 years and may be a sensitive marker for global change. Waves also seem to interact with atmospheric tides all over the globe to create winds in the upper mesosphere and lower thermosphere in excess of 500 km/hr, which is nearly supersonic. Models do not come close to explaining these high winds, which create conditions for velocity-shear-driven instabilities of the neutral atmosphere in regions long thought to be quite stable. Understanding the role of medium- and short-period waves in the energy and momentum budget of Earth's upper atmosphere is a crucial goal of the next decade.

Mesoscale Auroral Features and Their Role in Auroral Oval and Magnetosphere Processes

Thermospheric vertical mixing in the auroral oval is driven by both global and local processes. The global process establishes an equilibrium for each constituent. It is driven by the temperature structure of the auroral thermosphere. By this mechanism alone, molecular species can be elevated to F-region heights when the thermospheric temperatures are raised during an auroral substorm (see next section). Vertical winds cause localized upwellings and downwellings and are much more effective at redistributing the thermosphere during heating events. Coupled regions of upwelling and downwelling tend to neutralize each other on global scales but serve to raise molecular species very rapidly at the center of auroral heating. Upward fluxes of molecular species are significantly enhanced due to the upwelling process, with a resulting increase in molecular concentrations at F-region heights, affecting the time constants for change in ionospheric concentration.

Recent results from the FAST and Polar satellites have indicated that the narrow-scale plasma structures that produce auroral arcs are confined to altitudes below about 6,000 km, where the bulk of auroral acceleration appears to occur. These results, along with theoretical considerations, suggest that energy enters the auroral zone at intermediate and large scales and then filaments into narrow structures due to magnetosphere-ionosphere interaction. This filamentation is probably due to a combination of ionospheric conductivity effects and the formation of density holes on a variety of scales in the auroral zone. Understanding these complicated interactions is a high priority for future research.

Understanding the scale sizes of the parallel electric fields that accelerate auroral particles is an important and active area of research. Models based on the reflection of electrons out of a converging magnetic field produce smoothly varying, weak electric fields. On the other hand, auroral zone observations in regions of both upward- and downward-going current indicate that localized regions of weak parallel electric fields are distributed along auroral field lines. Some recent observational and theoretical results suggest that the auroral potential drop may occur as a result of large parallel electric fields, usually referred to as double layers or sheaths. Determining the relative contribution of these various structures would probably require a multispacecraft auroral mission.

Convection in the outer magnetosphere is strongly affected by the ionospheric drag caused by collisions with the neutral thermosphere. When parallel electric fields are present, however, magnetospheric convection can become uncoupled from this ionospheric drag, leading to the possibility of fast flow channels. Investigations of magnetosphere-ionosphere coupling and its relation to substorms are still in an early stage; however, it will be a promising area of research in the next 10 years as the dynamics of the magnetosphere as a coupled system becomes better understood. The multiscale dynamics of the auroral zone, in which phenomena from narrow auroral arcs to large-scale magnetospheric convection are clearly interrelated, is a prime example of how smaller-scale processes can have large-scale effects.

Connection and Reconnection of Magnetic Field Lines in the Magnetosphere

As noted in the section on electrodynamic coupling, the highly conductive magnetic field lines electrodynamically couple different magnetospheric regions. However, this frozen-in condition does not hold everywhere, and its violation in localized regions has a strong impact on magnetospheric dynamics. For example, a strict application of the frozen-in condition implies that the magnetopause is an impenetrable boundary since infinitely conductive plasmas cannot penetrate or diffuse across field lines. However, a continuous transport of mass, momentum, and energy from the solar wind into the magnetosphere is observed. To understand this transport has become one of the most important challenges in magnetospheric physics.

Magnetic reconnection of oppositely directed magnetic field lines usually takes place in strong current sheets such as those that occur at the magnetopause and in the tail. Observations in the magnetosphere have shown that the orientation of the interplanetary magnetic field is a major factor in controlling geomagnetic activity, ionospheric convection, the structure of the magnetic field, and the dynamics of plasma flows at the magnetopause. Accordingly, magnetic reconnection has been widely accepted as the main mechanism for coupling the solar wind to the magnetosphere-ionosphere system. At Earth's magnetopause, reconnection results in the formation of open magnetic field lines, leading to exchange of mass along open flux tubes and an efficient transfer of solar wind momentum and energy by magnetic stresses. Nightside reconnection is considered to be critical for releasing magnetic energy stored in the tail.

The breakdown of the frozen-in condition is often described in terms of the electron behavior as described by a generalized Ohm's law. In the recent GEM reconnection challenge, magnetic reconnection was studied in a simple current sheet configuration under a specified set of initial conditions. These conditions were modeled by different codes, ranging from fully electromagnetic particle codes to conventional resistive magnetohydrodynamic codes. The rate of reconnection obtained from the different models was found to be about the inflow speed, which depends on external magnetic field strength and density, and to be insensitive to the mechanism for breaking the frozen-in condition.

The observations indicate that the coupling of the solar wind and the magnetosphere by reconnection is controlled by global parameters rather than localized microinstabilities or other kinetic effects. Observations also show that magnetic shear is not the only factor that controls the dynamic processes of the solar wind–magnetospheric interaction. Larger-scale causes and effects of reconnection are important since most measurements sample the external region around the reconnection site. Both case studies and statistical studies based on numerous in situ observations provide evidence for large-scale quasi-steady reconnection and patchy intermittent reconnection. Because observations in the narrow region where magnetic field and plasma disconnect are difficult to obtain, observational tests of the theoretical predictions of reconnection have focused on locations outside this diffusion region. Comparison of the observations with the testable predictions given by theory and simulation provides valuable information for reconnection research.

Over the next decade, a systematic effort will be needed to establish and develop a fundamental theoretical understanding of the dynamics of magnetic reconnection. The lack of a complete dynamical model for interactions at the magnetopause has led to a number of controversies in magnetopause physics and our understanding of substorms. In addition to considering the localized breakdown of the frozen-in condition, attention also needs to be focused on the reconfiguration of the plasma after reconnection has taken place. Observational studies that can refute or verify the results from such theory and simulations are clearly necessary. Multipoint spacecraft missions such as Cluster II and Magnetospheric Multiscale are required to separate space and time variations and to simultaneously sample the plasmas in the inflow and outflow regions of a reconnection site. A direct interaction between the theoretical predictions and analysis of data from spacecraft and ground-based instruments will be needed to obtain a complete understanding of reconnection processes.

Magnetic reconnection is a universal process that also occurs in phenomena such as solar flares, coronal mass ejections, the galactic dynamo, and astrophysical accretion disks, as well as in laboratory plasmas. However, the magnetosphere is the only cosmic plasma where reconnection processes can be observed in considerable detail. Understanding of the reconnection process in the magnetosphere will be a significant step toward the understanding of related phenomena occurring in astrophysics, solar physics, and laboratory plasma physics.

DYNAMICS OF GEOMAGNETIC STORMS, SUBSTORMS, AND OTHER SPACE WEATHER DISTURBANCES

There are a number of different modes of geomagnetic activity, including magnetic storms and substorms. A magnetic storm occurs when merging of the interplanetary magnetic field with Earth's magnetic field causes deep and intense circulation of the magnetospheric plasma, building up the energy content of the ring current to unusually high levels. The close relationship between strong southward interplanetary magnetic field that lasts for 3 hours or more and large magnetic storms is a consequence of the central role of convection. Substorms, on the other hand, are impulsive events that unload energy stored in the magnetotail. A period of southward IMF triggers enhanced reconnection at the dayside magnetopause, decreasing magnetic flux on the dayside and adding magnetic flux (and thus stored magnetic energy) to the magnetotail. Substorms return magnetic flux to the plasma sheet and dayside magnetosphere from the magnetotail. They can occur multiple times within a magnetic storm period but also several times per day even in the absence of storm activity. The nature of the coupling between storms and substorms is an unresolved issue that is receiving much attention.

It should be noted that storms and substorms are not the only dynamic occurrences in the magnetosphere. Other features that have been identified include poleward boundary intensifications and associated plasma sheet flow bursts, the magnetospheric response to solar wind dynamic pressure enhancements ("dynamic pressure disturbances"), and steady magnetospheric convection intervals. Poleward boundary intensifications, which are enhancements in the aurora that occur at the poleward edge of the auroral oval, can strongly affect the entire plasma sheet and auroral oval. Dynamic pressure disturbances lead to dramatic increases in the magnetopause, tail, field-aligned, and ionospheric current systems. They also lead to large enhancements in essentially all energetic particle populations within the magnetosphere, in auroral particle precipitation, and in large-scale reconnection rates along the open-closed field line boundary. Periods of steady enhanced convection (e.g., convection bays) lead to large increases in fluxes in the inner plasma sheet and to large and long-lasting enhancements in morningside auroral electron precipitation.

We have known for some time that the nature of an effective solar wind driver depends on the type of geophysical response in question. For example, high-speed solar wind streams with embedded magnetic field fluctuations produce strong radiation belt enhancements and prolonged substorm activity but very weak magnetic storm responses. However, the A-I-M system is not a passive element responding in a prescribed way to the solar wind. The state of the system is a function of coupling and feedback processes determining the effectiveness of solar wind forcing. Over the next decade much work will be done on the underlying physical processes and coupling mechanisms that alter the geoeffectiveness of solar wind drivers and possibly modify the distribution of solar wind energy within the A-I-M system. It is expected that fresh insights into the interpretation of statistical data sets and of correlations between magnetic and solar wind indices will result and that previously unappreciated geoeffective elements in the solar wind will be identified. Future progress in understanding the response of the magnetosphere, ionosphere, and upper atmosphere to the energy inputs from the Sun and solar wind requires an integrated approach that views the A-I-M as a tightly coupled system.

Sources and Losses of Radiation Belt Electrons

Earth's radiation belts are situated in the closest region of near-Earth space, which hosts the majority of commercial and military spacecraft in orbit around our planet. While the particle populations in this region have been measured since the dawn of the space age, the physical mechanisms underlying some of the most fundamental properties of the belts are not yet understood. In particular, we do not know (1) the primary means by which the radiation belt electrons are accelerated to energies of hundreds of keV to many MeV, (2) the relative roles of different types of waves in the loss of the particles via resonant wave-particle interactions, or (3) why some geomagnetic storms result in flux enhancements while others do not. While theoretical models of radiation belt dynamics do exist, they are not yet established enough to allow accurate, reliable predictions of dynamical response to solar and geomagnetic activity.

From the point of view of basic plasma physics, the radiation belts constitute the A-I-M region within which mechanisms and effects of wave-particle interactions are manifested in their richest variety. This region harbors a variety of plasma waves that can resonantly interact with the particles, scattering them in energy and pitch angle and alternately accelerating them or causing them to precipitate out of their trapped orbits. The waves involved range in frequency from mHz to the electron gyrofrequency (kHz), and the underlying physical processes are typically highly complex, requiring a combination of macroscale and microscale modeling.

The ionosphere and the solar wind are two plasma reservoirs which are the ultimate sources of radiation belt electrons. Since the temperature of both of these plasmas is <10 eV, significant acceleration must occur for radiation belt electrons to attain energies up to several MeV. The physical mechanisms by which electrons are accelerated are not yet understood. The identification of the primary acceleration mechanisms is needed for a quantitative understanding of radiation belt dynamics and for a proper assessment of radiation effects on satellites. While there is general agreement that electron acceleration must involve the violation of one or more of the adiabatic invariants (approximate constants of cyclotron-, bounce-, and drift-motion), candidate acceleration mechanisms under consideration include radial diffusion and scattering by wideband electric or magnetic field fluctuations, drift resonance with ULF (mHz) waves, and resonance with waves near the electron gyrofrequency. At present, we do not know (1) the wave modes that can resonate with the particles over the very wide range of particle energies involved, (2) the locations of the primary acceleration regions and the importance of the plasma parameters and propagation conditions for efficient acceleration, or (3) the time scales for wave-driven acceleration. Resolving these questions will require simultaneous and multipoint measurements of waves and particles before, during, and after storm-time acceleration.

The primary loss processes for trapped radiation belt electrons consist of precipitation resulting from pitch angle scattering of the particles via wave-particle interactions and/or Coulomb collisions. Coulomb collisions with the dense atmosphere dominate at low altitudes, while wave-particle interactions are generally believed to be the agents of loss at higher L-shells. The waves that are responsible include whistler-mode waves (named for their descending frequency vs. time at a fixed point away from the source), generated in the magnetosphere and by lightning discharges, as well as by VLF transmitter signals. However, the relative roles of different waves and the regions over which they are dominant are not yet known. Quantitative understanding of loss processes is important for modeling the relaxation and quiet time structure of the belts. Second, quantification of electron loss during disturbed times sets constraints on the minimum amount of acceleration required. Electron loss is thus taking on a new significance in the effort to understand the processes responsible for electron acceleration. Third, precipitating radiation belt particles can penetrate to altitudes of between 40 and 70 km, where they can affect the chemistry and dynamics of the lowest regions of the A-I-M system. Assessment of the global effects of such precipitation must be based on a quantitative understanding of electron loss rates under different conditions.

Response of the Ring Current

Observations of the response of the magnetosphere to strong driving by interplanetary magnetic clouds have revealed new features of ring-current dynamics and coupling that will be a focus of research efforts in the next decade. IMCs can have time scales for southward IMF as long as 12 hours and a slow, smooth rotation of southward to northward IMF as the cloud moves past Earth. Large solar wind dawn-to-dusk electric fields associated with these southward interplanetary magnetic fields are mapped down along magnetic field lines across Earth's polar cap and magnetotail. These electric fields are associated with a strong convection that draws plasma from the magnetotail deep into the inner magnetosphere. Ions moving into the inner magnetosphere are energized as they drift across electric equipotentials to form the bulk of the storm-time ring-current populations. Thus the magnitude of the cross-tail electric potential is an indicator of the energy available for accelerating ring-current ions. The ring-current ions generate an electric field that shields the inner magnetosphere from the convection electric field. In the time-dependent case, there can be overshielding or undershielding of the convection field. Evidence for overshielding has been found in images of a shoulder feature in the outer boundary of the plasmasphere by the IMAGE spacecraft.

As a consequence of the long time scales associated with the IMC, plasma sheet ions, which move on open drift paths into the inner magnetosphere, are not captured on closed drift paths but make one pass through the inner magnetosphere before being lost at the dayside magnetopause. This produces an asymmetric ring current, with up to 90 percent of the particle energy flowing along open drift paths during the buildup of the ring current, referred to as the main phase of a storm. When the solar wind east-west electric field weakens, open drift paths are converted to closed, trapping the ring current in the inner magnetosphere. This trapped ring current is symmetric, requiring no ionospheric closure current, and decays slowly through collisions with neutrals and the ionized plasmasphere and through scattering of ions in the fields of plasma waves. New energetic neutral atom (ENA) images of the ring-current structure and evolution from the Polar and IMAGE missions clearly reveal the asymmetric open drift paths of the main-phase ring current and its gradual conversion to a symmetric form as the IMF turns northward. Calculations of global ring-current loss rates using ENA images document the rapid loss lifetimes associated with the asymmetric ring current. Clearly ring-current intensity and evolution are a function of both the injection strength, as represented by the large-scale electric field, and the source strength, which includes solar wind and ionospheric contributions.

New information from recent superstorms indicates that the injection strength as measured by the convection potential is not simply a function of the solar wind east-west electric field strength, but is modified significantly by the A-I-M system. This surprising result indicates that the A-I-M system actively limits the amount of solar wind energy input. The situation is further complicated by variations in plasma sheet density, which may determine the relative strength of superstorms when the convection potential has become saturated. Clearly the structure and dynamics of the large-scale electric field and its impact on the inner magnetospheric populations must be a major focus of magnetic storm and inner magnetospheric research in the coming decade.

It is also clear that ion source strength is important in determining ring-current intensity, composition, and dynamical variation during magnetic storms. Responding to strong solar wind driving, the plasma sheet can become enriched in ionospheric ions and structured in the radial and azimuthal directions, affecting the ring-current strength and subsequent decay. If a superdense plasma sheet population moves into the inner magnetosphere in response to a storm-time convection electric field, it has a major impact on the strength of the ring current being formed at this time. A sharp decrease in plasma sheet density can trigger ring-current decay even in the presence of strong solar wind driving. A major focus of the next decade will be on understanding the dynamical variations processes responsible for loss of magnetotail plasma and refilling (from both ionospheric and solar wind sources) during the elevated magnetic activity associated with storms.

Coupling of Magnetotail Dynamics to the Ionosphere

Earth's magnetotail couples to the ionosphere at high latitudes. A dynamic, visual reminder of this coupling is provided by auroras in the Northern and Southern Hemispheres. In the past decade, progress has been made in understanding pieces of this coupling, much of which occurs at the boundary of the plasma sheet and the low-density region of the tail called the lobes, the plasma sheet boundary layer (PSBL), shown in Figure 3.1a. Although this region was identified more than 15 years ago, many early observations were made by equatorial spacecraft at large radial distances. From such measurements, the boundary was shown to have spiky, large-amplitude electric fields, significant plasma flow, including bursty bulk flows, counterstreaming electron and ion beams, enhanced broadband electrostatic noise, and field-aligned currents.

Many of the same features are also observed at low altitude, such as field-aligned particle beams, field-aligned currents, and large-amplitude electric fields. More recently, observations from the Polar spacecraft have confirmed the connection between the PSBL in the tail and the ionospheric auroral zone. These observations have included detailed comparison of plasma distribution functions, which show evidence of auroral acceleration. There is also evidence for field-aligned current systems and large electric fields. The large electric fields are located predominantly at the poleward edge of the high-latitude plasma sheet observed by Polar and appear to be correlated with substorm activity. The electric and magnetic fields have also recently been shown to carry significant energy flux directed toward Earth, which powers the aurora at ionospheric heights.

To date, the existing studies of the magnetotail-ionosphere connection are mainly statistical, with a few case studies. The circumstantial case for this coupling is excellent, but there remain many questions about the details of the process. Just how electromagnetic energy flux is generated in the tail and converted to particle energy is not known. How the field-aligned currents are generated in the tail and how they connect to the auroral zone is understood only qualitatively. NASA's existing and planned multi-spacecraft missions will shed light on these processes; however, most of the envisioned missions involve a small number (between 2 and 5) of spacecraft at relatively small separations. To more clearly tie down the connection between the magnetotail and ionosphere, spacecraft at high and low altitudes along roughly the same field line are needed. Within existing data sets, few conjunctions exist, reducing the likelihood of clear signatures for the coupling processes. A multispacecraft magnetosphere-ionosphere coupling mission, combined with suborbital and ground-based measurements, is within the cost range of NASA MIDEX and suborbital resources and NSF resources. The planned MagCon mission will provide much improved temporal and spatial resolution giving global coverage of the tail region.

Storm–Plasma Sheet Coupling

One of the most important dynamical processes in the magnetosphere is the substorm cycle, which produces intense electric fields and energized plasma distributions moving earthward and tailward. New information that bears on the relationship between storms and substorms has recently been obtained by ENA imaging. A comparison between ENA images of isolated and storm-time substorm plasma injections reveals important differences between these types of events. Storm-time ion injections are more prolonged and suggest that a strong, quasi-steady, cross-tail electric field is a cause of the transport rather than the periodic stretching and collapsing of the magnetic field, which is characteristic of the substorm cycle. In addition, the magnetospheric region participating in the storm-time ion injections is much broader, encompassing most of the nightside inner plasma sheet region. Both ENA images and ring-current models suggest that the enhanced convection electric field during storm activity moves ion injections deep into the inner magnetosphere, where they are strongly energized to form the storm-time ring current. In the absence of strong convection, during isolated substorms, these injections have essentially no impact on ring-current formation but act to enhance the cross-tail current at large radial distances.

We already know that long intervals of strong convection are necessary to produce an intense ring current. Variable substorm electric fields may play an important role in diffusing energetic ions inward to form the high-energy tail of the ring-current distribution but are not responsible for the major portion of ring-current energization. It appears that substorms modify the plasma distributions in the inner plasma sheet, creating high-temperature source populations for the ring current. Even more dramatic modifications to the plasma sheet density occur in association with severance and loss of the outer portion of the plasma sheet, which sometimes occurs. This substantive loss of magnetotail plasma is important to magnetotail energetics.

Cold and Hot Plasma Interactions at the Plasmasphere–Ring Current Interface

Innovative imaging techniques are already revealing new details of the strong coupling between hot and cold plasma populations in the inner magnetosphere. These details have implications for ring-current decay; plasma wave generation, propagation, and damping; and thermal plasma energetics and dynamics. Ring-current precipitation loss is enhanced on occasion in a region coincident with the dusk bulge, and plasmaspheric tails (also called drainage plumes) are formed in response to changes in the global convection pattern (Figure 3.8). An array of troughs and bite-outs in the thermal plasma structure appear to be signatures of the structures in the global electric field produced by ring-current penetration electric fields. Finally, observational evidence exists that large-scale electric fields associated with partial ring-current closure and ring-current shielding effects dramatically alter the duskside thermal plasma dynamics. The structuring of the plasmasphere has important consequences for the growth and propagation of a variety of plasma waves. These waves are believed to be important in producing loss of ring current and radiation belt ions and electrons as well as in redistributing energy among different plasma species.

The ring current is a major heat source for the plasmasphere. However, the temperature structure of the plasmasphere is not related in a simple way to magnetic activity. The details of this relationship are important for global models of the inner magnetosphere. The contribution of ring-current ion precipitation to the heating

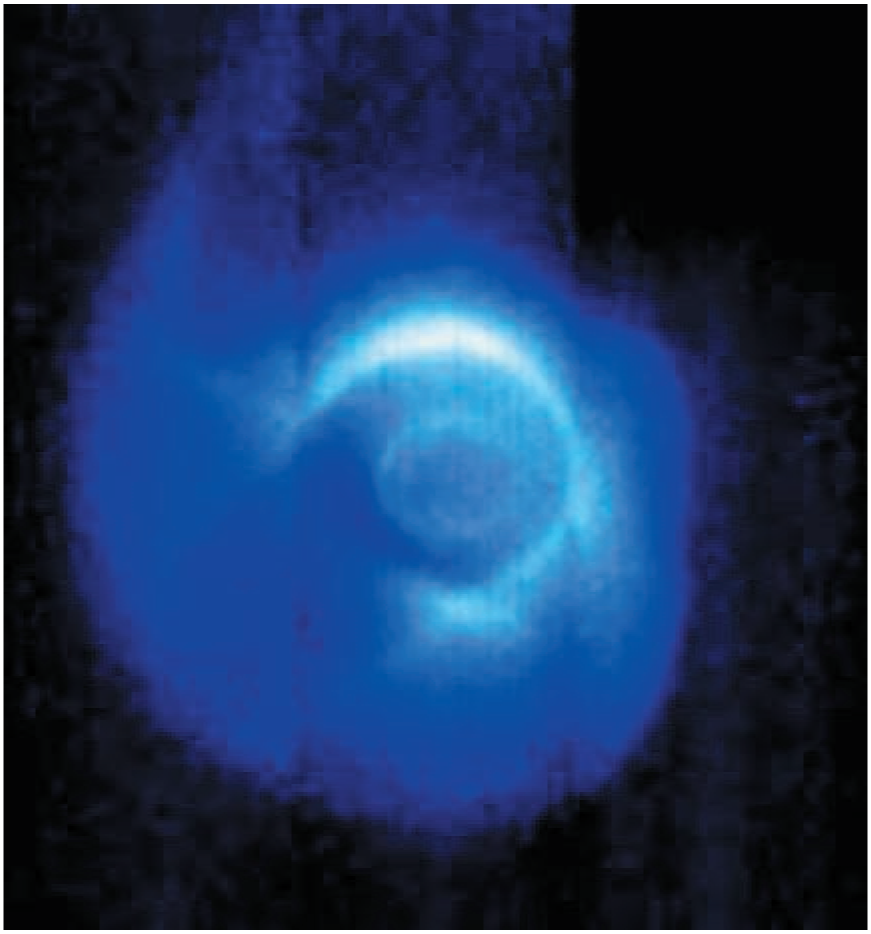


FIGURE 3.8 Plasmasphere and auroral oval as viewed by EUV instrument on the IMAGE satellite, showing plasma on dusk side (left) swept toward magnetopause (top) during a period of enhanced magnetospheric convection electric field. Image taken on August 11, 2000, 1755 UT. Courtesy of B.R. Sandel, University of Arizona.

and ionization of the neutral atmosphere (at subauroral- and midlatitudes) and to alterations in the minor species chemistry is not yet understood. Subauroral ion precipitation zones intensify and move equatorward as magnetic activity increases. This ion precipitation is a time-varying mixture of oxygen and proton components, which have very different interactions with thermal ion and neutral populations. The rates of volume heating and ionization of the neutral atmosphere by precipitating ring-current oxygen ions can equal or exceed daytime solar heating and ionization rates. Precipitating protons produce lower heating rates for comparable characteristic energies and energy fluxes, but even higher ionization rates, which peak at lower altitudes. In certain regions, protons have been shown to make an important contribution to E-region ionization.

SOLAR VARIABILITY AND CLIMATE

Solar Variability and Middle Atmospheric Chemistry

The middle atmosphere, extending from about 10 or 15 km altitude upward to about 90 km, is a region whose chemistry is dominated by minor constituents, and especially by ozone. Ozone (O_3) is formed through dissociation of O_2 by solar ultraviolet (UV) radiation, producing atomic oxygen which reacts with O_2 again to form ozone. The concentration of ozone is determined by a balance between its production rate and its loss rate through photodissociation, chemical reactions, and transport processes. Photodissociation of ozone by solar UV radiation has two important effects: It is the principal source of heat for the middle atmosphere, and it protects Earth's surface from UV radiation, which can destroy proteins and nucleic acids such as DNA. Indeed it is widely thought that life could not have emerged from the protective ocean until the ozone concentration had developed sufficiently to allow life to exist on land.

Destruction of ozone by catalytic chemical reactions has become the focus of worldwide interest in recent years, since the halogen constituents involved (mainly chlorine and bromine) are almost entirely anthropogenic in origin. Variations in solar UV radiation, however, also produce changes in ozone concentration, and the magnitude of these changes remains poorly understood. An additional complication is that changes in ozone concentration also cause changes in the stratospheric temperature distribution and consequent changes in the atmospheric transport of ozone. This feedback can also influence the dynamics of the upper troposphere, as has been shown by simulations using general circulation models.

In these circumstances it is vitally important that we understand the chemistry of ozone and the other minor constituents of the middle atmosphere in detail, and in particular the variability that is likely to occur. Changes in ozone production will take place primarily as a result of variability in solar UV radiation in the 170- to 242-nm wavelength range. The variation in this wavelength range over the 11-year solar-activity cycle has been estimated using data from spacecraft missions, particularly the Upper Atmosphere Research Satellite, but a knowledge of possible variation on longer time scales with high spectral resolution is required for a full understanding. To a large extent this will be accomplished by the Solar Radiation and Climate Experiment mission, currently scheduled for launch in 2003 and also intended as a component of the NPOESS series, a merger of DOD and NOAA monitoring capabilities.

Solar Irradiance Variability and Climate

Since the Sun is the ultimate driver of our climate system, it is reasonable to suspect that there might be a link between solar variability and the changes in climate that are known to have taken place in the distant and recent past and are probably continuing to take place at the present time. Despite many claims of correlations between solar-activity indicators and climate variables, however, the existence of such a link has been hard to prove. Part of the difficulty has been identifying a physical mechanism to couple solar variability to the lower atmosphere and the surface of the Earth. In the case of variability in the Sun's luminosity (total irradiance), the difficulty is showing that the variations are large enough to have a significant effect on the surface climate. Variability in the Sun's irradiance in the ultraviolet is substantial on a decadal time scale—however, the radiation is largely absorbed in the middle atmosphere and its influence on surface climate is not completely clear.

Interest in variations in the Sun's total irradiance (the solar "constant") was sparked by two developments: the need to identify the natural sources of climate variability so that the anthropogenic impact of greenhouse-gas emissions could be estimated, and the discovery that total irradiance does indeed vary on time scales at least as long as the 11-year solar-activity cycle. The climatic effects of variations on shorter time scales are probably negligible, while the variations over the 11-year time scale that have been identified seem too small to have a significant impact. There is an increasing amount of evidence from paleoclimate studies and isotopic measurements, however, that total irradiance probably varies at longer time scales and may have been an important factor in past climate changes. Since such variations can either weaken or amplify the anthropogenic effects, and thereby increase or reduce the time available to address these effects, it is of the utmost importance to estimate their magnitude.

Part of the variation in the Sun's irradiance takes place in the ultraviolet region of the spectrum, which is responsible for both the formation and the destruction of ozone in the stratosphere. The relative UV variations on the 11-year solar-activity time scale are an order of magnitude larger than the variations in total irradiance, and recent modeling work has shown that the former can have a significant effect on tropospheric dynamics and hence on climate. Presumably, total irradiance variations at longer time scales would be accompanied by UV variations of larger magnitude, which could amplify climatic effects. This possibility is relevant to the prediction of future climate change, and further studies are needed.

Solar irradiance measurements can only be carried out from spacecraft, since absorption and scattering by clouds and the atmosphere make ground-based measurements unreliable. Since long-term measurements are needed, the instruments will have to be replaced from time to time. The inevitable differences in absolute calibration of different instruments make it essential that successive instruments have a period of overlap.

Frontier Issues

While changes in solar irradiance provide a direct means of changing the surface climate, other less direct mechanisms have been suggested, and further observational and modeling studies are needed to test their feasibility. The basis for most of these mechanisms is solar-induced variability in cloud formation and cloud properties. Since clouds have a major influence on climate and the terrestrial radiation budget, they could provide a powerful means of amplifying the effects of solar variability, if a connection to solar variability can be shown to exist.

The flux of galactic cosmic rays striking Earth varies in intensity over 11-year intervals, in antiphase with the solar magnetic activity cycle, as a result of interactions between the cosmic ray particles and the interplanetary magnetic fields. It has frequently been suggested that the ionization produced by these particles could influence the nucleation of cloud particles, which would serve to couple solar activity to weather and climate. An alternative suggestion is that changes in the global electric field influence the freezing of supercooled water droplets and the release of latent heat. While observations have occasionally been taken as support for these mechanisms, the mechanisms remain controversial. If they are real, solar-induced changes in cloudiness are likely to produce regional rather than global changes, in contrast to the global effects of variations in irradiance. The societal impacts of an improved ability to predict regional climate changes are important, however, and further study is justified.

Space Climate

Recent work has pointed out the likelihood of an anthropogenic effect on the upper atmosphere and ionosphere, mainly as a result of the enhanced CO_2-induced cooling by radiation to space. The subject is still in its infancy, but the predicted changes could have a significant societal impact by changing the density of the mesosphere and thermosphere. This would in turn affect the decay of satellite orbits and, possibly, the global circulation in these upper regions. The changes could also affect the radio-propagation characteristics of the ionosphere. Since the anticipated changes are long term, "space climate" is an appropriate name for the range of effects. Observations of upper-atmosphere parameters have not yet shown a clear trend, and further observational and modeling efforts are clearly needed.

MAGNETOSPHERIC, IONOSPHERIC, AND ATMOSPHERIC PROCESSES IN OTHER PLANETARY SYSTEMS

The physical and chemical processes that control the atmospheres and plasma environments of the planets and their satellites are essentially the same throughout the solar system, but they are manifested in very different ways owing to differing heliocentric distances, planetary sizes, atmospheric compositions, and intrinsic magnetic fields. A comparative approach to understanding the processes controlling atmospheres, ionospheres, and magnetospheres throughout the solar system can be very effective. We have learned much about the neutral and plasma environments of the planets and satellites in our solar system over the past four decades, but our knowledge of these environments remains far from complete. Over the next decade a systematic effort will be needed to address the most basic deficiencies in this knowledge.

Structure and Dynamics of the Upper Atmospheres and Ionospheres of the Planets and Satellites

The structure and dynamics of the upper atmosphere (mesosphere, thermosphere, and exosphere) and ionosphere of a planet are strongly affected by inputs of energy and momentum from both the lower atmosphere, below it, and the magnetosphere and solar wind above. Our knowledge of planetary atmospheres has advanced considerably over the past three decades yet lags far behind our knowledge of the terrestrial system. The structure of the ionospheres of Venus and Earth is known rather well, but our knowledge of the other ionospheres in our solar system is limited. In situ measurements have been made in the upper atmospheres and ionospheres of Venus, Earth, Mars, and comets. Ionospheres have also been detected remotely with the radio occultation technique at Jupiter, Saturn, Uranus, and Neptune and at the satellites Io, Europa, Ganymede, and Titan. However, this technique only provides total electron density, and measurements have not been made of other important quantities such as ion composition and temperatures. At Mars, retarding potential analyzers on the Viking 1 and 2 landers each provided a single set of ion density profiles for the major ion species, and the Mars Global Surveyor measured the total neutral density, the magnetic field, and superthermal electron fluxes.

The lower atmospheres of most planets, and perhaps a few satellites such as Titan, can deliver significant inputs of energy and momentum into the upper atmosphere via the upward propagation and breaking of internal gravity waves. On Mars it appears that gravity wave production, and consequent upper atmosphere heating, is associated with dust storms that sometimes occur in the lower atmosphere near the surface. Neutral density measurements by the accelerometer onboard the Galileo probe provided evidence for gravity wave propagation into the upper atmosphere of Jupiter, but the source mechanism is not known. Energy inputs into upper atmospheres also come from the external plasma environment such as magnetospheres. Auroral emissions are one manifestation of such inputs. Missions are needed to study the structure and dynamics of planetary atmospheres and ionospheres and their coupling to the lower atmosphere and to the magnetosphere or solar wind. The Nozomi mission to Mars should provide some of this information, but it remains to be seen what shape the spacecraft and instruments will be in after the unanticipated 4-year flight delay before arrival at Mars. The ESA's Mars Express mission will also study Mars, but further missions will be required at both Venus and Mars to unravel the neutral dynamics and its role in A-I-M coupling. Such missions should include Venus and Mars climate orbiters (VCO and MCO) and a Mars Thermosphere Ionosphere Dynamics Orbiter (MTIDO). An appropriate mission to Jupiter, such as a Jupiter polar orbiter or a Jupiter auroral orbiter, should also be considered; it might emphasize the complex A-I-M issues at auroral latitudes. A mission to Jupiter's satellite Europa has been suggested; any spacecraft sent to Europa should include ion and neutral mass spectrometers in its instrument complement in order to characterize the composition of the atmosphere and ionosphere. Such measurements would also provide valuable information on the surface composition of this important satellite. Among ongoing missions, the Cassini mission should provide much information about the ionosphere of Titan and some information about Saturn's ionosphere.

Characterization of the Structure, Dynamics, and Composition of the Planetary Magnetospheres

Planets with sufficiently large intrinsic magnetic fields, such as Mercury, Earth, Jupiter, Saturn, Uranus, and Neptune, and Jupiter's satellite Ganymede, possess magnetospheres that act as obstacles to the solar wind. By contrast, the external plasma environment (e.g., solar wind) interacts rather directly with the upper atmospheres and ionospheres of bodies that have small intrinsic magnetic fields, such as Venus, Mars, Titan, and comets. In this case, "induced" magnetospheres form in which the external magnetic field is enhanced and drapes around the body.

The terrestrial magnetosphere has been the subject of extensive study over the past 40 years or so by dozens of Earth-orbiting satellites, including Dynamics Explorer, ISEE, Geotail, Polar, and IMAGE. The magnetospheres of Mercury, Jupiter, Saturn, Uranus, and Neptune have been explored by the Mariner, Pioneer, Voyager, and Galileo missions. The Ulysses and Cassini spacecraft also made measurements within the magnetosphere of Jupiter, and the Cassini mission will provide extensive data on Saturn's magnetosphere.

Our knowledge of Jupiter's magnetosphere has advanced greatly as a result of the Galileo mission and earlier missions, although many questions concerning its structure and dynamics remain. Recent global MHD models have helped to put the data returned from these various missions into context. Among the many outstanding questions concerning this magnetosphere are the following: What are the relative roles of planetary rotation and the solar wind in driving the dynamics? Where are the locations of reconnection region(s) on the magnetopause? What are the acceleration mechanism(s) for the observed energetic particles? What causes the energetic particle bursts coming from the magnetosphere but observed outside it? What are the identity and nature of the particle populations responsible for the Jovian aurora? What is the role of the ionosphere (and ionospheric electrical conductivity) in the A-I-M interaction? What are the mechanisms for the interaction of the magnetospheric plasma with the various satellites (e.g., the Io plasma torus)? One mission that would fill many of the gaps left from earlier missions, particularly in the inner magnetosphere and high-latitude regions, would be a Jupiter polar orbiter or a Jupiter Auroral Orbiter (JAO).

Planetary auroras are observable manifestations of atmosphere-ionosphere-magnetosphere coupling and occur when energetic charged particles from an external plasma environment interact with a neutral atmosphere. Very often, auroral particle precipitation is a sign of field-aligned electrical currents, which result from dynamical/MHD stresses somewhere in the A-I-M system. Jupiter provides a particularly interesting example of a planetary aurora. The total auroral power is a few times 10^{13} watts, and the auroral emissions are observed in the x-ray, ultraviolet, visible, infrared, and radio regions of the spectrum. For example, observations of auroral ovals by the Hubble Space Telescope (HST) reveal a complex spatial and temporal morphology (Figure 3.9). A JAO-type mission and regular remote monitoring at all wavelengths (e.g., by HST and other facilities) will be needed to understand this aurora.

Mercury has an intrinsic magnetic dipole moment but only a thin neutral exosphere, and it has a small magnetosphere that interacts with the solar wind. Almost all our information on this unique magnetosphere comes from instruments onboard the Mariner 10 spacecraft. It is time that we explore this magnetosphere more thoroughly and look for phenomena such as substorms. The proposed Messenger mission is a start in this direction.

Electrodynamical Coupling Processes at Weakly Magnetized Planets and Bodies

The solar wind is able to interact directly with the atmosphere and ionosphere of a planet or object possessing only a weak intrinsic magnetic field. Examples of weakly magnetized bodies include Venus, Mars, Pluto, comets, and satellites such as Io and Titan. For the satellites, the relevant external plasma flow is associated with the parent planet's magnetosphere. A thorough review of the solar wind interaction with weakly magnetized planets and bodies can be found in the report by the Panel on Solar Wind and Magnetosphere Interactions, but a concise review of this topic will also be given below.

The solar wind interaction with Venus was most recently studied from 1978 through 1992 by the Pioneer Venus mission. The instrument complement on the Pioneer Venus Orbiter was well designed, and much was learned about this interaction. However, significant gaps in our knowledge remain, such as the cause of the magnetic flux ropes observed in the ionosphere by the magnetometer and the cause of ion outflow on the nightside. No mission is currently planned for Venus.

Although many missions to Mars have been undertaken over the past three decades, we still do not have a comprehensive picture of the solar wind interaction with

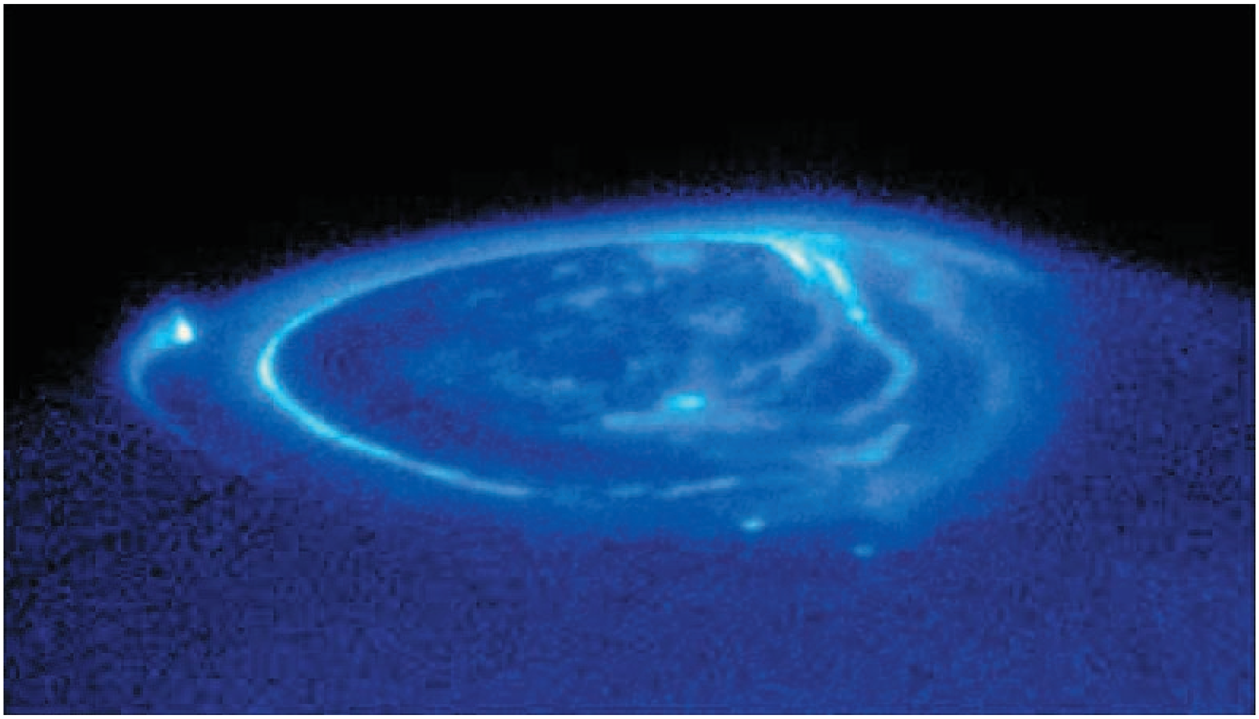


FIGURE 3.9 Jupiter's aurora as seen with the Hubble Space Telescope. Emission is evident from a main auroral oval, which is thought to be associated with field-aligned electrical currents originating in the outer part of the jovian magnetospere. Auroral emission is also present at lower latitudes at the magnetic footprints of the satellites Io, Europa, and Ganymede and is evidence for the electrodynamical coupling of these satellites with the jovian atmosphere. Courtesy of John Clarke, Boston University.

this planet. All the NASA missions lacked comprehensive instrument packages that could have provided coordinated information on the upper atmosphere, ionosphere, and plasma environment. Only recently have we even learned (from the magnetometer onboard the Mars Global Surveyor) that strong localized, crustal magnetic anomalies exist on Mars that apparently create mini-magnetospheres, affecting the interaction with the solar wind.

The solar wind interaction with comets is dominated by the mass-loading process, in which neutrals distributed over a large volume of space are ionized by solar radiation and contribute heavy cometary pickup ions to the solar wind flow. Halley's Comet was intensively studied by many spacecraft (and especially the European Giotto spacecraft, which approached to within 600 km of the nucleus). The U.S. International Cometary Explorer (ICE) spacecraft flew through the plasma tail of comet Giacobini-Zinner. Remote observations of comets continue to be an important source of information on these primitive objects. For example, the ROSAT and Chandra observatories have recently seen soft x rays coming from a large number of comets. It has been demonstrated that these x rays are produced by the charge transfer of solar wind heavy ions with cometary neutrals. Currently, the only approved cometary mission with instruments relevant to A-I-M is ESA's Rosetta mission.

The solar wind interaction with Pluto might be cometary in nature, but nothing is really known about it. Any mission to Pluto now being contemplated should include an instrument capable of measuring pickup ions in the vicinity of Pluto and learning their composition.

Other interesting objects with weak intrinsic magnetic fields and interesting A-I-M effects associated with external plasma interactions include the Galilean satellites—Io, Europa, Ganymede, and Callisto—and Saturn's largest moon, Titan. The jovian satellites were studied by the Galileo mission, as well as by the earlier Voyager mission and by remote observations, but much remains to be learned. As noted above, any spacecraft sent to Europa should be equipped with ion and neutral mass spectrometers to characterize the atmosphere and ionosphere.

Saturn's magnetospheric plasma strongly interacts with Titan's upper atmosphere and ionosphere, perhaps helping to drive the complex hydrocarbon chemistry in the lower ionosphere that contributes to haze formation. Most of the important A-I-M issues at Titan will be addressed by the Cassini mission starting in 2004.

3.3 SOCIETAL IMPACT OF SPACE WEATHER

The maturity of the atmospheric-ionospheric-magnetospheric discipline has made possible a close connection between A-I-M science and applications for the benefit of society. The application of space physics and aeronomy to societal needs is now referred to as space weather. While relevant activities have been pursued by NOAA and DOD for many years, the science community has only recently become more closely connected with this pursuit, since the inception of the National Space Weather Program in 1995 and NASA's Living With a Star program.

The NSWP is a multiagency endeavor to understand the physical processes, from the Sun to Earth, that result in space weather and to transition advances in science into operations in order to assist users affected by the space environment. This transitioning of knowledge has been an obstacle for many years and is still a major challenge. The scientific community should be encouraged to bridge the gap between the science and user needs by developing operational products. It is important that resources be available to help scientists prepare their models and data for transitioning to operations, pursue rigorous model validation, ensure robustness of computer codes, and provide the documentation essential for efficient and appropriate use of their models and data. Similarly, resources must be available to the commercial and defense agencies, such as NOAA and DOD, so they can transition advances in modeling and data and put to use the new understanding of the physics of space and the upper atmosphere. In the area of policy research, it is important that cost-benefit analyses of the societal impacts of space weather be conducted, along with annual updates to keep pace with rapidly changing technological systems.

NASA's new LWS program is an excellent opportunity to provide measurements and develop models that will clarify the relation between sources of space weather and the ultimate impact of space weather on society. Those space weather phenomena that most directly affect life and society include radiation exposure both in space and at commercial airline altitudes; communications and navigation errors and outages; changes in the upper atmosphere that affect satellite drag and orbital decay; radiation effects on satellite electronics and solar panels; and power outages on the ground due to geomagnetically induced currents, to name a few.

The LWS science architecture team has proposed an end-to-end approach to understanding how solar variability affects life on our planet. In the A-I-M area, a geospace missions network is being defined. The network will include geosynchronous-transfer-orbit satellites to study the radiation environment across L-shells, essential to understanding the rapid buildup of outer zone electron fluxes and the access and trapping of solar energetic protons in the inner magnetosphere. The network will also include low-Earth orbiting spacecraft with higher inclination, designed to study both thermosphere-ionosphere effects and precipitation of energetic particles from the radiation belts, and, finally, a high-inclination elliptical remote-sensing spacecraft to provide global imaging relevant to current systems in the high-latitude ionosphere that affect power outages through GICs. LWS will also contribute to missions-of-opportunity flights, leveraged programs, instrument development, and the development of end-to-end models that emphasize boundaries and linkages, above and beyond the scope of existing SR&T grants. The Panel on Atmosphere-Ionosphere-Magnetosphere Interactions endorses these goals in its recommendations, which are set forth in Chapter 6, and stresses the need to develop solar and geospace missions in parallel.

The goals of space weather modeling and observation are multifaceted. The first is the production of a suite of accurate climatological models describing the mean conditions, together with an estimate of the range of states as a function of conditions such as season, and solar and geomagnetic activity. Such models would be a valuable tool for spacecraft design—for instance, to estimate the expected environmental extremes of radiation belt particles on a polar-orbiting spacecraft. A second goal is to capture the "weather" of the system—that is, the minute-to-minute or hour-to-hour changes in the space environment. Space weather can be divided further into specification and forecast, the latter being an extreme test of our physical understanding of the complete solar-terrestrial system. It is imperative that we embrace all available techniques to achieve these goals. For instance, we must adopt the advanced data assimilation techniques that have been the mainstay of conventional weather forecasting for many years. We must endeavor to develop comprehensive satellite and ground-based observing systems to provide measurements to drive the data assimilation models. It is important that research and operations work together to achieve maximum benefit from new multispacecraft observing systems. For instance, every effort must be made to disseminate measurements from science missions for operational use in real time. An excellent example of this has been the availability of solar wind data from NASA's Advanced Composition Explorer (ACE) spacecraft. We should strive to make this the norm, thereby maximizing the use of scarce community resources. It is also important that a system be developed to allow data from existing operational satellites, such as DMSP and NOAA-POES, to be downloaded in real time rather than once per orbit. The 2-hour delay in orbit-by-orbit satellite communication reduces the value of the data for operational use.

The maturity of our understanding of the physics of the upper atmosphere enables us to move beyond qualitative science to a true quantitative description of the environment. Such an advance involves developing appropriate metrics to measure the ability of models or theories to make predictions. When we develop or improve empirical models such as the International Reference Ionosphere, it is no longer sufficient to specify the parameters themselves—instead, we must strive to characterize the variability of the system under different conditions.

The societal impacts of space weather are broad, affecting communications, navigation, human health, power distribution, and satellite operations (Figure 3.10). Space weather is of international concern, and other nations are pursuing parallel activities that can be leveraged through collaborations. The European Space Agency and individual nations within ESA have programs for studying space weather. For example, the ESA's 2001 report on space weather[1] articulated possible elements of a future space segment for an operational, service-oriented, European space weather system. Likewise, Japan has a solar cycle's worth of energetic particle data from the polar-orbiting Akebono satellite to contribute to radiation belt studies, and comparisons between Yohkoh x-ray images of the Sun during the last solar cycle and images that will be obtained from the new x-ray imager to be flown on the GOES spacecraft will provide an important cycle-to-cycle and long-term baseline on x-ray input to the atmosphere. A brief review of the principal space weather effects follows.

COMMUNICATIONS

Communications can be separated into near-ground and space-based systems; both are affected by the space

---

[1] ESA. 2001. *Space Weather: Concurrent Design Facility Study Report,* CDF 11(A), December. Available online at <http://www.estec.esa.nl/wmwww/wma/spweather/esainitiatives/>.

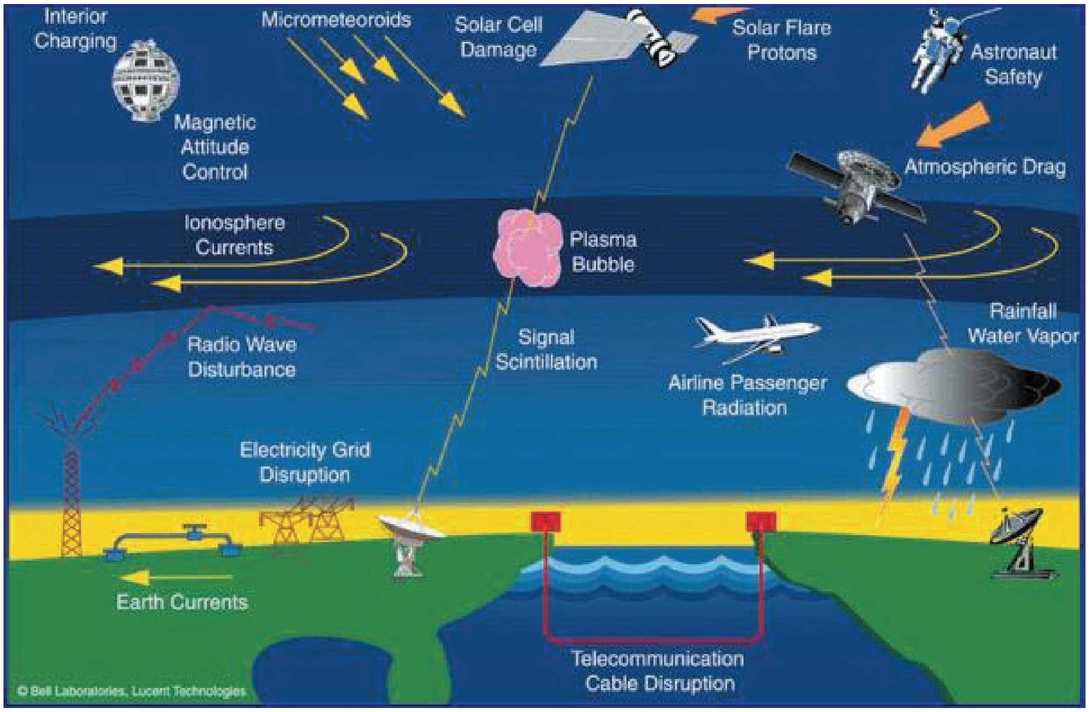


FIGURE 3.10 Space weather hazards, including space, air, and ground effects. Space weather describes events in space that affect Earth and our technology. Severe solar eruptions can cause disturbances that dramatically affect both the ionosphere and magnetosphere. Courtesy of L. Lanzerotti.

environment but in quite different ways. Near-ground communication systems (ground-to-ground and aircraft-to-ground) utilize high frequencies (3–30 MHz) and use the ionosphere actively as a reflector to propagate a radio signal from transmitter to receiver. Satellite communications transmit at higher frequencies (VHF, UHF, L-band, etc.), where it is irregularities in the ionosphere that are most likely to disrupt the fidelity of the radio signal.

For high frequencies, two main space environment factors affect the propagation: absorption of the signal in the D-region and changes in the reflecting properties of the ionosphere in the F-region (near 300 km altitude). Absorption is the process by which the energy of radio waves is converted into heat and electromagnetic noise through interactions between the radio wave, ionospheric electrons, and the neutral atmosphere. Most of the absorption occurs in the ionospheric D-region (50–90 km altitude), where the product of the electron density and the electron/neutral collision frequency attains a maximum. Within this region the neutral density is relatively constant over time, so variations in the local electron density drive the total amount of absorption. The electron density is a function of many parameters and normally varies with local time, latitude, season, and stage of the solar cycle. These natural changes are predictable and affect absorption only moderately at the lowest HF frequencies. Much more significant changes in the electron density, and therefore the absorption strength, are seen in response to solar x-ray flares (the classic short-wave fade) and solar proton events (the classic polar cap absorption), both of which change the lowest usable frequency. Routine monitoring of the solar x-ray flux by the GOES series of operational spacecraft provides a reasonable specification of the current condition of absorption on the dayside of Earth. Forecasting an x-ray flare a day or two in advance would be an ideal situation for HF operators, but even a short-term forecast of a few hours presents a huge challenge for solar physics research. Likewise, specifying the spatial extent and intensity of polar cap absorption during a solar proton event is possible with existing polar orbiting and geostationary monitors, but a reliable forecast poses a similar challenge given our current understanding of solar and space physics. With the ever-present need for commercial airlines to maintain good communication links, particularly with the increasing number of polar routes, there has been a renewed emphasis on accurate predictions of HF propagation conditions.

The reflecting properties of the F-region ionosphere are influenced by another space weather event: the geomagnetic storm. During a storm the high-latitude ionosphere is impacted directly, but the energy injected into the upper atmosphere drives winds and composition changes in the neutral atmosphere that produce both increases and decreases in F-region electron density across the whole globe. These changes in the ionosphere affect the maximum usable frequency. At present, HF users rely on climatological models of the ionosphere to choose which frequency to use along a particular propagation path and have little or no forecast capability. Day-to-day variability is difficult to capture, and the more severe disruptions during a storm are not yet adequately quantifiable. Advances in combining models and data using data assimilation techniques are in progress and hold the best hope for improved specification of the F-region conditions; they also may allow at least a short-term forecast (6–12 hours). The main driver of this activity is the promise of new global satellite observing systems such as the Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC), or the Geospace Missions Network, which is a component of NASA's LWS program. Growing networks of ground-based Global Positioning System receivers complement the satellite observations and are essential to providing global coverage for specification and initialization for short-term forecasts that make use of physical models.

Many military and civilian systems depend on HF and UHF communications and will continue to do so for the foreseeable future. Localized conflicts in future military confrontations very likely will require secure communications for locating downed pilots. Complex operations may be delayed if ionospheric conditions are threatening. The airline industry uses the HF band extensively for communications. Communications systems at higher frequencies that are less susceptible to space weather effects are also used by airlines, but these additional capabilities are not available at all locations, particularly in the northern polar region. Near the poles, the ionosphere becomes strongly modified during large solar proton events. At the highest latitudes, these energetic protons arriving from the Sun have direct access to the ionosphere, where their energy is deposited. The HF communications systems are unable to operate well at these times, so airlines must be rerouted to lower latitudes, at substantial cost to the airlines and ultimately to the consumers. The use of imaging holds great promise for tracking space weather disturbances, much as weather systems on Earth are now tracked from geostationary orbit.

The Drug Enforcement Agency uses over-the-horizon (OTH) radars in its battle against smuggling. These systems were important in cold war defense against cruise missiles and continue to play a huge role in the surveillance of regions outside the United States. They are now used to detect airplanes and ships from the velocity information encoded in the HF signals as they bounce off the ionosphere. Since the system uses HF waves that refract from the bottom of the ionosphere, a range of space weather effects can disrupt it. Allies such as Australia use OTH systems for wide-area surveillance in Asia and are keenly aware of space weather.

Ground-to-satellite communication signals are only weakly affected by the ambient ionospheric densities. However, there are times, at low latitudes and in the auroral oval and polar cap, when plasma irregularities are predominant. Irregularities diffract even the very-high-frequency communication signals, leading to scintillation in both the amplitude and phase of the waves. Severe scintillation can result in complete loss of the radio transmission. Forecasting the occurrence of ionospheric irregularities and scintillations is a major challenge, but physical understanding of the system is advancing rapidly. Renewed efforts in this area hold distinct promise for the future. The Air Force, with NASA's help, has embarked on an ambitious satellite program called the Communications/Navigation Outage Forecast System to predict these events and the scintillation they cause. The Navy also has a sizable program in ionospheric characterization and prediction.

NAVIGATION

High-frequency satellite-based navigation systems, of which the GPS is an example, rely on propagating signals through the ionosphere from ground to satellite. Ionospheric irregularities are therefore just as important for these navigation applications as they are for satellite communication. Precise positioning can also be affected by the total electron content of the ionosphere, due to changes in the refractive index of the medium. As the navigation signals pass through the plasma the signal is retarded by the medium, so the arrival time of the signal, and hence the calculation of position, is compromised. If the ionosphere is invariant, this correction can be easily defined, but the highly dynamic nature of the system introduces an uncertainty for single-frequency GPS users. Commercial airlines, for instance, that rely on single-frequency GPS navigation systems require accurate specification of the current conditions. The Federal Aviation Administration attempts to make corrections to TEC climatology by using the Wide Area Augmentation System (WAAS) over the continental United States. Of particular concern are local gradients in electron content, which are difficult to capture in the coarse observing grid of the WAAS system. Specifying and forecasting the state of the TEC has the potential to greatly advance the adoption of new data assimilation techniques and observing programs. Ionospheric models and the science on which they are based must continue to improve if a WAAS-type system is to be able to capture these steep gradients in electron content.

The GPS was invented for military purposes but it now has widespread application in both military and civilian uses. The ability to pinpoint one's position is of great value, and any outage or degradation due to space weather can be a threat to life and military advantage. The combination of GPS and cell phone technology and links to the 911 system hold great promise for safety and emergency situations.

Low-frequency ground-based navigation systems, of which loran is an example, are normally unaffected by the ionosphere. However, during solar x-ray and proton events, the reflecting height of the radio wave is lowered, introducing an ambiguity between the directly propagated signal and the signal reflected from the ionosphere (known as the sky wave). The utility of the system is compromised, requiring the system to be turned off for the duration of the event.

ELECTRIC POWER ISSUES

On March 13 and 14, 1989, one of the largest geomagnetic storms on record plunged Quebec Province, Canada, and its 6 million inhabitants into darkness for over 9 hours and cost the utilities millions of dollars. In this case, as well as others like it, geomagnetically induced currents resulting from disturbances in the space environment have been the cause of power outages and damage to electric power generation and distribution systems. More recently, the smaller but still significant geomagnetic storm on July 15 and 16, 2000, caused voltage variations and tripping of protective devices at many locations in the United States.

During large geomagnetic storms, rapidly time-varying electric fields and currents that are often co-located with the auroral oval intensify and expand to lower latitudes and broaden in area to expose large portions of Earth's surface to this destructive source of energy. Given enough warning, power companies can take protective measures to reduce effects on their systems by better management of connections and power transfer among grids, by deferring maintenance, by using a more conservative load margin, and by adjusting protective devices. During the current solar cycle, the power industry may have become more vulnerable to space environment disturbances because of regulatory changes that have resulted in increased utility connectivity and smaller operating margins. At the same time, new research, observations, and models are beginning to contribute to improved warnings and alerts of geomagnetic storms provided by the NOAA Space Environment Center, and private industry products are also being offered for use by the power utilities.

Primary among the improvements are observations from the ESA/NASA SOHO spacecraft that provides solar observations of so-called halo events, or Earth-directed coronal mass ejections, that often result in geomagnetic storms in a few days. However, we do not yet have the scientific understanding necessary to interpret these observations to skillfully predict the magnitude and timing of a disturbance reaching Earth, especially with the long lead times of several days that are desired by the power utilities and others affected by solar disturbances. Another important observation is from the NASA Advanced Composition Explorer (ACE) spacecraft that is located at the L1 Lagrangian point between Earth and the Sun, about 1 percent of the way to the Sun. NASA and NOAA, along with collaborators from many nations worldwide that are tracking ACE, work together to make these data available in real time for the detection of geoeffective solar wind conditions and input to magnetospheric models that forecast geomagnetic activity. These data are critical for providing short-lead-time warnings, about one-half to one hour, of impending geomagnetic storm conditions. After completion of the ACE mission, it is imperative to continue these measurements with new solar-wind-monitoring measurements for both research and space weather operations.

There are challenging opportunities to improve the magnetosphere-ionosphere-upper atmosphere models that use solar and solar wind data as input to determine the location and severity of geomagnetic disturbances. Coupled models of the entire Sun to upper atmosphere system are needed to test our understanding of solar-terrestrial processes and to improve our ability to forecast geoeffective disturbances. Finally, globally distributed ground-based magnetometer measurements are needed to validate model predictions and to provide nowcasts of geomagnetic disturbances with improved spatial resolution.

ASTRONAUT, AIRLINE, AND SATELLITE HAZARDS

The impacts of space weather on society are also increasing rapidly as we maintain a permanent presence in space, as our economy and lifestyle depend more and more on satellite technology, and as high-latitude airline flights become more frequent. Astronauts are subject to considerable radiation even in low-Earth orbit. When the International Space Station (ISS) changed its orbit to accommodate Russian launches, it moved to higher latitudes, where it is more susceptible to solar energetic protons, and it entered the dangerous portion of the radiation belt. As detailed in the 2000 NRC report *Radiation and the International Space Station*, astronauts could receive a significant radiation dose (reaching allowable limits) during ISS construction due to large solar proton events that may occur during space walks. This study emphasized that the radiation risk to astronauts could be greatly reduced by employing data and models to adjust scheduled space walks. Missions to Mars will be very dangerous indeed, since solar flares give very little time to take safety precautions and the radiation levels could be life threatening. Space walks for construction and maintenance of the ISS can also be somewhat dangerous during geomagnetic storms as the radiation belt fluxes increase. Even polar airplane flights at times have high radiation levels. Although Earth's atmosphere still provides substantial protection from solar radiation, at high latitudes during large solar proton events, the radiation dose to airline passengers and crews becomes enhanced.

Similarly, our satellites in space are vulnerable to a wide variety of environmental effects. Low- to medium-energy electrons create an electrical charging layer on the surface of spacecraft that can discharge and damage satellite systems. High-energy electrons can penetrate deep into the satellite's components, including onboard computers, and create electrical discharges that can cause problems ranging from minor anomalies to total failure of the spacecraft. Similarly, solar energetic protons and heavy nuclei can penetrate deep into a spacecraft's components and disrupt satellite operation. In addition, some spacecraft rely on the relatively steady magnetic field within the magnetosphere to maintain their orientation and control their momentum. During times when the solar wind is enhanced, the protective shield of Earth's magnetic field can be compressed to within some satellite's orbits, leaving the spacecraft outside the protective magnetosphere and unable to maintain normal control.

SATELLITE DRAG AND COLLISION AVOIDANCE

Satellites in low Earth orbit are significantly affected by atmospheric drag. During a magnetic storm the atmosphere is heated and expands. The dynamic nature of the energy input gives rise to the launch of neutral density waves, and the momentum forcing at high latitudes can lead to deep neutral density holes. In the past, after intense dynamic events, NORAD lost track of many space objects and had to laboriously reacquire them over the following days and weeks. Our prediction of such events could greatly assist in keeping track of all the space debris in orbit.

Day-to-day neutral density perturbations contribute the largest error in the position determination of space objects and debris. With accurate knowledge of neutral density the uncertainty windows on predicted position could be significantly reduced and so limit the need for ISS to maneuver to avoid collisions.

3.4 EXISTING PROGRAMS AND NEW INITIATIVES

Our understanding of the A-I-M system is being advanced through a number of vigorous programs, ranging from international and multiagency programs to smaller scale programs. Current examples are NASA's International Solar-Terrestrial Physics (ISTP) program, the TIMED Solar Terrestrial Probe, the IMAGE MIDEX mission and the SAMPEX and FAST Small Explorer (SMEX) satellites; the National Space Weather Program; the NSF-initiated SHINE, GEM, and CEDAR programs and their international counterparts; NOAA's space weather programs; the International Space Environment Services program; and numerous DOD activities. These programs have supported satellite and ground-based monitoring instruments and the related data analysis and theory and modeling efforts. Through these targeted programs and the critically important base programs funded by NSF, NASA, NOAA, and the DOD, much important scientific progress has been made, helping us to focus our research effort on high-priority issues. The progress that has been made and the important new questions that have been addressed are serving to excite and to motivate the next generation of young scientists who will lead our future explorations of the solar-terrestrial environment. Our progress has also led to direct application of research models and data for operational space weather prediction and new efforts to model the Sun-Earth system from a systemic and holistic perspective. While progress is being made, however, there are still very few areas where our space weather modeling capabilities have sufficient accuracy to influence economic decisions being made by private industry and the government. Future programs and interagency activities will continue to refine our understanding of the A-I-M system and should lead to an improved and quantitative understanding of the complex coupling between the many different physical regimes extending from the Sun to Earth.

Future progress in understanding the response of the magnetosphere, ionosphere, and upper atmosphere to the energy inputs from the Sun and solar wind as well as to forcing from the lower atmosphere requires a new approach that goes beyond the ISTP program's method of tracing the flow of energy through geospace with single-point observations. Future progress requires a systems approach that views the A-I-M as a tightly coupled system with microscale drivers and feedbacks that modify the first-order global response. We need a more detailed view of the dynamical behavior of boundary regions in geospace such as the dayside magnetopause (to view the entry of solar wind energy and momentum), the nightside inner plasma sheet (to view the current disruption/acceleration boundary region), the midtail region (to view the nightside, near-Earth, magnetic merging site that releases stored magnetotail energy), and key coupling regions between the ionosphere and magnetosphere (to view the closure of currents through the low-altitude ionosphere). In these regions, microscale processes are known to control the global system behavior. Clusters of satellites flying in close formation can resolve the dynamical response of the regions (by measuring gradients, curls, and divergences of important quantities) as well as distinguish spatial from temporal variations. In effect, they give us a space- and time-resolved view of the dynamical behavior of key regions rather than of the single-point, basic-state quantities measured in past missions.

To maximize the information gained from these missions, creative coordination of approved programs across agencies (e.g., NSF, NASA, NOAA, ONR, and DOD) is needed to place the system's microscale drivers in the context of the solar wind energy input that they mediate and the global system response that they engender. Such a cluster in equatorial orbit would give us a global knowledge of the zonal electric field at the magnetic equator—the single most important parameter needed for predicting space weather disturbances there. New, movable, state-of-the-art ground-based radars planned by NSF in conjunction with existing sophisticated arrays provide a powerful network for sensing the global magnetospheric configuration as mapped onto Earth's upper atmosphere. These radio-wave systems must be augmented by relocatable optical systems as well.

New instrumentation on the NOAA geosynchronous satellites allows continuous monitoring of the solar disk in x rays to view oncoming explosive events. The launch in 2004 of STEREO, part of NASA's Living With a Star program, will allow stereo viewing of explosive events leaving the solar surface. Other elements of LWS include a multisatellite radiation belt mission and a mission to study magnetosphere-ionosphere coupling. The deployment of an operational geosynchronous global ionospheric imager by ONR is under discussion. DOD-operated Los Alamos National Laboratory geosynchronous satellites provide important multipoint monitoring of plasmas entering the inner magnetosphere from the magnetotail, the source populations for the ring current and radiation belts. New instrumentation on the DOD Defense Meteorological Satellite Program satellites will provide snapshots of the auroral oval that can be inverted to provide energy input and conductivity estimates. We also need assured access to large-aperture telescopes capable of collecting laser light scattered back to Earth from metallic atom layers. These lidar systems are the only reliable tools for ground-based study of the 80–120 km component of the atmosphere; they can be augmented by rockets and ground-based imagers to provide the data we need to understand this important interface zone between the atmosphere and space.

A systems view of the solar wind-A-I-M interaction motivates a search for new ways of providing snapshots of global quantities rather than single-point measurements in key regions of the A-I-M system. The IMAGE mission is the first to use new technologies to provide global snapshots of the changing spatial configuration of the ring current and plasmasphere. Imaging of auroral emissions, which are the signatures of electrodynamic coupling between a planet's magnetosphere and ionosphere, remains the only realistic means by which to examine the global dynamics of such extended systems and is an excellent tool for placing other observations in the global context. A vigorous development program for new ground- and space-based measurement techniques and instrumentation must be part of the ongoing effort. These include both miniaturization of instrumentation for small satellite constellations and new remote-sensing instruments that provide snapshots of large-scale regions from space or ground observatories.

The systems view requires an enhancement of efforts aimed at developing global theoretical models of the Sun-Earth system, including the simultaneous development of new software technologies for efficient use of parallel computing environments. Of particular importance is the development of adaptive grid technologies for models that address the large range of spatial and temporal scales characteristic of the global system structure and response. However, the A-I-M system is not simply multiscale, but it also requires inclusion of additional physical processes in boundary layers, transitioning between weak and strong magnetic field regions and coupling ionized and neutral populations across substantial gradients in density, flow velocity, energy, pitch angle, and species—for example, along magnetic field lines. Single-particle drift motion, which is species-dependent, is required to resolve features of the ring-current buildup and recovery in a geomagnetic storm not contained in single-fluid MHD codes. Likewise, parallel electric fields are an essential feature of the codes that capture auroral-arc-scale transverse structure, an important part of magnetosphere-ionosphere coupling at high latitudes, which mediates the large-scale conductivity of the ionosphere, in turn affecting global MHD structure.

A final example of where additional physical processes must be included in boundary regions is the decoupling of electron and ion motion, so-called Hall currents, which must be taken into account in any code that accurately computes the reconnection rate near an x-point of magnetic field line merging. Yet the global flows into and out of the merging region are well described by single-fluid MHD codes, as has been demonstrated by extensive data comparison. Hall MHD, multifluid, and particle codes that incorporate gradient-curvature drifts and parallel electric fields are examples of computational approaches that include the subgrid physics needed to accurately describe the non-Maxwellian features of the A-I-M system. GEM's Geospace General Circulation Model efforts and the multiagency-sponsored Coordinated Community Modeling Center rapid prototyping activity are examples of important ongoing programs aimed at maximizing the usefulness of and access to global modeling efforts by the scientific community.

NSF's highly successful SHINE, GEM, and CEDAR programs and the recent coordination of these groups into Sun-to-Earth analysis campaigns highlight the need to focus this broad range of expertise on issues involved in coupling between and among the Sun, the solar wind, the magnetosphere, and the ionosphere/atmosphere. To facilitate collaboration between scientific communities that rarely come together and among scientists, space-weather forecasters, and the user communities, new information technologies are needed that provide seamless virtual interactions. The ongoing Scientific Committee on Solar Terrestrial Physics (SCOSTEP) Solar Terrestrial Energy Program—Results, Applications, and Modeling Phase (S-RAMP) space weather campaign, which seeks to draw together all of these communities in analyzing an international set of observations, is a large-scale, ongoing experiment in the uses of virtual collaboration technologies. NSF's information technology initiatives should be utilized as much as possible to develop important collaboration technologies in support of such major community analysis efforts.

3.5 TECHNOLOGIES FOR THE FUTURE

DATA ASSIMILATION

The many civilian and military users of transionospheric and subionospheric communication channels and navigational systems are in need of physics-based ionosphere models that are driven by data assimilation. By the end of the decade we may have access to the requisite quantity of data, but we must aggressively test and validate these tools as they become more sophisticated. A goal of the National Space Weather Program is to complete and test data assimilative models similar in design to those of the meteorological community. Successful prediction of ionospheric disturbances in the equatorial zone seems well within our capability in this decade. At middle latitudes, long thought to be quiescent, recent results show that fundamental physics questions remain unanswered. At high latitudes, many challenges remain.

With large and distributed data sets being collected by multiple agencies through various techniques and covering diverse regions in geospace, new information technologies are required to improve data access from a single coordinating site, to allow remote searching of these distributed data sets, and to simplify data assimilation into global models in real time and for postevent analysis.

Improvements in meteorological weather forecasting have demonstrated the utility of adopting sophisticated data assimilation techniques. It is essential that the space physics and aeronomy communities learn, further develop, and implement these methods for both operational and scientific use. The potential benefit of new observing systems can come to fruition only if maximum use is made of the data; this, in turn, will happen only if a comprehensive data assimilation program is developed. This is not a trivial task, and the effort involved should not be underestimated. Assimilation models can also provide the tools to visualize a wide network of data and provide guidance on when and where to target observations to ensure efficient use of resources.

Data assimilation is the optimal combination of data with the physical understanding embedded in physical models. It is distinct from data synthesis, where a small number of observations are used to adjust a model output. There are numerous data assimilation methods available in meteorology and oceanography that can be applied to the space environment.

SPACECRAFT AND INSTRUMENT TECHNOLOGY

There are several areas of technological development that are needed to implement A-I-M objectives on future missions efficiently. The needed technology splits into two areas: spacecraft subsystems and instrumentation. Both areas require active funding by NASA to foster improvement. It should be emphasized that the best return on NASA's investment will come from a peer-reviewed competition open to universities and industry as well NASA centers. In the case of spacecraft subsystems, the return will be immediate and far-reaching. Improvements in telemetry, command and data handling (CDH), attitude control, and power systems can be directly transferred to private industry, where they will make American spacecraft more competitive in a global economy. Instrumentation development has a less direct impact, but the innovations that come from better scientific instruments often lead the way for incorporating new technology into spacecraft subsystems. It is important to realize that developments in these areas must support traditional, highly instrumented spacecraft as well as smaller, more simply instrumented spacecraft (micro- and nanosatellites).

Future NASA space science missions will increasingly rely on a multispacecraft approach, as is amply discussed in this report and those of the other panels. For such missions to be achieved at reasonable cost, spacecraft subsystems must become more efficient in their use of mass and power. For every scientific spacecraft built in recent memory, the majority of mass and power go to spacecraft subsystems, not instrumentation. To improve the performance of these systems, NASA will need to foster research and innovation. Development of highly power-efficient CDH systems with good flexibility and increased-capacity mass memories with small impact on spacecraft resources will be critical to these missions. NASA is currently working on high-efficiency thruster systems. Multispacecraft missions can require frequent station keeping to maintain optimum spacecraft positioning; more efficient thrusters can have a direct impact on spacecraft size and mission longevity. This work should be continued and expanded.

Both power generation and power storage improvements are needed. Although solar cells have become more efficient, development of still higher efficiency solar cells should be encouraged. For planetary missions to the outer solar system, radioisotope thermal generators (RTGs) are needed. The development and use of RTGs is a politically sensitive issue, but if NASA is to continue exploration of Jupiter and beyond, these power sources must be developed so that the public is satisfied that they are safe. Without safe, politically acceptable RTGs, exploration of the solar system and beyond will be significantly limited. Battery development should also be encouraged. By decreasing the mass requirements for a given amount of stored power, smaller, more efficient spacecraft can be constructed.

Multispacecraft missions will also place significant new demands on telemetry reception capacity. Much of this reception capacity is currently contained within the Deep Space Network (DSN). However, the DSN, as presently configured, consists of relatively large, expensive receiving antennas. Recent experiences with Cluster and other missions suggest that DSN is stretched very thin. The next generation of multispacecraft missions will not fly at extremely large distances from Earth; most are envisioned to be within its magnetosphere. This suggests that NASA should consider augmenting the DSN with arrays of smaller, less-expensive antennas that can handle missions that stay within 20–30 R_E of Earth. If each of the main DSN stations (Canberra, Goldstone, Madrid) had two, three, or four smaller dishes, then the upcoming multispacecraft missions could have their telemetry reception offloaded onto these smaller dishes, freeing the larger ones for planetary or other low-signal-level missions. As part of making such a system efficient, automation of receivers must be considered to reduce operating costs. NASA should develop options for managing data reception to find solutions that will give efficient and sufficient capacity.

On the instrumentation side of new technology, the key is to provide funding to scientists to develop the instrumentation. This support must be for the development of new basic technologies and materials as well as specific instrumentation designs. As an example, basic research in magnetoresistive materials may lead to new, highly efficient magnetometers. However, we also need novel designs for electrostatic detector optics to improve the efficiency of detection of low- and mid-energy ion mass analyzers. In the current environment, instrumentation innovation does occur, but it tends to be sporadic, and nonuniversity groups that have internal overhead return tend to be favored, because they can internally finance modest development efforts. By providing steady, regular funding for instrument development that is openly competed for and peer-reviewed, along with support for suborbital and UNEX-class flight opportunities, NASA will allow researchers to count on continuing funding for their efforts to come up with new detector designs that will measure more accurately and more efficiently.

An important part of the development of new, compact systems for both instrumentation and spacecraft is the microelectronics that is at the heart of these systems. Central to microelectronics is the development of modern parts with good radiation tolerance. In particular, NASA must foster the development of rad-hard microprocessors, programmable gate arrays, and other digital and analog electronic components so that the United States can remain competitive with the rest of the world.

3.6 RECOMMENDATIONS

Future research in atmosphere-ionosphere-magnetosphere science must prominently support projects, theories, and models that address the three-dimensional, dynamic behavior of the coupled A-I-M system. Crucial to understanding dynamic, complex geophysical phenomena such as magnetic reconnection, auroral processes, and electrodynamic ionosphere-thermosphere coupling are measurements from multiple platforms (e.g., the recently launched four-satellite Cluster II mission and the planned Magnetospheric Multiscale mission). Also critical to achieving such understanding are advances in the area of numerical simulation, including the development of mature coupled ionosphere models and the incorporation in global models of proper physical representations of sub-grid-scale effects.

Future measurements and models must pay even greater attention to these essential aspects of near-Earth space. The overarching goals are these:

1. To understand how Earth's atmosphere couples to its ionosphere and its magnetosphere and to the atmosphere of the Sun and
2. To attain a predictive capability for those processes in the A-I-M system that affect human ability to live on the surface of Earth as well as in space.

We currently have a tantalizing glimpse of the physical processes controlling the behavior of some of the individual elements in geospace. We must now address cross-cutting science issues, which include

- **the instantaneous global system response of the A-I-M system to the dynamic forcing of the solar atmosphere.**

For example, how does the magnetosphere limit solar wind power input, manifest in saturation of the polar cap potential? How do the neutral atmosphere and the ionosphere respond to sudden and long-term changes on the Sun? In view of the multiple temporal and spatial scales we must understand

- **the role of micro- and mesoscale processes in controlling the global-scale A-I-M system.**

The exchange of mass, momentum, and energy between the geophysical domains (e.g., connection of solar wind plasma at the magnetopause, ionospheric outflow, upward propagation of electromagnetic and mechanical energy from the lower atmosphere) is a key element in the coupled A-I-M system. It is now imperative that we understand

- **the degree to which the dynamic coupling between the geophysical regions controls and impacts the active state of the A-I-M system.**

The Sun is now recognized as one of the important factors in global change. Accordingly, we must resolve

- **the physical processes that may be responsible for the solar forcing of climate change.**

These critical science issues thread the artificial boundaries between the disciplines, but within each discipline, important science questions remain. For example, Earth's outer magnetosphere, acting as a powerful particle accelerator, is often populated by a surprising degree of relativistic electrons that pose a radiation hazard to space-based systems. It is important that we determine

- **the origin of the multi-MeV electrons in the outer magnetosphere and the cause of the pronounced fluctuations in their intensity.**

In the thermosphere and ionosphere, one of the fundamental science issues that must be resolved is to determine

- **the balance between internal and external forcing in the generation of plasma turbulence at low latitudes.**

To accomplish our goals we note that simultaneous, multiplatform remote-sensing observations of the A-I-M system—as well as in situ measurements—are urgently needed in order to specify the many interconnected dynamic, thermodynamic, and composition variables. As our understanding of the complexity of the thermosphere, ionosphere, and magnetosphere grows, so does the requirement to capture observations of these multiple facets of the coupled media.

In the next decade, NASA should give highest priority to multispacecraft missions such as Magnetospheric Multiscale (MMS) (Box 3.1), Geospace Electrodynamics Connections (GEC), Magnetospheric Constellation (MagCon), and Living With a Star's geospace missions, which take advantage of adjustable orbit capability and the advancing technology of small spacecraft. Missions that involve large numbers of simply instrumented spacecraft are needed to develop a global view of the system and should be encouraged. NSF, for its part, should support extensive ground-based arrays of instrumentation to give a global, time-dependent view of this system. Ground- and space-based programs should be coordinated—as, for example, is being done in the Thermosphere-Ionosphere-Mesosphere Energetics and Dynamics (TIMED)/CEDAR program—to take advantage of the complementary nature of the two distinct viewpoints. NASA, NSF, DOD, and other agencies should encourage the development of theories and models that support the goal of understanding the A-I-M system from a dynamic point of view. Furthermore, these agencies should work toward the development of data analysis techniques, using modern information technology, that assimilate the multipoint data into a three-dimensional, dynamic picture of this complex system. Funding for the NASA Supporting Research and Technology (SR&T) program should be doubled to bring the proposal success rate up from 20 percent to the level found in other agencies. Solar Terrestrial Probe (STP) flight programs should have their own targeted postlaunch theory, modeling, and data analysis support.

MAJOR NSF INITIATIVE

Simultaneous, multicomponent, ground-based observations of the A-I-M system are needed in order to specify the many interconnecting dynamic and thermodynamic variables. As our understanding of the complexity of the A-I-M system grows, so does the requirement to capture observations of its multiple facets. The proposed Advanced Modular Incoherent Scatter Radar (AMISR) (Box 3.2) will provide the opportunity for coordinated radar-optical studies of the aurora and coordinated investigations of the lower thermosphere and mesosphere, a region not well accessed by spacecraft. Initial location at Poker Flat, Alaska, will allow coordination of radar with in situ rocket measurements of auroral processes. Subsequent transfer to the deep polar cap will enable studies of polar cap convection and the mapping of processes deeper in the geomagnetic tail.

1. The National Science Foundation should extend its major observatory component by proceeding as quickly as possible with Advanced Modular Incoherent Scatter Radar (AMISR) and by developing one or more lidar-centered major facilities. Further, the NSF should begin an aggressive program to field hundreds of small automated instrument clusters to allow mapping the state of the global system.

Ground-based sensors have played a pivotal role in our understanding of A-I-M science and must continue to do so in the coming decade and beyond. Anchored by a state-of-the-art, phased-array scientific radar, the $60 million AMISR is a crucial element for A-I-M. A distributed array of instrument clusters would provide

BOX 3.1 MAGNETOSPHERIC MULTISCALE MISSION

The Magnetospheric Multiscale (MMS) mission is a multispacecraft Solar Terrestrial Probe to study magnetic reconnection, charged particle acceleration, and turbulence in key boundary regions of Earth's magnetosphere. These three processes—which control the flow of energy, mass, and momentum within and across plasma boundaries—occur throughout the universe and are fundamental to our understanding of astrophysical and solar system plasmas. Only in Earth's magnetosphere, however, are they readily accessible for sustained study through the in situ measurement of plasma properties and of the electric and magnetic fields that govern the behavior of the plasmas. But despite four decades of magnetospheric research, much about the operation of these fundamental processes remains unknown or poorly understood. This state of affairs is in large part attributable to the limitations imposed on previous studies by their dependence upon single-spacecraft measurements, which are not adequate to reveal the underlying physics of highly dynamic, highly structured space plasma processes.

To overcome these limitations, MMS will employ four co-orbiting spacecraft, identically instrumented to measure electric and magnetic fields, plasmas, and energetic particles. The initial parameters of the individual spacecraft orbits will be designed so that the spacecraft will form a tetrahedron near apogee. Thus configured, the MMS "cluster" will be able to measure three-dimensional fields and particle distributions and their temporal variations and three-dimensional spatial gradients with high resolution while dwelling in the key magnetospheric boundary regions, from the subsolar magnetopause to the high-latitude magnetopause, and from the near tail to the distant tail. Adjustable interspacecraft separations—from 10 km up to a few tens of thousands of kilometers—will allow the cluster to probe the microphysical aspects of reconnection, particle acceleration, and turbulence and to relate the observed microprocesses to larger-scale phenomena. MMS will uniquely separate spatial and temporal variations over scale lengths appropriate to the processes being studied—down to the kinetic regime beyond the approximations of MHD. From the measured gradients and curls of the fields and particle distributions, spatial variations in currents, densities, velocities, pressures, and heat fluxes will be calculated.

In order to sample all of the magnetospheric boundary regions, MMS will employ a unique four-phase orbital strategy involving carefully sequenced changes in the local time and radial distance of apogee and, in the third phase, a change in the inclination of the orbit from 10 degrees to 90 degrees. In the first two phases, the investigation will focus on the near-Earth tail and the subsolar magnetopause (Phase 1; 12 R_E apogee) and on the low-latitude magnetopause flanks and near-Earth neutral line region (Phase 2; apogee increasing from 12 to 30 R_E). In Phase 3, MMS will use a lunar gravity assist to achieve a deep-tail orbit with apogee at 120 R_E and to effect the inclination change to 90 degrees. In this phase, MMS will study plasmoid evolution and reconnection at the distant neutral line. In the final, high-inclination phase, perigee will be increased to 10 R_E and apogee reduced to 40 R_E on the night side, and the MMS cluster will skim the dayside magnetopause from pole to pole, sampling reconnection sites at both low and high latitudes.

The nominal MMS mission has an operational duration of 2 years. While some mission-enhancing technologies such as an interspacecraft ranging and alarm system are desirable, no new mission-enabling technologies are required for the successful accomplishment of the MMS science objectives.

MMS is a mission of both exploration and understanding. Its primary thrust is to study on the meso- and microscales the basic plasma processes that transport, accelerate, and energize plasmas in thin boundary and current layers—the processes that control the structure and dynamics of the magnetosphere. With sensitive instrumentation and variable spacecraft orbits and interspacecraft spacing, MMS will integrate for the first time observations and theories over all geomagnetic scale sizes, from boundary layer processes that operate at the smallest scale lengths to the mesoscale dynamics that couple solar wind energy throughout the Earth's space environment.

The major science goals of the MMS mission include an understanding of the following:

- Reconnection at the magnetopause at high and low latitudes,
- Reconnection in the magnetotail and the associated magnetotail dynamics,
- Plasma entry into the magnetosphere,
- Physics of current sheets,
- Substorm initiation processes, and
- Cross-scale coupling between micro- and mesoscale phenomena.

BOX 3.2 ADVANCED MODULAR INCOHERENT SCATTER RADAR

The Advanced Modular Incoherent Scatter Radar (AMISR) is a state-of-the-art phased-array incoherent scatter radar (ISR). This highly versatile instrument will be ringed by less expensive complementary systems, typically optical in nature. The science plan is to target unsolved problems in aeronomy by placing AMISR in appropriate geographic locations for periods of 3 to 5 years. The first science goal is to understand the coupling between the neutral atmosphere and the high-speed current-carrying plasma in the auroral oval. This interaction involves electrodynamic forcing via momentum transfer from the plasma to the neutrals, Joule heating due to the currents that flow, composition changes of the thermosphere, and particle impact ionization associated with the aurora, to name just a few aspects. The first AMISR site will be in the Fairbanks area to take advantage of existing instrumentation and the Poker Flat Rocket Range.

Subsequent sites will be decided on the basis of community input by a panel of research scientists. Candidates include a location in the deep polar cap, which has never been studied using the ISR technique, and a location in the off-equatorial zone to study development of the ionospheric anomaly and its severe effects on communications systems.

The full AMISR will have three faces, each of which is a phased-array ISR capable of pulse-to-pulse beam swinging. The system will provide measurements of electric fields, ion and electron temperatures, electron density, ion composition, and neutral winds in the meridian plane. Three faces will allow a very wide area to be studied from a single location. Alternatively, the faces can be deployed separately since each is in its own right a very powerful system. A complementary set of optical- and radiowave-based sensors will accompany the deployment of the AMISR and extend its capabilities.

The design of the AMISR is completed and a prototype element has been constructed and tested successfully. Once the project is approved, a first face can be constructed in about 2 years. Subsequent faces will be online in about the same time scale. The total cost is $60 million, including the associated additional instrumentation.

the high temporal and spatial resolution observations needed to drive the assimilative models, which the panel hopes will parallel the weather forecasting models we now have for the lower atmosphere. Much of the necessary infrastructure for such a project has already been demonstrated in the prototype Suominet, a nationwide network of simple Global Positioning System/meteorology stations linked by the Internet. The proposed program would add miniaturized instruments, such as all-sky imagers, Fabry-Perot interferometers, very-low-frequency receivers, passive radars, magnetometers, and ionosondes in addition to powerful GPS-based systems in a flexible and expandable network coupled to fast, real-time processing, display, and data distribution capabilities. Instrument clusters would be sited at universities and high schools, providing a rich hands-on environment for students and training with instruments and analysis for the next generation of space scientists. Data and reduced products from the distributed network would be distributed freely and openly over the Internet. An overall cost of $100 million over the 10-year planning period is indicated. Estimated costs range from $50,000 to $150,000 per station depending on the instruments to be deployed. Adequate funding would be included for the development and implementation of data transfer, analysis, and distribution tools and facilities. Such a system would push the state of the art in information technology as well as instrument development and miniaturization.

Extending the present radar-centered upper atmospheric observatories to include one or more lidar-centered facilities is crucial if we are to understand the boundary between the lower and upper atmosphere. Fortunately, a number of military and nonmilitary large-aperture telescopes may become available for transition to lidar-based science in the next few years. Highest priority would be given to a facility at the same geographic latitude as one of the existing radar sites.

NASA ORBITAL PROGRAMS

The Explorer program has since the beginning of the space age provided opportunities for studying the geospace environment just as the Discovery program now provides opportunities in planetary science. The continued opportunities for University-Class Explorer (UNEX) (Box 3.3), Small Explorer (SMEX), and Medium-Class Explorer (MIDEX) missions, practically defined in terms of their funding caps of $14 million, $90 million, and $180 million, allow the community the greatest creativity in developing new concepts and a faster response time to new developments in both science and technology. These missions also provide a crucial training ground for graduate students, managers, and engineers.

BOX 3.3 KEY ROLE OF UNEX SPACE MISSIONS

Small spacecraft missions can be extremely productive scientifically and can also provide a fertile training ground for students of science and engineering. NASA has attempted to establish several new lines of small-end missions, including the UNEX mission line in space science. Of course, the Small Explorer (SMEX) program (at a larger scale size and cost) has been a remarkable success, and the smaller sounding rocket and balloon programs have in the past been immensely rich and rewarding programs. In carrying out the Student Explorer Demonstration Initiative (STEDI) program, the Universities Space Research Association set an excellent tone for how to manage small missions. Appropriate levels and numbers of reviews were employed and key types of help were provided to STEDI teams, as needed.

It has been widely acknowledged that small-spacecraft missions can provide a profound educational experience for university students. It has been from the ranks of such highly trained students that many present-day principal investigators of NASA space science missions have emerged. To have a future space science program with strong experimental content, the United States must ensure that student training continues to be a high priority. This demands that small, focused spacecraft missions be available to the university research community, which in turn means that an ample number of spacecraft payloads must be made available to researchers.

The NASA UNEX program was generally viewed as a direct successor to the STEDI program. However, UNEX has been effectively cut from future NASA budgets. It is regrettable that this program and the opportunities afforded by the STEDI concept will not be available to university scientists for research and educational opportunities. Moreover, the stresses that apparently continue to occur in the sounding rocket and balloon programs of NASA suggest that the suborbital program also is very limited in the access to space it gives for young scientists and engineers and as a hands-on training ground for them. See *A Space Physics Paradox* for further discussion.[1]

It would seem that NASA has identified larger-spacecraft missions as its primary focus of attention and funding. This means that very small, PI-class spacecraft missions are not a high priority for it. NASA and other agencies could serve the university community in a most beneficial and effective way if they would offer low-cost launch possibilities to university groups. This would allow the community to revivify the UNEX program, establish appropriate small-spacecraft launch capabilities, strengthen the engineering and science education program, and fully develop this nation's small satellite program potential. In carrying out these steps, the agencies would perform an immense service for university researchers throughout the nation. At a cost of ~$20 million per mission and with launches once or so per year, the program would make very modest resource demands.

---

[1] NRC. 1994. *A Space Physics Paradox: Why Has Increased Funding Been Accompanied by Decreased Effectiveness in the Conduct of Space Physics Research?* National Academy Press, Washington, D.C.

Imager for Magnetopause-to-Aurora Global Exploration (IMAGE), launched in March 2000, is an example of a highly successful MIDEX mission; it was preceded by the first two ongoing SMEX missions, Solar Anomalous and Magnetospheric Particle Explorer (SAMPEX) and Fast Auroral Snapshot Explorer (FAST), launched in 1992 and 1996, which have provided enormous scientific return for the investment. The Aeronomy of Ice in the Mesosphere (AIM) SMEX was recently selected for launch in 2006. The UNEX program, after the great success of the Student Nitric Oxide Explorer (SNOE), launched in February 1998, has effectively been cancelled. This least expensive component of the Explorer program plays a role similar to the sounding rocket program, with higher risk accompanying lower cost and a great increase in the number of flight opportunities. An increase in funding to $20 million per mission with one launch per year would make this program viable with modest resources.

2. The SMEX and MIDEX programs should be vigorously maintained and the UNEX program should quickly be revitalized.

The Solar Terrestrial Probe (STP) line of missions defined in the NASA Sun-Earth Connection (SEC) Roadmap (strategic planning for 2000 to 2025) has the potential to form the backbone of A-I-M research in the next decade. The missions that are part of the current program include TIMED, launched in February 2002, Solar-B, Solar Terrestrial Relations Observatory (STEREO), MMS, GEC, and MagCon. After TIMED, launched in February 2002, the next A-I-M/STP mission, MMS, is in the process of instrument selection for a 2009 launch.

The STP cadence, with one A-I-M mission per decade (TIMED was significantly delayed), has fallen behind the NASA SEC Roadmap projections.

3. The panel heartily endorses the STP line of missions and strongly encourages an increase in the launch cadence, with GEC and MagCon proceeding in parallel.

The A-I-M research community has very successfully utilized the infrastructure developed within the ISTP program. The integration of the data from spacecraft and ground-based programs beyond those funded by the ISTP project itself—such as those of NOAA, LANL, and the DOD—have contributed substantially to our understanding of the global system.

Comparisons between the Sun-Earth system and other Sun-planet or stellar-planet systems provide important insights into the underlying physical and chemical processes that govern A-I-M interactions. Improved understanding of A-I-M coupling phenomena such as planetary and terrestrial auroras would benefit from such an approach.

4. The Sun-Earth Connection program partnership with the NASA Solar System Exploration program should be revitalized. A dedicated planetary aeronomy mission should be pursued vigorously, and the Discovery program should remain open to A-I-M-related missions.

NASA SUBORBITAL PROGRAM

The NASA Suborbital program has produced outstanding science throughout its lifetime (Box 3.4). Many phenomena have been discovered using rockets, rockoons, and balloons, and many outstanding problems brought to closure, particularly when space-based facilities are teamed with ground-based facilities. These phenomena include the auroral acceleration mechanism, plasma bubbles at the magnetic equator, the charged nature of polar mesospheric clouds, and monoenergetic auroral beams. This program continues to generate cutting-edge science with new instruments and data rates that are more than an order of magnitude greater than typical satellite data rates. Both unique altitude ranges and very specific geophysical conditions are accessible only to sounding rockets and balloons, particularly in the campaign mode. Many current satellite experimenters were trained in the Suborbital program, and high-risk instrument development can occur only in such an environment. To accomplish significant training, it is necessary that a graduate student remain in a project from start to finish and that some risk be acceptable; both are very difficult in satellite projects. The high scientific return, coupled with training of future generations of space-based experimenters, makes this program highly cost-effective.

The sounding rocket budget has been level-funded for over a decade, and many principal investigators are discouraged about the poor proposal success rate as well as the low number of launch opportunities. The sounding rocket program was commercialized in 2000; in this changeover, approximately 50 civil service positions were lost and the cost of running the program increased. Approved campaigns were delayed by up to a year, and it is not yet clear whether the launch rate will ever return to precommercialization levels. Effectively, commercialization has meant a significant decline in funding for the sounding rocket program. An additional concern is that, as currently structured—i.e., with a fixed, 3-year cycle for all phases of a sounding rocket project—funding is not easily extended to allow graduate students to complete their thesis work, because it is generally thought that such work should fall under the SR&T program, already oversubscribed. The rocket program has a rich history of scientific and educational benefit and provides low-cost access to space for university and other researchers. Further erosion of this program will result in fewer and fewer young scientists with experience in building flight hardware and will ultimately adversely affect the much more expensive satellite programs.

5. The Suborbital program should be revitalized and its funding should be reinstated to an inflation-adjusted value matching the funding in the early 1980s. To further ensure the vibrancy of the Suborbital Program, an independent scientific and technical panel should be formed to study how it might be changed to better serve the community and the country.

SOCIETAL IMPACT PROGRAM

The practical impact on society of variations in the A-I-M system falls into two broad categories: the well-established effects of space weather variations on technology and the less clear yet tantalizing influence of solar variability on climate. The societal impacts of space weather are broad—communications, navigation, human radiation hazards, power distribution, and satellite operations are all affected. Space weather is of international concern, and other nations are pursuing parallel activities, which could be leveraged through

BOX 3.4 SUBORBITAL PROGRAM

NASA's Suborbital program provides regular, inexpensive access to near-Earth space for a broad range of space science disciplines, including space plasma physics, astronomy, and microgravity. The program has been extremely successful throughout its history, consistently providing high scientific return for the modest funding invested. Phenomena from auroral physics to supernovae have been investigated with sounding rocket and balloon experiments, and many key discoveries in these fields have come from this program. For the science of lower-altitude regions such as equatorial ionospheric irregularities and mesospheric physics, this program provides the only access because these regions are too high for airplanes and too low for satellites. Moreover, the use of new, advanced instrumentation concepts coupled with data rates that exceed those of satellites by a factor of more than 10 have provided measurements of a resolution and quality that are simply unobtainable elsewhere. This science is exciting and central to NASA's mission.

The program also has a spectacular track record in training future scientists and engineers. More than 300 Ph.D. theses have been based on rocket data alone. This has provided a regular flow of technically adept individuals who have first-hand experience in building space flight instrumentation. Indeed, most of today's successful satellite experimenters were trained in this program. The experience that students receive—developing, flying, and interpreting data from instruments they themselves have built—is only possible within this program. This is an unparalleled learning experience that satellite programs cannot match because they are too risk-averse and span too long a period for a graduate student to be involved from start to finish.

Despite this tremendous track record in which the Suborbital program has continually demonstrated its scientific validity (made clear through a series of reviews over the past decade), funding for the program has seriously eroded. Its conversion to a government-owned, contractor-operated program, coupled with loss of many civil servant positions, has left the program severely underfunded for operations. Additionally, funding for scientific investigations has remained stagnant, resulting in a significant decline in the number of funded investigations over the past 15 years. There is great concern among the scientific community that NASA management does not deem the program sufficiently important to restore and protect its funding. This attitude must be changed and the program must be restored to a healthy level that will allow it to continue to play its important scientific and student training roles in which it is so uniquely effective. See *A Space Physics Paradox* for further discussion.[1]

---

[1] NRC. 1994. *A Space Physics Paradox: Why Has Increased Funding Been Accompanied by Decreased Effectiveness in the Conduct of Space Physics Research?* National Academy Press, Washington, D.C.

collaboration. The role of solar variability in climate change remains an enigma, but it is now at least being recognized as important to our understanding of the natural—as opposed to anthropogenic—sources of climate variability.

6. To maximize the societal impact of studies and knowledge of the A-I-M system, the study of solar variability—both of its short-term effects on the space radiation environment, communications, navigation, and power distribution and of its effect on climate and the upper atmosphere—should be intensified by both modeling and experimentation.

NASA's Living With a Star program, as defined by the Science Architecture Team report, should be implemented, with increased resources for the geospace component. The share of resources required for the Solar Dynamics Observatory, already defined before the start of LWS, has resulted in an unbalanced portfolio.

Missions such as the National Polar-orbiting Operational Environmental Satellite Systems (NPOESS) and Solar Radiation and Climate Experiment (SORCE) should provide vital scientific data for monitoring long-term solar irradiance, and NPOESS should provide ionosphere and upper atmosphere observations to fill gaps in measurements needed to understand the A-I-M system.

An L1 monitor should be a permanent facility, to provide solar wind measurements crucial to determining the response of the A-I-M system to its external driver.

The National Space Weather Program should be strengthened and used as a template for interagency cooperation. International participation in such large scope programs as LWS and NSWP is essential.

NASA's new Living With a Star program can, over the next decade, provide substantial new resources to address these goals. It is crucial that there be overlap between the geospace and solar mission components of LWS for the A-I-M system to be studied synergistically, that resources be adequate for the geospace component, and that theory, modeling, and the comprehensive data system that will replace the ISTP infrastructure be defined at the outset, as called for in the Science Architecture Team report.[2] The National Space Weather Program, a multiagency endeavor established in 1995, addresses the potentially great societal impact of the physical processes from the Sun to Earth that affect the near-Earth environment in ways as diverse as terrestrial weather. The program specifically addresses the need to transition scientific research into operations and to assist users affected by the space environment. Such multiagency cooperation is essential for progress in predicting response of the near-Earth space environment to short-term solar variability.

Several potential mechanisms for a solar variability-climate connection have been suggested: (1) changes in the Sun's total irradiance or luminosity, which is the basic driver of the climate system; (2) changes in spectral irradiance, particularly in the UV, which drives the chemistry and dynamics of the middle atmosphere and has been shown by modeling studies to influence the dynamics of the troposphere; and (3) the possible influence of cosmic-ray and electric-field variations on cloud nucleation, which could significantly modify Earth's radiation balance.

7. The NOAA, DOE/LANL, and DOD operational spacecraft programs should be sustained and DOD launch opportunities should be utilized for specialized missions such as geostationary airglow imagers, auroral oval imagers, and neutral/ionized medium sensors.

The interagency cooperation established in the NSWP is outstanding and is a model for extracting the maximum benefit from scientific and technical programs. It has also been effective at bringing together different scientific disciplines and the scientific and operations communities. Interagency cooperation has worked well in the AFOSR/NSF Maui Mesosphere and Lower Thermosphere Program, and it has been key to the success of the NOAA GOES and NPOESS programs of meteorlogical satellites with space environment monitoring capabilities. International multiagency cooperation has been very successful for the ISTP program, which involves U.S., European, Japanese, and Russian space agencies. Global studies require such international cooperation. The panel recognizes that much more science can be extracted by careful coordination of ground- and space-based programs.

MAXIMIZING SCIENTIFIC RETURN

Funding for NASA Supporting Research and Technology, including guest investigator studies and focused theory, modeling, and data assimilation efforts, is essential for maximizing the scientific return from large investments in spacecraft hardware.

Supporting Research and Technology

While spacecraft hardware projects are concentrated at relatively few institutions, the NASA SR&T program is the primary vehicle by which independent investigations can be undertaken by the broader community. Likewise, NSF helps individual investigators to carry out targeted research through its Division of Atmospheric Sciences base programs—SHINE, CEDAR, and GEM. Such individual PI-driven initiatives are the most inclusive, with data analysis as well as theoretical efforts and laboratory studies, and often lead to the highest science return per dollar spent. The funding for such program elements falls far short of the scientific opportunities, with the current success rate for submitted NASA SR&T proposals being 10 to 20 percent. Furthermore, limited available SR&T funds have been used for guest investigator participation in underfunded STP-class flight programs. Without adequate MO&DA funding for NASA orbital and suborbital programs, the SR&T budget intended for targeted research on focused scientific questions has been utilized to support broader data analysis objectives.

8. The funding for the SR&T program should be increased, and STP-class flight programs should have their own targeted postlaunch data analysis support.

9. A new small grants program should be established within NSF that is dedicated to comparative atmospheres, ionospheres, and magnetospheres (C-A-I-M).

---

[2] NASA, Living With a Star, Science Architecture Team. 2001. *Report to the Sun-Earth Connection Advisory Subcommittee,* August. Available online at <http://lws.gsfc.nasa.gov/docs/lws_sat/sat_report2.pdf>.

BOX 3.5 COMPARATIVE ATMOSPHERES, IONOSPHERES, AND MAGNETOSPHERES (C-A-I-M) PROGRAM

The comparison of the Sun-Earth system to other Sun-planet systems can provide unique insights into how atmospheres, ionospheres, and magnetospheres (A-I-M) respond to solar inputs. The physical and chemical processes controlling these responses manifest themselves very differently at each solar system body, yet are essentially the same at a basic level. All the techniques that have been used so successfully to understand solar-terrestrial physics (e.g., modeling, ground-based and remote observations, and in situ measurements) need to be applied to other planets and bodies, so that the study of solar-planetary relations becomes the natural extension of the terrestrial space weather effort. Achieving this goal will require several elements, including NASA planetary missions dedicated to A-I-M goals (or including significant A-I-M capabilities), a Discovery program in which A-I-M missions are included, and a grants program within NSF that is dedicated to comparative atmospheres, ionospheres, and magnetospheres (C-A-I-M).

A new C-A-I-M grants program at NSF would play a key role in addressing the interdisciplinary issues needed to understand and relate A-I-M processes throughout our solar system, or even at other stellar systems. Such a grants program would provide much needed resources for analysis of both past and future data sets (from ground- or space-based observatories, or from in situ missions), modeling and data interpretation related to A-I-M objectives, telescope time, special meetings devoted to terrestrial-planetary issues, and the nonmission research support needed to encourage C-A-I-M science activities in the community. Such a grants program would need about $5 million per year in order to adequately develop and explore the linkages between the terrestrial and planetary manifestations of atmosphere-ionosphere-magnetosphere physics.

A new C-A-I-M grants program at NSF (Box 3.5) would allow the techniques that have been applied so successfully to A-I-M processes at Earth (modeling, ground- and space-based observations, and in situ measurements) to be used to understand A-I-M processes at other planets. Such a comparative approach would improve understanding of these processes throughout the solar system, including at Earth. Presently, a modest $2 million Planetary Science program at NSF covers all of solar system science (except for solar and terrestrial studies), with only a small fraction going to planetary A-I-M research.

Theory, Modeling, and Data Assimilation

Theory and modeling provide the framework for interpreting, understanding, and visualizing diverse measurements at disparate locations in the A-I-M system. There is now a pressing need to develop and utilize data assimilation techniques not only for operational use in specifying and forecasting the space environment but also to provide the tools to tackle key science questions. The modest level of support from the NSF base programs (CEDAR, GEM, SHINE) and NASA SR&T has been inadequate to build comprehensive, systems-level models. Rather, individual pieces have been built and first stages of model integration achieved with funding from such programs as NASA's ISTP and its Sun-Earth Connections Theory Program, the AFOSR/ONR Multidisciplinary University Research Initative program, NSF Science and Technology Center programs, and the multiagency support for such efforts as the Community Coordinated Modeling Center. Such programs enable the development of theory and modeling infrastructure, including models to address the dynamic coupling between neighboring geophysical regions. Their value to the research community is clearly their provision of longer-term funding, which has been essential to developing a comprehensive program, outside the purview of SR&T.

10. The development and utilization of data assimilation techniques should be enhanced to optimize model and data resources. The panel endorses support for theory and model development at the level of the NASA Sun-Earth Connections Theory Program, the AFOSR/ONR MURI program, NSF Science and Technology Center programs, and the multiagency support for such efforts as the Community Coordinated Modeling Center (CCMC). Support should be enhanced for large-scope, integrative modeling that applies to the coupling of neighboring geophysical regions and physical processes, which are explicit in one model and implicit on the larger scale.

The preceding science recommendations are grouped into three cost categories and prioritized in Table 3.1. Equal weight is given to STP and LWS lines, as indicated by funding level. Small programs are ranked by resource allocation, while the Advanced Modular Incoherent Scatter Radar is the highest-priority moderate initiative at lower cost than others.

BIBLIOGRAPHY

GEM documents. Available online at <http://www-ssc.igpp.ucla/gem/Welcome.html>.

National Aeronautics and Space Administration (NASA), Office of Space Sciences. 2000. *Strategic Plan*. NASA, Washington, D.C.

NASA. 2000. *Sun-Earth Connection Roadmap, Strategic Planning for 2000–2020*. NASA, Washington, D.C.

NASA, Living With a Star, Science Architecture Team. 2001. *Report to the Sun-Earth Connection Advisory Subcommittee*, August. Available online at <http://lws.gsfc.nasa.gov/docs/lws_sat/sat_report2.pdf>.

National Research Council (NRC). 2000. *Radiation and the International Space Station: Recommendations to Reduce Risk*. National Academy Press, Washington, D.C.

NRC. 1994. *A Space Physics Paradox: Why Has Increased Funding Been Accompanied by Decreased Effectiveness in the Conduct of Space Physics Research?* National Academy Press, Washington, D.C.

National Science Foundation (NSF). 1995. *National Space Weather Program: The Strategic Plan,* FCM-P30-1995. Office of the Federal Coordinator for Meteorology, Silver Spring, Md.

NSF. 1997. *National Space Weather Program: The Implementation Plan,* FCM-P31-1997. Office of the Federal Coordinator for Meteorology, Silver Spring, Md.

NSF. 1997. *CEDAR Phase III Document*. Available online at <http://cedarweb.hao.ucar.edu/index.html>.

4

Report of the Panel on Theory, Modeling, and Data Exploration

SUMMARY

Today, space and solar physics presents both great opportunities and challenges, stemming from a science that is changing character from strongly exploratory and discovery driven to more mature and explanatory driven. Furthermore, the societal and economic impacts of solar and space physics, through the forecasting of space weather, have become increasingly important. The opportunities, challenges, and societal and economic impacts place significant new demands on theory, modeling, and data exploration. Theory and modeling act to interpret observations, making them meaningful within the context of basic physics. Frequently, theory reveals that seemingly disparate observed phenomena correspond to the same physical processes in a system (or better yet, in many systems). Furthermore, besides their roles in organization and understanding, theory and modeling can predict otherwise unexpected but important or relevant phenomena that may subsequently be observed and that might otherwise not have been discovered.

This report of the Panel on Theory, Modeling, and Data Exploration provides a brief survey of key aspects of theory, modeling, and data exploration and offers four major recommendations. The survey is not exhaustive but instead emphasizes basic issues concerning the nature of theory and its integrative role in modeling and data exploration. The panel offers a new synthesis for the organization and integration of space physics theory, modeling, and data exploration: coupling complexity in space plasma systems. "Coupling complexity" refers to the class of problems or systems that consist of significantly different scales, regions, or particle populations and for which more than one set of defining equations or concepts is necessary to understand the system. For example, the heliosphere contains cosmic rays, solar wind, neutral atoms, and pickup ions, each of which interacts with the others but needs its own set of equations and coupling terms. Similarly, the ionosphere-thermosphere and magnetosphere are different regions governed by distinct physical processes.

From this synthesis, the four major recommendations flow naturally. They are designed to (1) dramatically improve and expand space physics theory and modeling by embracing the idea of coupling complexity (or, equivalently, nonlinearity and multiscale and multiprocess feedback) within space plasma systems, (2) increase access to diverse data sets and substantially augment the ability of individual investigators to explore space physics phenomenology by redesigning the archiving, acquisition, and analysis of space physics data sets, (3) strengthen the role of theory and data analysis in space-based and ground-based missions, and (4) support and strengthen the application of space physics research to economic, societal, and governmental needs, especially in areas where space weather and space climatology impact human activities and technological systems.

THE COUPLING COMPLEXITY RESEARCH INITIATIVE

For theoreticians, modelers, and data analysts, the great challenges of space physics often result from the closely intertwined and integrated coupling of different spatial regions, disparate scales, and multiple plasma and atomic constituents in the solar, interplanetary, geospace, and planetary environments. To embrace the demands imposed by hierarchical coupling or coupling complexity—nonlinearity and multiscale, multiprocess, multiregional feedback in space physics—space physicists must address a number of challenges:

- Formulation of sophisticated models that incorporate disparate scales, processes, and regions and the development of analytic theory;
- Computation;
- Incorporation of coupling complexity into computational models;
- Integration of theory, modeling, and space- and ground-based observations;
- Data exploration and assimilation; and
- Transition of scientific models to operational status in, for example, space weather activities.

Recommendation 1. NASA should take the lead in creating a new research program—the Coupling Complexity Research Initiative—to address multiprocess coupling, nonlinearity, and multiscale and multiregional feedback in space physics. The research program should be peer reviewed. It should do the following:

- **Provide long-term, stable funding for a 5-year period.**
- **Provide sufficiently large grants that critical-mass-sized groups of students, postdoctoral associates, and research scientists, gathered around university and institutional faculty, can be supported.**
- **Provide funding to support adequate computational resources and infrastructure for the successfully funded groups.**

- **Facilitate the development and delivery of community-based models.**
- **Use the grants to leverage faculty and permanent positions and funding from home institutions such as universities, laboratories, institutes, and industry.**

This research program would emphasize the development of coupled global models and the synergistic investigation of well-chosen, distinct theoretical problems that underlie the basic physics inherent in the fully general self-consistent space physics problem. For major advances to be made in understanding coupling complexity in space physics, sophisticated computational tools, fundamental theoretical analysis, and state-of-the-art data analysis must all be brought under a single umbrella program. Thus, computational space physicists, theoreticians working with pen and paper, and data analysts need to be part of a single research program addressing a major problem in space physics. The models and algorithms developed by these research groups will make a major contribution to future National Aeronautics and Space Administration (NASA), National Science Foundation (NSF), and National Oceanic and Atmospheric Administration (NOAA) activities, especially those that focus on remote sensing and multipoint measurements. The models/algorithms will (1) couple measurements made at different times and places, (2) integrate the effect of multiscale processes so that the processes can be related to line-of-sight (column-integrated) remote-sensing measurements, and (3) provide a framework within which large multispacecraft data sets can be organized.

A fundamental component of this recommendation is that the award of a grant will carry with it the expectation of a commitment from the home institution (university, laboratory, industry) to develop a stable, long-term program in space physics by creating permanent positions; this would provide an intellectual environment within which large research efforts can flourish and would allow for critical-mass efforts.

Since nearly 30 groups submitted proposals to the most recent NASA Sun-Earth Connections Theory Program, the panel used this as an indication of the number of large groups that exist currently in the United States. Accordingly, it recommends that the Coupling Complexity Initiative should support 10 groups, each with funding of between $500,000 and $1 million per year. This would require committing $7.5 million to $10 million per year in funding. The panel recommends the formation of a cross-agency commission, with NASA possibly taking the lead through its Living With a Star program, to examine the implementation of a cross-agency Coupling Complexity Initiative.

THE GUEST INVESTIGATOR INITIATIVE

Related to the five tasks listed in Recommendation 1, data and theory face challenges in two areas:

- Integrating theory, modeling, and space- and ground-based observations and
- Data exploration and assimilation.

To address these points in the context of solar and space physics modeling and data analysis, the panel offers a second recommendation:

Recommendation 2. The NASA Guest Investigator program should (1) be mandatory for all existing and new missions, (2) include both space- and ground-based missions, (3) be initiated some 3 to 5 years before launch, and (4) be peer reviewed and competed for annually, with grant durations of up to 3 years. Funding, at a minimum equivalent to 10 percent of the instrument cost, should be assigned to the Guest Investigator program and should explicitly support scientists working on mission-related theory and data analysis. Further, the Guest Investigator program for each mission should have the same status as a mission instrument. Other agencies should also consider guest investigator initiatives with their programs.

The panel strongly supports and endorses the current NASA Guest Investigator program and would like to see it strengthened, with similar programs created in other agencies. The implementation of this recommendation would address the very real concerns expressed by many experimentalists that too few theorists play an active role in exploring, interpreting, refining, and extending the observations returned by expensive missions. The panel notes that in an era of "fast missions," an already active cadre of theorists and data explorers should be in place to take full advantage of a newly launched mission. Furthermore, a robust Guest Investigator initiative may also address the concern that NASA expects principal investigators for experiments to submit proposals with extensive science goals but does not provide sufficient funding to support the science.

At least 10 percent of the instrument cost should be assigned to the Guest Investigator program and should be budgeted in the mission costs from the outset. The panel recommends that Guest Investigator programs begin a few years prior to launch.

A DISTRIBUTED VIRTUAL SPACE PHYSICS INFORMATION SYSTEM

This recommendation is intended to increase access to diverse data sets and substantially augment the ability of individual investigators to explore space physics phenomenology by redesigning the archival, acquisition, and analysis of space physics data sets.

Recommendation 3. NASA should take the lead in convening a cross-agency consultative council that will assist in the creation of a cross-agency, distributed space physics information system (SPIS). The SPIS should link (but not duplicate) national and international data archives through a suite of simple protocols designed to encourage participation of all data repositories and investigator sites with minimal effort. The data environment should include both observations and model data sets and may include codes used to generate the model output. The panel's definition of data sets includes simulation output and supporting documentation.

Among other tasks, the system should do the following:

1. Maintain a comprehensive online catalog of both distributed and centralized data sets.
2. Generate key parameter (coarse resolution) data and develop interactive Web-based tools to access and display these data sets.
3. Provide higher-resolution data; error estimates; supporting platform-independent portrayal and analysis software; and appropriate documentation from distributed principal investigator sites.
4. Permanently archive validated high-resolution data sets and supporting documentation at designated sites and restore relevant offline data sets to online status as needed.
5. Develop and provide information concerning standard software, format, timing, coordinate system, and naming conventions.
6. Maintain a software tree containing analysis and translation tools.
7. Foster ongoing dialogues between users, data providers, program managers, and archivists, both within and across agency boundaries.
8. Maintain portals to astrophysics, planetary physics, and foreign data systems.
9. Survey innovations in private business (e.g., data mining) and introduce new data technologies into space physics.
10. Regularly review evolving archival standards.
11. Support program managers by maintaining a reference library of current and past data management plans, reviewing proposed data management plans, and monitoring subsequent adherence.

The primary objective of this recommendation is to establish a research environment for the disparate data sets distributed throughout the space physics community. By providing a cross-agency framework that encompasses agency-designated archives and PI sites, the proposed space physics information system will enable users to identify, locate, retrieve, and analyze both observations and the results of numerical simulations. The community-maintained SPIS will facilitate the introduction of innovative data mining techniques and methodology from private business into the scientific community and improve communication between users and data providers. It will assist agency managers and PIs by archiving project data management plans and documenting best practices.

The panel recommends that SPIS begin at a modest level and grow only with demonstrated need and success. Periodic competitions for all tasks within the data system will help impose cost constraints. The central node would consist of a full-time project scientist, project manager, project engineer, and administrative assistant. The four discipline nodes employ half-time scientists, quarter-time administrators, and half-time programmers. The central and discipline nodes provide the framework to link the various data sets. With the help of advisory committees from the research and operational communities, the central and discipline nodes will identify and fund tasks at more ephemeral subnodes. A fully operational system might cost $10 million annually.

THE TRANSITION INITIATIVE

The applications of space physics research to economic, societal, and governmental needs, especially in areas where space weather and space climatology impact human activities and technological systems, need to be supported and strengthened. For models to be an effective bridge between space physics theory and economic, societal, and governmental needs, they have to address the demands imposed by coupling complexity and be sufficiently robust, validated, documented, standardized, and supported. These additional demands impose significant challenges to the groups that develop models, and the criteria for transitioning a model successfully to operational use are frequently very different from those needed to develop models purely for research purposes.

Recommendation 4. NOAA and the Air Force should initiate a program to support external research groups in the transitioning of their models to NOAA and Air Force rapid prototyping centers (RPCs). Program support should include funding for documentation, software standardizing, software support, adaptation of codes for operational use, and validation and should allow for the research group to assist in making the scientific codes operational. The RPC budgets of the NOAA Space Environment Center and the Air Force Space and Missiles Center/DSMP Technology Applications Division (SMC/CIT) should be augmented to facilitate the timely transition of models.

Despite the many solar, heliospheric, and geospace models that can potentially be used for operational space weather forecasting, relatively few have so far been transitioned into operation at the NOAA and Air Force space weather centers. This is due to inadequate resources to support transition efforts, in particular at the NOAA Space Environment Center. The recommended transition initiative addresses this problem directly.

Costs, particularly for the Air Force, are difficult to forecast precisely. For NOAA, the panel estimates that $1 million would be required to start and support the first year of operation, with $500,000 per year to support three permanent NOAA staff and an additional $500,000 for operational support, real-time data links, software standardization, documentation, and related expenditures. Since the panel anticipates that between three and five codes per year will be readied for transition at a cost of ~$200,000 each, an additional $1 million per year should be available for competition. A conservative annual budget is therefore between $2 million and $2.5 million.

The transition initiative should be funded through a partnership between the Air Force and NOAA, with the precise levels of funding determined by the needs of each. The NOAA Space Environment Center and Air Force RPC budgets will have to be augmented and greater industrial and business support developed.

4.1 INTRODUCTION

Space physics is positioned at a critical juncture. The past decade has witnessed a shift from a strongly exploratory and discovery-driven science to a more mature explanatory science. Furthermore, the societal and economic impact of forecasting space weather has become increasingly important. With its maturing, solar and space physics offers exciting opportunities and places significant new demands on theory, modeling, and data exploration. Theory and modeling act to interpret observations, making them meaningful in the context of basic physics. Frequently, theory reveals that seemingly disparate observed phenomena correspond to the same physical processes in a system (or better yet, in many systems). Furthermore, besides their roles in organization and understanding, theory and modeling can predict otherwise unexpected yet important or relevant phenomena that may subsequently be observed and that might otherwise not have been discovered. The dozen boxes throughout the report illustrate the complementary roles of theory and observation in space physics.

In this report, the panel briefly surveys key aspects of theory, modeling, and data exploration and offers four major recommendations. The recommendations are based on the deliberations of panel members, all of whom have been actively involved in the application of theory, modeling, and/or data assimilation to contemporary problems in solar and space physics, and on numerous interactions with the community of researchers and agency officials concerned with the issues outlined in this report. At its meetings, the panel received presentations from a wide range of space physicists and agency officials. In reaching its findings, the panel drew on these experiences and on an extensive public dialogue with the broad space physics community. For example, the panel held town-hall-style meetings at various conferences and met with participants at the fall 2001 American Geophysical Union meeting, the Advanced Composition Explorer (ACE) workshop, and the NSF-sponsored meetings of the Geospace Environment Modeling (GEM) program and the Coupling, Energetics, and Dynamics of Atmospheric Regions (CEDAR) program. The panel also solicited comments via advertisements in electronic publications such as SPA News.

This report is not intended to be exhaustive; instead, it emphasizes basic issues surrounding the nature of theory and its integrative role in modeling and data exploration. Since examples illustrate themes far better than dry narrative, the panel offers representative space physics success stories (Boxes 4.1 to 4.10). The stories illustrate the importance of theory in synthesizing observations and describe its role in driving, and being driven by, data exploration—processes that can lead to the identification of new scientific frontiers.

From the survey, the panel offers a new synthesis for the organization and integration of space physics theory, modeling, and data exploration: coupling complexity in space plasma systems. Putting it another way, the panel explores the science of nonlinearity and multiscale and multiprocess feedback in space plasma systems. A precise definition of coupling complexity is given in Section 4.4, "A Coherent Framework for Theory, Modeling, and Data Exploration." From this synthesis flows a set of four major recommendations designed to achieve four goals: (1) dramatically improve and expand space physics theory and modeling by embracing the idea of coupling complexity (or, equivalently, nonlinearity and multiscale and multiprocess feedback) within space plasma systems, (2) increase access to diverse data sets and substantially augment the ability of individual investigators to explore space physics phenomenology by redesigning the archiving, acquisition, and analysis of space physics data sets, (3) strengthen the role of theory and data analysis in space-based and ground-based missions, and (4) support and strengthen the application of space physics research to economic, societal, and governmental needs, especially where space weather and space climatology impact human activities and technological systems.

4.2 THEORY, COMPUTATION, AND SUCCESS STORIES

Theory serves to identify fundamental problems and processes that cross traditional boundaries or divisions. It acts as a framework within which to interpret observations; it can integrate apparently disparate observations and data sets, and it provides guidance in the development of tools, data sets, and the observations needed to obtain a complete understanding of heliospheric plasma physics processes. As space physics matures from an exploratory science, theory is beginning to assume an increasingly important role in defining the frontiers of the science, introducing new concepts that provide the framework for observational missions, observatories, and discoveries. Finally, space physics in general, as interpreted and guided by theory, occupies a unique place within the broader field of astrophysics in that it offers the only accessible laboratory in which to develop theories, build models, and test by observation plasma physics processes on all scales.

All these ideas are exemplified by the success stories presented in boxes throughout the text. Some of the stories highlight the role that theory has taken in developing new fields, and some show how theory and modeling can have an impact on society and the economy. Others address, resolve, and develop new questions in basic theory, and yet others illustrate how questions that arise in space physics frequently translate into basic questions for physics or astrophysics.

Numerical modeling has made dramatic advances over the past decade and is now considered by some space physicists to be a fourth branch of scientific methodology besides theory, experiment, and observation. Several factors have contributed to this advance:

1. Moore's law continues to hold: Computational power has increased and continues to increase exponentially, doubling approximately every 18 months.
2. New and more accurate and efficient numerical algorithms have been developed. Examples include flux-limited fluid and magnetohydrodynamics (MHD) algorithms and implicit particle codes.
3. Many models are now three-dimensional (spatial), an essential capability for addressing many kinds of problems.
4. Numerical models have become more realistic by including additional physical processes and by using more realistic model parameters.
5. Models with different algorithmic strengths or models of interaction regions are being coupled. Examples are codes addressing the coupling of the magnetosphere-ionosphere and ionosphere-thermosphere.
6. Numerical models are directly compared with observations, and their predictive capabilities are tested.

Computer modeling can be divided roughly into three classes, characterized as follows:

1. *Modeling of a limited or local system or process with as much realism and/or mathematical accuracy as possible.* The goal is to isolate a process, study it in detail, and extract the underlying physics. Sometimes the problem may be well posed mathematically, in which case the computer modeling approach tries to utilize the best numerical algorithms possible; but other times the problem may not admit a simple or tractable mathematical formulation, and the computer modeling may take the form of simulating the underlying physical processes in the hope that the collective behavior of the system is captured. Examples might be kinetic simulations of a reconnection-diffusion region, shock interaction modeling using high-order shock-capturing schemes, or turbulence modeling based on highly resolved spectral codes.

2. *Modeling that simulates large, possibly spatially or temporally complicated systems that can be described at a satisfactory level by a relatively simple, well-formulated set of mathematical equations (e.g., the MHD equations)*. The goal here is to solve the system in its full complexity (three dimensions, time-dependent, with realistic boundary conditions and initial data). This class of modeling requires the development of stable, robust, accurate, and efficient numerical algorithms to allow understanding of the physics of the spatially and temporally complex system as a whole, based on a relatively complete mathematical description. An example would be the three-dimensional, time-dependent MHD modeling of the magnetosphere.

3. *Modeling that attempts to synthesize classes 1 and 2 by addressing complex coupled systems*. The physical systems that are of interest frequently do not admit a single global mathematical model. Instead, quite distinct physical processes, although coupled at some level, may describe different regions or physical constituents. This class of computation attempts to bridge interactions between different regions, different processes, and different scales using different numerical techniques and/or mathematical formulations. The goal is comprehensiveness and the development of an understanding of the complexity of the system. The modeling may include the coupling of different numerical techniques and codes, each of which addresses a different plasma process and/or region. Examples include global magnetosphere-ionosphere-thermosphere models, solar wind–CME models, and the interaction of the solar wind with the local interstellar medium. Currently, because of limitations in computing capabilities and resources, very complicated boundary or initial conditions or the long-time evolution of a global system cannot be included in this third class of modeling.

Numerical modeling of solar and space plasmas, along with theory, drives the synthesis of observations (and aids active experiments in space physics, albeit to a lesser extent, owing to the rarity of such experiments). Computer modeling aids in the interpretation of data, in particular where the data coverage is sparse and where space-time ambiguities exist. Furthermore, numerical modeling has become a valuable tool for mission planning. Examples include the development of orbital strategies in magnetospheric missions and plasma measurement requirements for Magnetosphere Multiscale. Finally, the panel observes that class 3 modeling forms the centerpiece of operational space weather forecasting. Currently, only a few models are used, but if meteorology is any indication, the number of models and their scope will increase dramatically in the next decade.

KEY ISSUES FOR THE NEXT DECADE

Continued advances in modeling will depend on the continued increase of available computational power. Moore's law is expected to hold over the next 10 years. Thus, CPU power will increase roughly 100 times (50 Gflops for the desktop, 50 Tflops for large systems), as will memory and disk storage. However, that power will have to be available to researchers in the field. This will require the availability of dedicated large-scale computers for grand challenge problems, as well as Beowulf-class computers for individual researchers and groups.

Computer hardware should be treated like hardware for experimentation, and sufficient funding should be made available. NSF and NASA computing centers are insufficient. The two NFS centers—the San Diego Supercomputer Center and the National Center for Supercomputing Applications—serve most of the supercomputing needs of the nation's academic research communities and are notoriously oversubscribed, typically by a factor of between 2 and 5. Consequently the turnaround time for codes at these centers is very long, and scientific progress is severely impeded. While the NSF centers will likely remain the resource of choice for the most demanding computations, smaller installations that are available to a research group exclusively are significantly more cost-effective and can provide for most of the more typical computing needs.

Dedicated funding lines for model development will also be needed. Current NASA/High Performance Computing and Communications (HPCC) and NSF/Information Technology Research (ITR) programs focus too closely on computational issues at the expense of basic science questions. The panel sees a danger in these programs becoming excessively computer science oriented. Certainly, issues such as version control, validation, and interoperability are of importance, especially in the context of the increasingly sophisticated codes that will be used by multiple users or groups, but it is essential that the focus remain on basic science and the computational tools needed to advance this science. The development of large scientific codes, rather like the development of experimental hardware, can require a lengthy gestation period. Most NASA, NSF, and DOE programs generally support the science that is done with large computational models but do not typically support the development of such codes. Increased funding for model development from group-size to principal investigator (PI)-size grants is essential.

There are currently very few dedicated funding lines for model development. For example, the Multidisciplinary University Research Initiative (MURI) program, which funds many projects, is not likely to be extended. Because funding for model development is scarce, most requests are hidden in proposals whose objective is a particular scientific objective. For example, NASA SR&T and Sun-Earth Connections Theory Program (SECTP) grants and NSF GEM and Space Weather grants are usually awarded for the investigation of specific scientific issues, yet often some fraction of these grants is used for model development. However, these relatively small grants lead to fragmented model development efforts when much more focused development efforts are needed. This concealment of model development costs makes it virtually impossible to assess how much funding is actually available for model development. However, since the primary objectives of such grants are answers to science questions, the fraction of the funding devoted to model development is small, probably at most 10 or 20 percent. Thus, there is a clear need to increase funding for model development from group-size to PI-size grants.

The modeling community will also have to become more open-source oriented. Codes should often be made available to the scientific community, but this requires clean coding, portability of codes, and documentation, aspects that are not generally addressed with scientific codes. Furthermore, there is little incentive for code developers to address these issues because the tasks are work-intensive and little or no funding is provided to document and disseminate codes where appropriate.

Data assimilation techniques, long considered essential in atmospheric and oceanic modeling, are only beginning to appear in solar and space plasma modeling. Naturally, data assimilation is appropriate for class 2 computer models. Assimilative models will be essential to ingest and analyze data from forthcoming multisatellite missions. Dedicated funding for data assimilation model development is required. As the panel describes below, such funding can be tied to the appropriate missions by funding, for example, model development during the B, C, and D mission phases. Class 3 computations need to continue to couple different physical models to properly address the complexity of the topical space plasma systems. Examples are solar dynamo-photosphere-chromosphere-corona-solar wind models and geospace models that include the ring current, radiation belts, and the plasmasphere.

Class 2 simulations have to be made more realistic by ensuring that parameters are as true to actual conditions as possible. For example, kinetic simulations have to achieve realistic ion-electron mass ratios.

In the long run there will be a confluence of class 2 and class 3 models. By way of example, fully three-dimensional kinetic hybrid models embedded in three-dimensional MHD models will be used to investigate Earth's magnetosphere within the next one or two decades.

4.3 THE ROLE OF DATA EXPLORATION

Projects supported by many U.S. government agencies have returned an unparalleled wealth of in situ and remote solar and space physics observations. Taken together, these data sets form the basis for innumerable scientific studies designed to characterize Earth's upper atmosphere and plasmas throughout the solar system.

Real-time and archived observations are in constant use to identify new phenomena and test the predictions of theory and simulation. Researchers use observations at ever-increasing temporal and spatial resolutions to identify the fundamental microscale physical processes that govern entire systems, and multipoint and multi-instrument observations to determine the relationships linking solar phenomena to processes throughout the heliosphere, magnetosphere, and ionosphere. Systematic analyses of archived data sets characterize variations in Earth's natural environment over periods both long and short (compared with the 11-year solar cycle) and identify the effects of manmade perturbations, including nuclear tests, ionospheric heating by ground radars, and plasma releases into Earth's atmosphere and magnetosphere.

Solar and space physics observations also represent a national resource of immediate practical value to the United States. Both NOAA and the DOD currently use real-time observations of the Sun, the solar wind, Earth's magnetosphere, and Earth's ionosphere to predict the disruptive effects of space weather on ionospheric communication, aircraft navigation, radar defense, geosynchronous satellite safety, and commercial electrical power supply. DOE monitors the geosynchronous radiation environment to test adherence to nuclear test ban treaties.

Data analysis has a number of other benefits. First, it defines the technology required for future instruments and spacecraft. Second, solar and space physics data sets are eminently suitable for training students at both undergraduate and graduate levels in scientific methodology. Third, certain data sets are intrinsically fascinating to the general public, in particular images of the Sun and the aurora. Finally, solar and space physics data sets are becoming increasingly large. The need for rapid exchange of similarly large data sets prompted the development of the Internet and the World Wide Web, and there is every reason to believe that this need will continue in the future.

ACCOMPLISHMENTS OF THE PAST DECADE

Progress in solar and space physics often requires the correlative analysis of multiple data sets. The past decade saw exciting improvements in our ability to access data sets from different instruments, ground stations, and spacecraft. With the advent of the Internet, each federal agency established or designated an online archive (see Table 4.1) from which users could download data sets, view prepared survey plots, and interactively generate plots according to their own specifications.

Cost constraints and a need for simplicity require that the designated archives limit the data sets and the options for their presentation. However, cutting-edge research often requires innovative analysis of higher-resolution observations using special software. In general, this software is best developed by members of the principal investigator's research team, who are most familiar with both the instrument characteristics and user needs.

TABLE 4.1 Designated Data Repositories

Agency	Data Center	Web Address
NASA	NSSDC	nssdc.gsfc.nasa.gov
NSF	NCAR	cedarweb.hao.ucar.edu (for upper atmospheric data)
NSF	UMD	www.polar.umd.edu (for Antarctic data)
NSF	NSO	www.nso.noao.edu/diglib/ (for solar data)
NOAA	NGDC	www.ngdc.noaa.gov
DOD	NGDC	www.ngdc.noaa.gov
DOE	LANL	Leadbelly.lanl.gov
USGS	NGDC	www.ngdc.noaa.gov

NOTE: NSSDC, National Space Science Data Center; NCAR, National Center for Atmospheric Research; UMD, University of Maryland; NSO, National Solar Observatory; NGDC, National Geophysical Data Center; LANL, Los Alamos National Laboratory.

A number of Web sites appeared during the past decade that allow remote researchers to interactively plot and even analyze high-resolution observations from individual experiments. Several programs also took up the challenge of providing real-time observations in a manner suitable for both scientists and the general public, including the Solar and Heliospheric Observatory (SOHO), the Transition Region and Coronal Explorer (TRACE), and Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI), which provide solar images; Polar, which gives auroral images; ACE, which observes solar wind plasma and magnetic field; geostationary operational environmental satellites (GOES), which observe geosynchronous magnetic fields and energetic particles; and Super Dual Auroral Radar Network (SuperDARN), which makes ionospheric radar observations.

PROBLEMS AND SOLUTIONS

Almost a decade ago, NASA and the research community convened an open workshop at Rice University to discuss problems hindering the widespread availability of solar and space physics data sets and to identify possible solutions.[1] Despite the progress noted above, many of the same problems confront researchers today.

Before data analysis can begin, researchers must first locate the data sets they wish to analyze. During the 1990s, efforts were made to develop a community-maintained catalog of online data sets. However, these efforts lapsed in the absence of any agency sponsorship. Consequently, the only available catalogs are those maintained by individual archives for their own data set holdings. Even when users know that the data sets they require exist, there are few pointers to where they might be obtained. The various federal agencies should work together to maintain one or more comprehensive catalogs pointing to all relevant data sets and documentation.

Many valuable data sets remain offline. Because multiple data requests impose a considerable burden on researchers at the locations where the data sets are stored, these data sets can be virtually inaccessible.

---

[1] NASA, Office of Space Science and Applications, Space Physics Division. 1993. *Concept Document on NASA's Space Physics Data System.* NASA, Washington, D.C. See also <http://spds.gsfc.nasa.gov/wkshprpt.html>.

When held solely within the community, offline data sets are in danger of being lost forever.[2] As the costs of storage space and high-speed communication continue to decrease, it is in the interests of the principal investigators themselves to place the data sets online in response to user requests and the need for data centers to permanently archive the data sets. Program officials can speed this process by requiring proper data management plans in new proposals and by monitoring the availability of data sets when selecting new projects and considering their continued funding or renewal. Even more importantly, program officials should reward public service with enhanced funding.

Even when data sets are publicly available, they may be unusable due to their complexity. A number of recent positive developments indicate ways in which this problem can be solved. Several research teams have begun making their own software available over the Internet. Particularly praiseworthy are efforts to enable outside users to access and portray complex high-resolution observations over the Web by running software remotely at the provider's institute. Efforts to develop platform-independent analysis and graphics tools for researchers inside and outside the principal investigator's team deserve to be given high priority by program managers. Since it can be costly to maintain software solely for external users, the best solutions provide the software used by the investigator team to external users.

The data sets needed for correlative studies are frequently provided from multiple sources in different formats. Interactive tools that can draw upon these sources to portray multiple data sets and software that can translate differing formats are urgently needed. While it is heartening to note that efforts to remedy these problems are under way at a number of institutions, it is also clear that there is considerable duplication of effort. To minimize costs, the research community needs to establish a software library in conjunction with archival centers. In particular, it would be desirable to initiate supervised online software trees like the solar community's Solarsoft to encourage software sharing and compatibility.

Finally, the rapid advances in numerical simulations described above imply that we will soon be making real-time space weather forecasts. Before becoming operational, the forecast models must be validated by comparing their predictions with in situ and remote observations. Any future data system must incorporate and support the results of the numerical simulations in a format and manner that facilitate this comparison.

AVAILABLE RESOURCES

At least two programs currently provide resources to improve the solar and space physics data environment: NSF's ITR program and NASA's Applied Information Systems Research Program (AISRP). The former supports research in systems design, information management, applications in science, and scalability. As this program is heavily oversubscribed, it has not had a significant impact on theory, modeling, and data exploration. By contrast, NASA's AISRP specifically targets the space sciences and is responsible for many of the success stories discussed in Boxes 4.1 to 4.10. Its funds have been used to restore valuable data sets and place them online and to develop cataloging, depiction, and format exchange tools. Particularly praiseworthy is the fact that the funds have often been used for projects that cross agency boundaries.

The recently inaugurated Virtual Solar Observatory (VSO)[3] illustrates well how the diverse data sets held within the solar and space physics discipline can be integrated and studied in a coherent manner. This initiative responds to a recommendation from the National Research Council's Committee on Ground-based Solar Research,[4] which called for NSF and NASA to collaborate in developing a distributed data archive with access through the Web by establishing a scalable environment for searching, integrating, and analyzing databases distributed over the Internet. Given the increasing emphasis on correlative space weather studies that seek to determine the terrestrial effects of solar disturbances, now is an appropriate time to expand this effort by establishing a data system that encompasses the full range of solar and space physics phenomena. In Section 4.4 the panel describes plans for a virtual, distributed space physics information system.

---

[2] NRC. 2002. *Assessment of the Usefulness and Availability of NASA's Earth and Space Science Mission Data*. National Academy Press, Washington, D.C.

[3] See <http://www.nso.noao.edu/vso/>.

[4] See Recommendation 7 on page 46 in NRC, Space Studies Board, 1998, *Ground-based Solar Research: An Assessment and Strategy for the Future,* National Academy Press, Washington, D.C.

BOX 4.1 GLOBAL MAGNETOSPHERE-IONOSPHERE-THERMOSPHERE MODELING

Over the past decade great progress has been made in modeling geospace—that is, the global magnetosphere-ionosphere-thermosphere system and its interaction with the solar wind (SW) and the interplanetary magnetic field (IMF). While the construction of cartoons depicting the global magnetosphere-ionosphere-thermosphere system was essentially state of the art a decade ago, today's most sophisticated models are run with measured SW/IMF data as input and compared extensively with in situ observations. Figure 4.1.1 shows results from the University of California at Los Angeles/NOAA geospace model for the Bastille Day geomagnetic storm in 2000 (July 14/15). The color rendering in the upper part of the figure shows the magnetosphere during the storm's main phase as it is compressed by the high solar wind dynamic pressure and eroded by the strong southward IMF B_z that is the hallmark of this storm. All three operational geosynchronous GOES satellites (marked by small red spheres in the figure) crossed into the magnetosheath at this time, as predicted correctly by the model. The lower part of the figure shows a comparison between the predicted (red) and observed (black) auroral upper (AU) and auroral lower (AL) indices of geomagnetic ground disturbance. The model is able to predict some of the intensifications and their magnitudes; however, it misses others, and clearly needs improvement. Global geospace models are now quantitative, and the new challenges are data assimilation, improving the physical realism, and metrics evaluations.

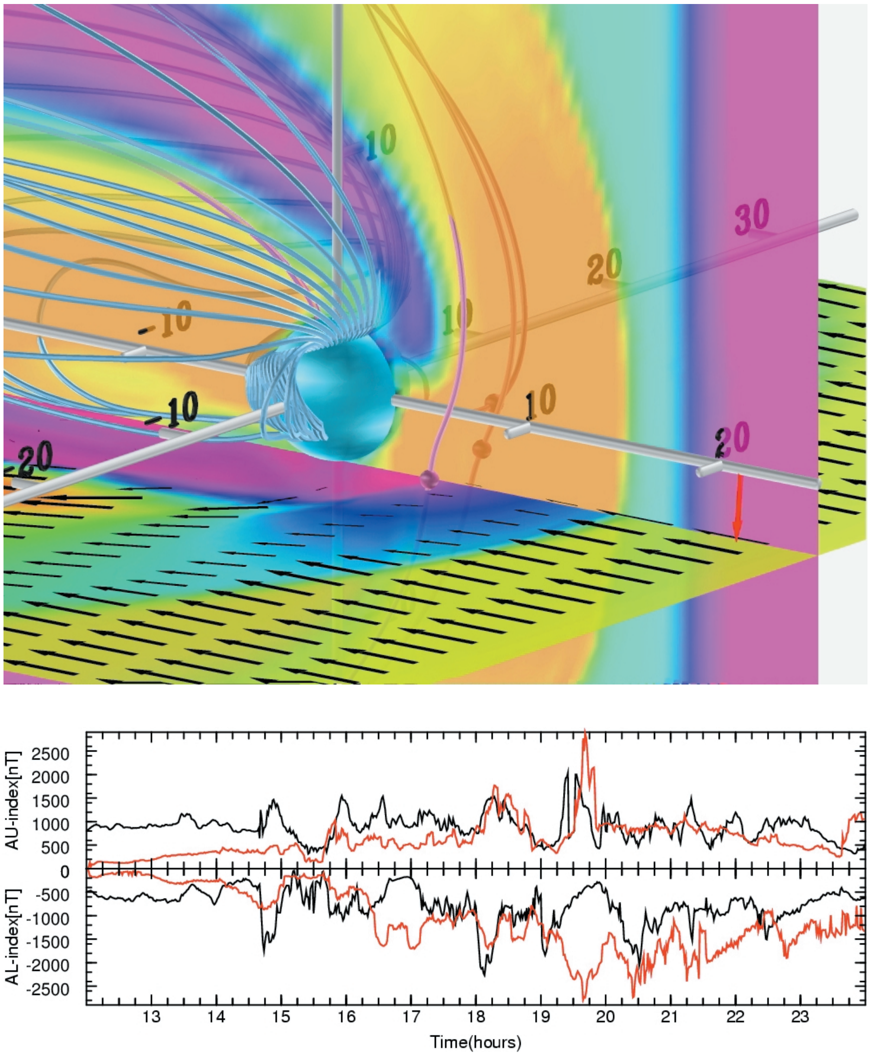


FIGURE 4.1.1 SOURCE: J. Raeder, Y.L. Wang, T.J. Fuller-Rowell, and H.J. Singer. 2001. Global simulation of magnetospheric space weather effects of the Bastille Day storm. *Solar Physics* 204: 325–338.

BOX 4.2 FROM HYDROGEN WALL TO STELLAR WINDS

The physics of the outer heliosphere and its interaction with the partially ionized local interstellar medium (LISM) has become an exciting growth area in space physics. An important recent development in outer-heliospheric research is the prediction and discovery of the "hydrogen wall." Theoretical models predict that the partially ionized LISM and the solar wind are separated by a complex set of plasma and neutral-atom boundaries of enormous scale, located between ~90 and ~250 AU from the Sun. A specific theoretical prediction by the models is that a wall of interstellar neutral hydrogen should exist upstream owing to the relative motion of the heliosphere and the LISM. This hydrogen wall is predicted to have a number density slightly more than twice the interstellar density, to be hotter than interstellar hydrogen, and to be some 100 AU wide. The physical reason for the hydrogen wall is the deceleration and diversion of the interstellar plasma flow about the heliosphere leading, through charge-exchange coupling, to a pile-up and heating of interstellar neutral hydrogen (Figure 4.2.1). The result is the formation of a giant wall, which acts to filter neutral hydrogen as it enters the heliosphere.

Confirmation of the hydrogen wall's existence was not expected for decades, but a serendipitous convergence of predictive theoretical modeling, observations to place limits on the cosmological deuterium/hydrogen ratio, and a multidisciplinary investigation spanning space physics and astrophysics led to the detection of the hydrogen! This is the first of the boundaries separating the solar wind and the LISM to be discovered, and it offers a glimpse into the global structure of the three-dimensional heliosphere. Radio emission from the heliopause, probably driven by global interplanetary shock waves, also provides an opportunity to probe deeply into the remotest reaches of the heliosphere. An unexpected astrophysics result to emerge from the recent work on the hydrogen wall is the first measurement of stellar winds from solarlike stars.

The research leading to the discovery of the hydrogen wall is an excellent example of theory driving the frontiers of space science and motivating the development of new observational techniques and methodology to complement traditional space physics tools.

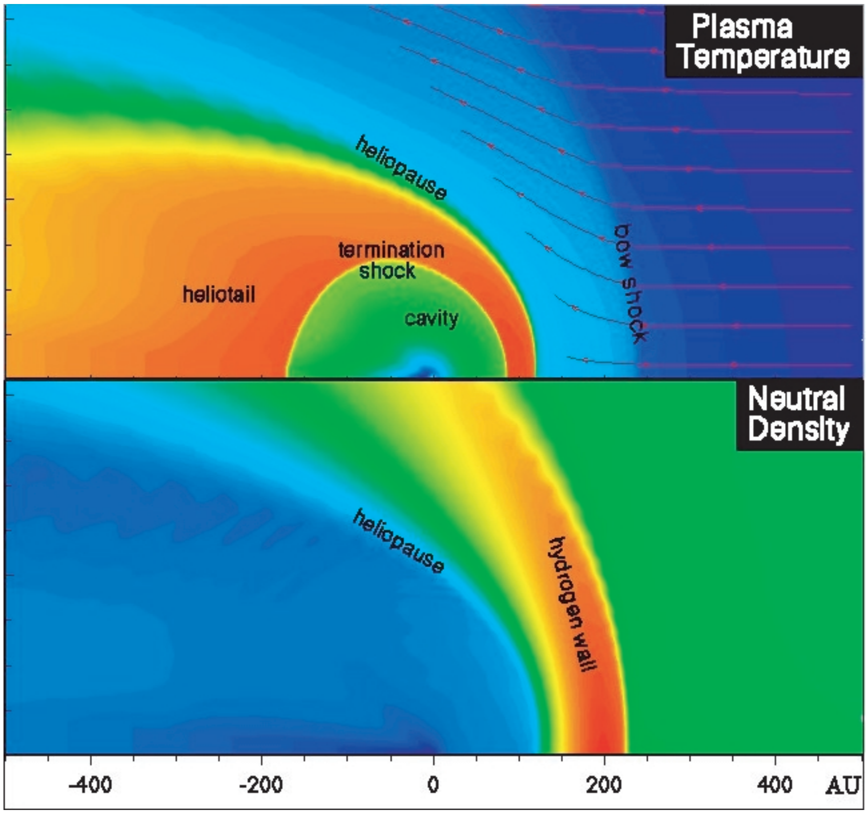


FIGURE 4.2.1 SOURCE: G.P. Zank. 1999. Interaction of the solar wind with the local interstellar medium: A theoretical perspective. *Space Science Reviews* 89: 413–688.

BOX 4.3 HOW ARE NEW RADIATION BELTS CREATED?

The unexpected creation of new electron and proton radiation belts during the March 24, 1991, geomagnetic storm challenged scientists to rethink the formation and stability of Earth's radiation belts. This storm was initiated by an interplanetary shock traveling at a remarkable 1,400 km/s. Fortuitous measurements by the Combined Radiation and Release Satellite showed that the belts appeared within minutes near the "slot region" at 2.5 R_E as the shock rapidly compressed Earth's magnetosphere (the largest compression on record). The newly formed belts persisted until 1994.

How did the belts form? The prevailing empirical diffusion models gave energization rates that were orders of magnitude too slow, and they only described the intensification of existing zones of MeV particles rather than the formation of new stable belts. Advances in theory identified a new process involving fast resonant acceleration by inductive electric fields accompanying a rapid compression of the geomagnetic field. This theory emphasized the need for new approaches to predicting Earth's dynamic radiation belts and to space weather forecasting.

One promising approach embeds a particle-pushing code for radiation belt particles in a global MHD simulation of the solar wind–magnetosphere interaction. Results for the 1991 storm (Figure 4.3.1) show that an initial outer zone electron source population, represented by the average NASA AE8MIN model in the upper right insert, was transported radially inward by the induction electric field to the sparsely populated slot region. This transport occurs on the MeV electron drift time of 1–2 min and produces a flux peak at 13 MeV for the 1991 event. A new proton belt was also formed at the same location by the same mechanism, but with greater energy (>20 MeV), trapping and transporting inward the extreme solar energetic proton source population produced by the interplanetary shock.

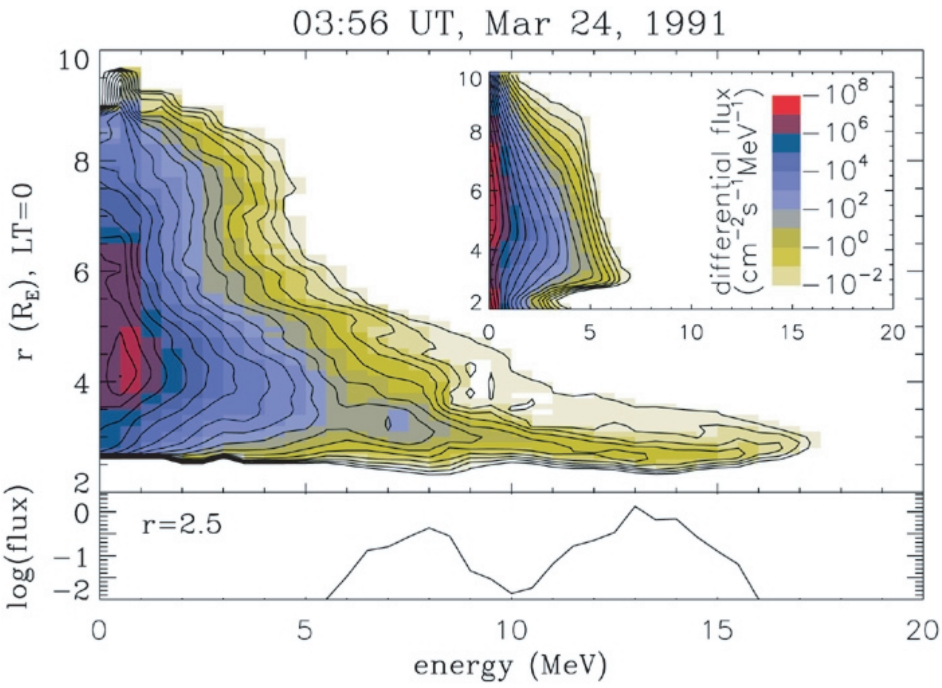


FIGURE 4.3.1 SOURCE: S.R. Elkington, M.K. Hudson, M.J. Wiltberger, and J.G. Lyon. 2002. MHD/particle simulation of radiation belt dynamics. *Journal of Atmospheric and Solar-Terrestrial Physics* 64(5–6): 607–615.

BOX 4.4 THE SOLAR DYNAMO PROBLEM

Dynamo theory has made great strides in recent years, exploring the variety of dynamo forms that might occur in planets, stars, and galaxies. Recent comprehensive massive simulations of the dynamo of Earth would appear to have that problem well in hand. On the other hand there is a serious problem with the dynamo of the Sun (and of other stars and of the galaxies). The extensive analytical and numerical models of the solar dynamo incorporate the notion of turbulent diffusion as an essential part of the dynamo. The concept of turbulent diffusion of the vector magnetic field is based on an intuitive analogy with the turbulent diffusion of a scalar field, such as a puff of smoke. The analogy provides just about the right diffusion rate to provide the 11- and 22-year magnetic cycle of the Sun, so it has been widely accepted. The difficulty is that the mean fields in the lower convective zone are estimated now to be 3,000 gauss or more, and such fields are as strong as the convection. Hence the field is too strong to be drawn out into the long thin filaments to which a puff of smoke would be subjected. So it is clear that we do not understand the diffusion that is essential for the solar dynamo. Combining this mystery with the mystery of the frail structure of the magnetic field, observed at the visible surface and inferred in the lower convective zone, we must be cautious in accepting the theoretical dynamo models. It is imperative that we address the problem of diffusion and the fibril structure of the field.

BOX 4.5 HELIOSEISMOLOGY

Helioseismology is the study of the structure and dynamics of the solar interior using the frequencies, phases, and travel times of resonant waves observable as oscillations at the photosphere. The evolution of the discipline is a striking example of the close and sometimes unpredictable interplay between theory and observation. Although the basic theory of nonradial stellar pulsations had been well developed in the early part of the 20th century to explain variable stars, it was never applied to quasi-static stars such as the Sun, and the first observations of solar oscillations in 1960 were in a sense fortuitous. Theorists had suggested evanescent acoustic waves propagating up through the atmosphere as a mechanism for the mysterious heating of the corona, and the unsuccessful search for such waves by looking for temporal variations in the Doppler shifts in Fraunhofer lines led to the discovery of surface oscillations in a band of frequencies around 3 mHz, lower than the frequencies of the coronal-heating waves sought. These so-called "5-minute oscillations" remained theoretically unexplained until 1970, when it was recognized that they could be the photospheric response to trapped interior waves of the acoustic-gravity mode class, a prediction that was confirmed by measurement of the wave dispersion relation. It was soon realized that the oscillations are part of an enormously rich spectrum of modes, now known to encompass wave numbers from 0 (purely radial) to well over 2,000 over the solar circumference, and ranging from purely surface modes to those penetrating to the center of the Sun. By properly decomposing and analyzing the observed spectrum, it became possible in principle to directly measure the depth dependence of sound speed, density, and bulk velocity in the interior of the Sun and to measure their departures from radial symmetry as well.

The last three decades have witnessed the emergence of the new discipline of helioseismology. Concurrent with the development of instruments and observing methods, a significant theoretical effort occurred to develop the complex analysis techniques required to find inversion methods with suitable resolution and precision, to enhance the resolution and physical complexity of the models of the structure and dynamics of the Sun and of its features, and to include wave phase information in the inferences. Finally, the necessity of analyzing many orders of magnitude more data than had hitherto been required has brought about major advances in the handling of data, and these have to some extent propagated throughout the field of solar physics.

The ability to directly determine physical conditions deep in the interior of a star with astonishing accuracy (the sound speed is now known to about one part in 10^4 through most of the solar interior) has importance well beyond the immediate fields of stellar structure and evolution. It allows us to use the solar interior for the exploration of physics of systems inaccessible to the laboratory. The verification of interior models to such precision meant that the solution of the problem of the deficit of solar neutrinos observed on Earth had to be sought in the realm of fundamental particle physics. The complex atomic physics required to calculate the radiative opacity, the equation of state, and the nuclear reaction rates is strongly constrained by the necessity of matching the observationally supported solar models. The description of turbulent convective flows and the magnetohydrodynamics of the near-surface regions of the Sun, crudely parameterized or ignored in structural models, are now subjects of active research supported by observational data. It will soon be possible to measure and analyze the internal oscillations of other stars in different states; a few have already been detected. This will truly allow stellar interiors to be used as a laboratory for physical exploration.

BOX 4.6 SUPERSONIC PLASMA BUBBLES IN THE UPPER ATMOSPHERE

Near Earth the geomagnetic field is a dipole whose axis is tilted by about 12 degrees from Earth's rotational axis. As a consequence, the geomagnetic field is horizontal at low latitudes in the upper atmosphere and ionosphere (altitudes of 100 to 2,000 km). During the day, upper atmospheric winds produce dynamo electric fields that cause plasma on these geomagnetic field lines to drift upward, which leads to an elevated ionosphere with peak densities of 10^6 cm^{-3} and peak altitudes as high as 600 km. At dusk, the ionosphere rotates into darkness, and in the absence of sunlight, the lower ionosphere rapidly decays. The result is that a steep vertical density gradient develops on the bottom side of the raised ionosphere. This produces the classical configuration for the Rayleigh-Taylor instability, in which a heavy fluid is situated above a light fluid (supported by the horizontal geomagnetic field). In this situation, a density perturbation can trigger the instability, and once triggered, the depleted densities at low altitudes bubble up through the raised ionosphere. The plasma density in the bubbles can be up to two orders of magnitude less than that in the surrounding medium, and the bubbles frequently drift upward with supersonic velocities. Typically, the bubbles develop into large structures. An example of a bubble structure is shown in Figure 4.6.1, which corresponds to a coherent backscatter measurement by the JULIA ground-based radar in South America. The bubbles can reach altitudes as high as 1,500 km, and the entire north-south extent of the magnetic field at low latitudes is usually depleted (30 degrees of latitude). The east-west extent of a disturbed region can be several thousand kilometers, with the horizontal distance between separate bubble domains being tens to hundreds of kilometers. Although plasma bubbles have been observed for more than 30 years, the exact trigger mechanism is still unknown, and the three-dimensional structure of the bubbles has not been modeled. The latter requires new models that rigorously incorporate the microphysics in large-scale models of the background ionosphere.

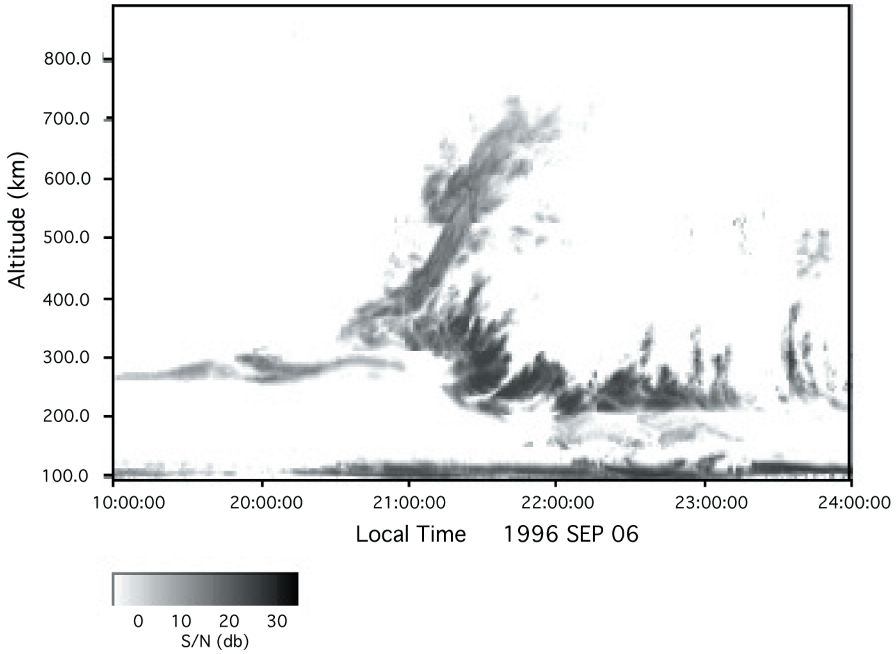


FIGURE 4.6.1 SOURCE: D.L. Hysell and J.D. Burcham. 1998. JULIA radar studies of equatorial spread *F*. *Journal of Geophysical Research* 103(A12): 29, 155.

BOX 4.7 DATA DISSEMINATION

No project required the coordinated analysis of multiple data sets more than NASA's International Solar-Terrestrial Physics program, which was predicated on the need to identify, describe, and explain the multitude of physical relationships linking processes on the Sun to those in the heliosphere, magnetosphere, and ionosphere. The Space Science Data Facility at NASA's Goddard Space Flight Center (GSFC) developed an Internet-based plotting and data downloading tool, CDAWeb, that allows interactive users to select key (low-resolution) parameters from multiple ground- and space-based instruments for plotting as lines or images versus time. When interesting intervals are identified, users can download the data sets seen in the plots in different formats, together with supporting documentation. Similar servers provide spacecraft trajectories and heliospheric and near-Earth solar wind observations.

CDAWeb provides over 1,300 parameters from 150 datasets and 25 missions. Usage has grown steadily with time and is now running at 6,000 plots and 2,000 data file listings per month. The server at GSFC has been mirrored by archive centers in Germany, the United Kingdom, and Japan. Figure 4.7.1 shows an example of multiple plot types. The plot was generated interactively over the Web by selecting the parameters, instruments, spacecraft, and time intervals for study.

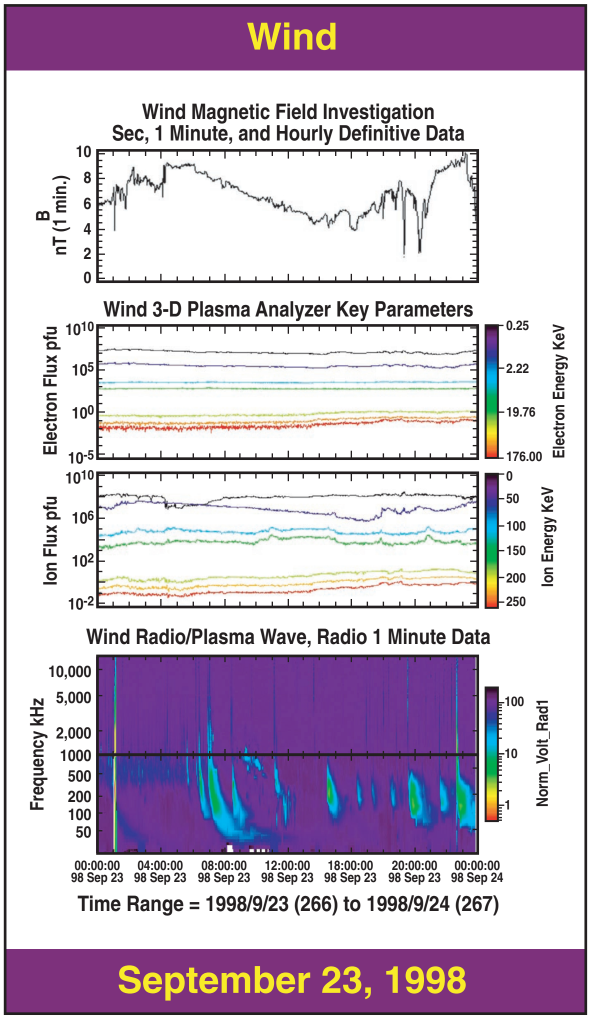


FIGURE 4.7.1 SOURCE: Plot by R.W. McGuire, Space Physics Data Facility at NASA Goddard Space Flight Center.

4.4 A COHERENT FRAMEWORK FOR THEORY, MODELING, AND DATA EXPLORATION

COUPLING COMPLEXITY IN SPACE PLASMA SYSTEMS

Space physics has typically been organized according to distinct or discrete events or physical processes, and recommendations for future directions have been expressed in the context of these very specific (and somewhat isolated and idealized) problems. A very influential example of this approach appears in the Colgate Report,[5] which identifies six problems[6] as vital to further understanding of space plasmas. Substantial progress has been made in solving the problems identified in the Colgate Report, but the basic questions remain with us.[7] Nevertheless, progress in the field has also served to crystallize the complexity and therefore the challenge of resolving many of the important problems of space physics. This complexity arises from the coupling across space and time scales (e.g., the generation of turbulence at boundary layers), the coupling of multiple constituents (e.g., the interaction of the solar wind with neutrals in the interstellar medium), and the linkage of different regions (e.g., the ejection of magnetized plasma from the surface of the Sun and its subsequent interaction with Earth's magnetosphere). The success stories given in Boxes 4.1 to 4.10 and the discussions in the text related to computation and data exploration illustrate these couplings, as do the examples in the following discussion.

Distinct plasma regions and regimes are invariably coupled in a highly nonlinear dynamical fashion, with the implication that each region or physical process cannot be considered in isolation. Multiple plasma regions can be coupled through events that transfer mass, momentum, and energy from one region to another, such as the eruption and subsequent propagation of a coronal mass ejection from the solar surface through interplanetary space to—perhaps—Earth's magnetosphere.

By contrast, a particular plasma region can admit multiple ion and atom populations, each governed by distinct plasma physical processes yet coupled to one another dynamically and self-consistently. An excellent example is the coupling of the solar wind to the local interstellar medium through the intermediary of neutral interstellar atoms (beyond some 10 to 15 AU, the dominant constituent of the heliosphere, by mass, is neutral interstellar H). Charge exchange serves to couple the plasma and neutral atom populations, yielding a highly nonequilibrated, nonlinear system in which the characteristics of both populations are strongly modified (pickup ions, anomalous cosmic rays, and very hot neutral atoms are some of the by-products created). Charge exchange is important, too, at solar system bodies such as Venus and Io, and of course energetic neutral atom imaging of the terrestrial magnetosphere has become an important new tool. The self-consistent coupling of disparate plasma regimes, each governed by possibly distinct plasma physical processes, is a challenge that must be addressed if the current fleet of satellites and ground-based observatories and ambitious new initiatives such as LWS, Solar Probe, and Interstellar Probe are to be successfully and fully exploited.

Furthermore, the space plasma environment typically possesses a multiplicity of spatial and temporal scales, and the nonlinear, dynamical, self-consistent feedback and coupling of all scales determines the evolution of the system through the creation of large- and small-scale structure. Excellent examples are reconnection and turbulence, where a marriage of large-scale, slow MHD behavior and fast, small-scale kinetic processes is needed to further our understanding of fundamental nonlinear processes that arise in space plasmas. Understanding multiscale feedback in space plasma systems will be one of the most important and challenging problems facing theorists and modelers over the next decade.

Different plasma regimes (and possibly even scales) are often separated by narrow boundaries, and the coupling and structure of large- and small-scale processes frequently control the evolution of the boundaries. The quintessential examples are collisionless shock waves and auroral arcs. The sharp gradients and the coupling of disparate scales in boundaries continue to challenge our understanding and modeling of space physics problems.

---

[5] National Academy of Sciences (NAS), Space Science Board. 1978. *Space Plasma Physics: The Study of Solar-System Plasmas,* Vol. 1. NAS, Washington, D.C.

[6] These were (1) magnetic field reconnection; (2) interaction of turbulence with magnetic fields; (3) the behavior of large-scale plasma flows and their interaction with each other and with magnetic and gravitational fields; (4) acceleration of energetic particles, and particle confinement and transport; and (5) collisionless shocks.

[7] See, for example, NRC, 1995, *A Science Strategy for Space Physics;* NRC, 1988, *Space Science in the Twenty-First Century: Solar and Space Physics;* and NRC, 1998, *Supporting Research and Data Analysis in NASA's Science Programs,* all from the National Academy Press, Washington, D.C.

BOX 4.8 FUTURE PROSPECTS: DATA ASSIMILATION

Data assimilation techniques were first used in numerical weather prediction, when meteorologists were confronted with having to solve an initial value problem without the right initial data. Specifically, there were insufficient synoptic observations to initiate a model run that could be used to predict the weather. To overcome this obstacle, the meteorologists developed a methodology that is now called data assimilation. This methodology uses data obtained at various places and times, in combination with a physics-based (numerical) forecast model, to provide an essentially time-continuous "movie" of the atmosphere in motion. During the last 40 years, meteorologists have continually improved their ability to predict the weather, both as a result of model improvements and because of a large infusion of new satellite and ground-based data. Following the example set by meteorologists, oceanographers began to use data assimilation techniques about 20 years ago. Recently, using a numerical model of the Pacific Ocean, in combination with a large number of distributed ocean measurements, oceanographers were able to successfully predict the coming of the last El Niño.

The solar and space physics community has been slow to implement data assimilation techniques, primarily because it lacks a sufficient number of measurements. However, this situation is changing rapidly, particularly in the ionosphere arena. It is anticipated that within 10 years, there will be several hundred thousand ionospheric measurements per day from a variety of sources, and these data will be available for assimilation into specification and forecast models. The data sources include in situ electron densities measured by NOAA and DOD operational satellites, bottomside electron density profiles from a network of 100 digisondes, line-of-sight total electron content (TEC) measurements between as many as 1,000 ground stations and Global Positioning System satellites, TECs between low-altitude satellites with radio beacons and specific ground-based chains of stations, TECs via occultations between various low-altitude satellites and between low- and high-altitude satellites, and line-of-sight optical emission data, which can provide information on plasma densities. Furthermore, the images of the plasmasphere and magnetospheric boundaries recently obtained by the IMAGE spacecraft can be used in data assimilation models. A magnetospheric constellation with more than 10 spacecraft would provide an invaluable dataset for data assimilation models.

There are numerous data assimilation techniques, but perhaps the one that has gained the most prominence is the Kalman filter. This filter provides an efficient means for assimilating different data types into a time-dependent, physics-based numerical model, taking into account the uncertainties in both the model and the data. Using a sequential least-squares procedure, the Kalman filter finds the best estimate of the state (ionosphere, neutral atmosphere, or magnetosphere) at time *t* based on all information prior to this time. Formally, the Kalman filter performs a recursive least-squares inversion of all of the measurements (TEC, in situ satellite, etc.) for the model variable (e.g., plasma density) using the physics in the model as a constraint. The net result is an improved estimate of the model variable; it has the least expected error, given the measurements, model, and error statistics.

Use of data assimilation techniques has recently been initiated in ionosphere-thermosphere studies, and it is clear that such techniques will have a major impact on the field during the next decade. With a physics-based model of the ionosphere assimilating hundreds of thousands of measurements per day, global ionospheric reconstructions will be available hour by hour throughout the year. With this information at hand, the scientific community should be able to resolve a host of long-standing basic science issues, and the operational community should have reliable ionospheric parameters for the various products. Eventually, a similar capability will be achieved for the study of Earth's upper atmosphere and magnetosphere, but perhaps not during the coming decade because of the lack of data. Because the development and testing of physics-based assimilation models are labor intensive, it is important that this effort begin now for other solar and space physics domains.

Finally, space plasmas, as illustrated in extraordinary detail by images of the Sun obtained by the spacecraft TRACE, can change their (magnetic field) configuration on extremely short time scales, with an accompanying explosive relaxation in the associated plasma. Like plasma boundaries, the rapid evolution of a plasma from one state to another continues to challenge theorists and modelers.

From the above discussion, a very natural classification or ordering of solar system plasma physics, and one that distances itself from the very regional- and event-based ordering of the past several decades, is (1) space plasma couplings across regions, (2) couplings across scales, (3) physics of boundaries, and (4) explosive relaxation of plasmas. In developing the high-priority theory and modeling challenges listed in Box 4.11, the panel focused on problems that span several categories. These are problems that are fundamental to the further development of space physics, that cut across the different panels of the Solar and Space Physics Survey Committee, and that have the potential to influence both astrophysics and laboratory plasma physics. Box 4.11 lists the space physics areas to which the panel gives highest priority; because the panel does not believe that the individual entries can admit a meaningful ranking, they are not listed in priority order. Improving our understanding of the problems identified in Box 4.11 will lead to major advances in the field of space physics over the next decade, and central to this advance will be the development of models and theories that embrace the highly nonlinear dynamical coupling and feedback of different and disparate scales, processes, and regions. As suggested early in this section, the notion of "coupling complexity" refers to a class of problems or systems that encompass significantly different scales, regions, or particle populations, the understanding of which requires more than one set of defining equations or concepts. As discussed above, for example, the heliosphere contains cosmic rays, the solar wind, neutral atoms, and pickup ions, each of which interacts with the others and requires its own set of equations and coupling terms. Similarly, the ionosphere-thermosphere and magnetosphere are different regions governed by distinct physical processes.

THE CHALLENGES OF COUPLING COMPLEXITY

To embrace the demands imposed by coupling complexity, as defined above, in resolving the problems listed in Box 4.11, theorists, modelers, and data analysts must address a number of challenges:

1. Formulation of sophisticated models that incorporate disparate scales, processes, and regions, the development of analytic theory, and the maintenance of a strong connection to basic science;
2. Computation;
3. Incorporation of coupling complexity into space physics models;
4. Integration of theory, modeling, and space- and ground-based observations;
5. Data exploration and assimilation; and
6. Transition of scientific models to operational status in, e.g., space weather activities.

Each of the above challenges requires some elaboration.

Challenge 1: Formulation of Sophisticated Models

In recognizing that multiple scales, regions, processes, and plasma populations are intrinsic to the challenging space physics problems of today, the correct

BOX 4.9 FUTURE PROSPECTS: TOMOGRAPHY

Tomography has been used extensively by the medical community for several decades, but it was not until about 1988 that this technique was first applied to the ionosphere. Ionospheric tomography is more difficult than medical tomography because the ionosphere varies with time, while a patient is generally motionless during a tomographic scan. Also, for ionospheric tomography, the scanning directions are limited. Nevertheless, to date, both radio and optical tomography have been used in ionospheric applications. With radio tomography, radio transmissions from a low-Earth-orbiting satellite (or satellites) are received along a chain of stations, with the stations typically distributed along a line. The signals received at the stations are used to measure the total electron content (TEC) along the ray paths. Each station records a large number of ray paths as the satellite traverses the station, with the pattern of ray paths taking the form of a partially opened fan. With multiple stations along a line, there are a large number of intersecting ray paths, and the associated TECs are inverted by a mathematical algorithm to obtain a two-dimensional reconstruction of the electron density as a function of altitude and distance along the chain of stations. Optical tomography works in a similar way, but instead of TECs integrated (line-of-sight) optical emissions are measured.

At the present time, tomography chains exist in the United States, South America, parts of Europe (including Scandinavia), Russia, and Asia, and these chains provide information about ionospheric weather in these local regions. Although tomography chains are relatively new, they already have been used very successfully in reconstructing several different ionospheric density features, including plasma troughs, auroral boundary blobs, traveling ionospheric disturbances, equatorial ionization crests, and equatorial plasma bubbles. During the next decade, tomography is anticipated to play an important role in elucidating ionospheric weather features.

mathematical and physical formulation is critical. In this, there is no substitute for time-honored analytical approaches to theoretical developments in plasma physics, fluid dynamics, and applied mathematics. Progress on highly nonlinear, coupled plasma problems may be made using techniques that range from the relatively standard to nonlinear, low-order reductive approaches and statistical methods. Current agency funding programs appear to be largely adequate (although subject to budgetary pressure) to support these basic theoretical efforts, although more innovation and bolder ideas, approaches, and techniques in proposed research should be encouraged and rewarded. *The panel cannot emphasize strongly enough that the program outlined in this report must maintain a strong and vigorous connection to basic space science*. Computation is no substitute for the development of rigorous theories and well-conceived models. Challenge 1 must be regarded as a critical component of the recommendations that are developed below.

Challenge 2: Computation

Many of the problems listed in Box 4.11 impose significant computational demands in terms of CPU power and the concomitant development of sophisticated and efficient codes. Two challenges face the community. The first is to further develop existing codes and algorithms, such as three-dimensional MHD codes that incorporate adaptive mesh refinement—for example, three-dimensional hybrid codes with improved electron/ion mass ratios or improved codes for data exploration. These problems do not demand the inclusion of new physics to handle coupling complexity but demand instead substantial progress in current research areas. The second challenge lies in developing and implementing new computational approaches for both model solving/simulation and data exploration that exploit advances made by numerical mathematicians, statisticians, and computer scientists. To meet both challenges, the panel strongly encourages funding agencies to augment the grants of modelers for a limited period to allow them to make their software public and to provide adequate documentation and support. At the same time, the panel warns against turning the open-source (or "community code") concept into an exercise in computer science with an excessive emphasis on standards, version control, interoperability, etc. Open-source codes should remain scientific codes first and foremost.

Challenge 3: Incorporation of Coupling Complexity into Space Physics Models

The coupling of different physical processes, scales, and regimes gives rise to the relatively new science of coupling complexity in space plasma systems. Obviously space physics has always attempted to incorporate as many physical processes as possible into a particular model. However, only in the last few years have our understanding of the basic underlying physical processes (through the acquisition of data and advances in theory, among others) and our access to powerful computers been sufficient to allow space physicists to make a reasonable effort to explore and model systems as opposed to processes.

The self-consistent incorporation of multiple scales, physical processes, and distinct regions into models will be the main challenge to theorists and modelers in the coming decade, demanding the formulation and development of sophisticated models and theory, the development of new and innovative algorithms, access to sophisticated computational resources, and the opportunity to test model predictions and validate theories against existing and future observations. The panel anticipates that models and theories that address coupling complexity will demand sophisticated new measurements, which will in turn drive and define new space- and ground-based missions (in situ, multipoint, remote, etc.) (Box 4.12).

The panel advocates both the synergistic investigation of well-chosen, isolated theoretical problems and the development of coupled global models. For major advances to be made in the space physics science of coupling complexity, fundamental theoretical analysis, sophisticated computational tools, and state-of-the-art data analysis must all be coupled intimately under a single umbrella program. Theoreticians working with pen and paper, computational space physicists, and data analysts will be needed collectively to achieve the major advances expected of space physics. Only by creating and maintaining major groups of this sort can a strong and vital connection between basic science, computation, and observations be achieved.

What is needed to develop a research program that addresses coupling complexity are the following:

- Time—that is, long-term stable funding;
- Synergistically interacting groups of students, postdoctoral associates, research scientists, and several university or institutional faculty who are able to inte-

BOX 4.10 FROM MODEL TO FORECASTING

The Magnetospheric Specification Model (MSM), as currently implemented by the NOAA Space Environment Center, computes fluxes of energetic electrons and ions in real time based on an estimated index of geomagnetic activity, Kp. The model provides a rough picture of present space environment conditions, in particular near geosynchronous orbit. The representation is imprecise, partly because particles are injected into geosynchronous orbit by substorms, which are not yet well understood, and partly because driving the model with the 3-hour Kp index causes it to miss short-term variability. Figure 4.10.1 shows an example of routine MSM output in the equatorial plane, which also exemplifies the high spatial variability of particle fluxes in the inner magnetosphere.

The MSM was developed at Rice University and is based on the research-grade RCM (Rice Convection Model), which was greatly simplified for operational use. Numerical solutions of some differential equations were replaced by simpler, observation-driven, semiempirical algorithms. It is noteworthy that the development of the MSM, exclusive of its scientific basis, took about 7 years and substantial resources.

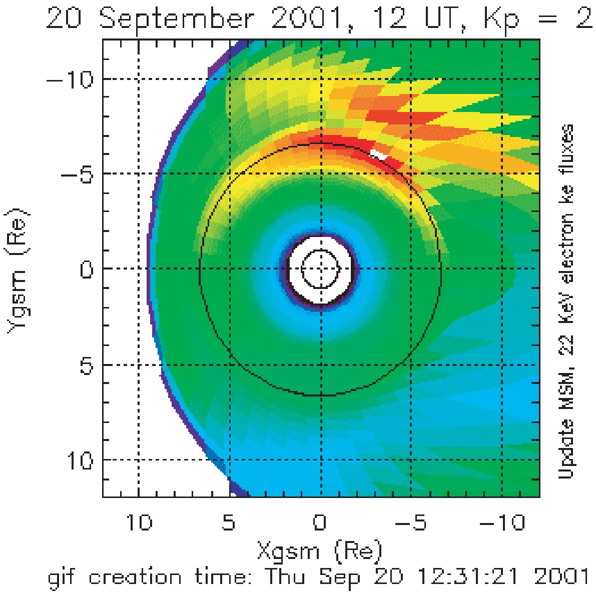


FIGURE 4.10.1 SOURCE: Unpublished figure courtesy of Richard Wolf, Rice University.

grate diverse scientific, theoretical, and computational viewpoints and to bring the complementary tools of theoretical and analytic techniques, computational physics, and data analysis to a major research problem;

- Sufficient computing and institutional resources (say, the ability to purchase and administer a Beowulf system or the ability to support computational physicists, theoretical physicists, and data analysts under a single umbrella program);
- The opportunity to develop highly sophisticated codes and models to extend and validate theories and models and to make predictions. These will eventually be transitioned to the community and will require resources for software development and support; and
- A commitment from the home institution (university, laboratory, industry) to develop a stable, long-term program in space physics by creating permanent positions. This would provide an intellectual environment in which large research efforts can flourish and would allow for critical-mass efforts.

Are the existing resources and programs adequate? The panel has carefully examined existing research programs and avenues supported by NASA, NSF, NOAA, DOE, DOD, the Air Force, the Navy, and the Army and concludes that the resources for a full-scale coupling complexity research program are inadequate or nonexistent. This conclusion resulted from the panel's examination of the existing research programs at the various agencies and from direct meetings with and presentations by NASA and NSF administrators. Programs such as the SECTP, MURI, the Supporting Research and Technology (SR&T) program, the HPCC program, ITR (an NSF program), and the CEDAR, GEM, and SHINE programs do not meet the five needs listed above and have goals that do not resonate well with a coupling complexity program. A clear distinction can be made between the recommended Coupling Complexity Initiative and the current SECTP. To simply enhance the existing SECTP is infeasible since the Coupling Complexity Initiative (1) is aimed primarily at systems throughout the heliosphere rather than isolated physical processes and does not emphasize the Sun-Earth connection; (2) would bring large numbers of theorists, computational modelers, and data analysts together in a synergistic fashion and let them work on different aspects of a complex physical system; (3) is intended to foster the growth of space physics at universities and institutes;

and (4) would have grants that are very much larger and of longer duration than those of the SECTP. Clearly, the intent of the Coupling Complexity Initiative is quite different from the goals of the SECTP. Programs such as SECTP are simply of insufficient duration, size, and scope and do not meet many of the needs identified here. Other programs, such as CEDAR/GEM/SHINE and SR&T, are excellent but clearly of very limited scope.

The challenges we will face in the coming decade from new and complicated missions (clusters, constellations, LWS), the piecing together of disparate data sets and results from space- and ground-based observations, and achieving the LWS goal of synthesizing and understanding the global coupled Sun-Earth system demand a far more sophisticated approach to theory and modeling that addresses coupling, nonlinearity, and multiscale, multiprocess, and multiregional feedback. Similarly, extraordinarily ambitious new missions such as Solar Probe and Interstellar Probe demand innovative theory and modeling since these missions will be unique and among the great scientific enterprises of the new century. Furthermore, coupling complexity will be of great importance to other fields, such as astrophysics and laboratory physics.

Recommendation 1. NASA should take the lead in creating a new research program—the Coupling Complexity Research Initiative—to address multiprocess coupling, nonlinearity, and multiscale and multiregional feedback in space physics. The research program should be peer reviewed. It should do the following:

- **Provide long-term, stable funding for a 5-year period.**
- **Provide sufficiently large grants that critical-mass-sized groups of students, postdoctoral associates, and research scientists, gathered around university and institutional faculty, can be supported.**
- **Provide funding to support adequate computational resources and infrastructure for the successfully funded groups.**
- **Facilitate the development and delivery of community-based models.**
- **Use the grants to leverage faculty and permanent positions and funding from home institutions such as universities, laboratories, institutes, and industry.**

The scope of a successful coupling complexity grant must recognize that a combination of remote sensing, in situ measurements, and advanced modeling will pro-

BOX 4.11 OUTSTANDING QUESTIONS

Examples of high-priority space science themes that are associated with coupling complexity are numerous and embrace many of the most significant challenges in space physics. The following have been identified as excellent examples of coupling complexity in that they are fundamental to the further development of space physics, they cut across the different areas addressed by the five panels of the Solar and Space Physics Survey Committee, and they have the potential to influence both astrophysics and laboratory plasma physics. The examples below all span three or more of the four classes that introduced the notion of coupling complexity (see p. 201).

Examples within solar physics are (1) coronal heating, which clearly requires that we address the coupling of physical processes across regions and scales as well as incorporate the explosive relaxation of plasmas; (2) coronal mass ejections and flares, which demand that all four classes be addressed; (3) the dynamo problem, which remains one of the major outstanding problems in solar physics and for which classes 1, 2, and 3 are all factors; and (4) solar variability (classes 1, 2, and 3), which is of both scientific and economic importance.

Examples within heliospheric physics are (1) the acceleration of the solar wind and the polar wind (classes 1, 2, and 3), both of which are major outstanding theoretical problems; (2) the interaction of the solar wind with the local interstellar medium (again, classes 1, 2, and 3), which is becoming a topic of increasing importance with fundamental implications for astrophysics; (3) turbulence in the interplanetary medium (classes 1 to 3), which remains as a great classical problem; and (4) transport phenomena (all four classes) for both particles and fields, which is another problem of classical origin.

Examples related to the interaction of the solar wind with planets are (1) the physics of planetary ionosphere-magnetosphere mass exchange; (2) magnetic storms; (3) substorms; and, of course (4) climate variability due to solar influences. All four examples here span all four classes.

Some examples are so broad ranging that they are of importance to solar, heliospheric, magnetospheric, and ionospheric physics. Examples include (1) current layers, boundaries, and shock waves (classes 1 to 4); (2) particle acceleration (also classes 1 to 4); (3) turbulence; and (4) changes in magnetic field topologies and plasma configurations.

BOX 4.12 EXAMPLE OF COUPLING COMPLEXITY

The ideas expressed in Challenge 3 are illustrated here by an example that has received recent and extensive discussion within the magnetospheric community. It has been recognized in the past few years that magnetohydrodynamics (MHD) cannot adequately model reconnection, so that implementing adaptive mesh refinement or related techniques to finer and finer scales is insufficient to clarify the physical processes. One approach that has been advocated is to develop methods for embedding kinetic models locally within larger-scale MHD simulations. At small scales, the MHD description fails and a new physics model (multifluid or kinetic) must be incorporated. As a preliminary approach, one can envisage using a hybrid model in a predetermined region of space in an MHD simulation. Eventually, the approach should be adaptive, so that as small scales develop in the MHD model, the grid refinement is coupled to the implementation of a higher-level physical model. This represents one aspect (global self-consistent computer simulations) of a program designed to further our understanding of reconnection. However, to completely elucidate this challenging problem will also require complementary studies that address theoretically related fundamental questions using, for example, analytical and other techniques and that might address aspects of reconnection that cannot be handled adequately within large-scale simulations. Similar complementary data studies are needed to both guide and refine the theoretical and computational studies.

vide the major advances in space physics over the next several decades. Successful grants will address problems by developing models and theories that define and drive missions, and the models will demand that observations and measurements reach new levels of sophistication, possibly through the use of new techniques.

The models and algorithms developed through the Coupling Complexity Initiative will make invaluable contributions to future NASA, NSF, and NOAA activities, which are expected to focus on remote sensing and multipoint measurements. Because processes acting within and on the various regions in the heliosphere occur at different locations and times, global models are needed to relate past events that occur in adjacent regions to the in situ measurements made along spacecraft trajectories and to remote-sensing scan planes. Also, since the processes operate over widely different spatial and temporal scales, models are needed to integrate the effects so that they can be related to line-of-sight (column-integrated) ground-based or spacecraft measurements. In addition, coupled global models are useful for providing a framework in which large multispacecraft, multisite data sets can be organized and interpreted. Finally, hybrid models that rigorously include both macroscopic and microscopic physical processes are needed to relate measurements obtained from multispacecraft missions that focus on boundary layers, so that important acceleration processes can be elucidated.

A fundamental component of this recommendation is that the award of a grant will carry with it the expectation of a commitment from the home institution (university, laboratory, industry) to develop a stable, long-term program in space physics by the creation of permanent positions; this provides an intellectual environment in which large, critical-mass research efforts can develop and flourish.

Since nearly 30 groups submitted proposals to the most recent NASA SECTP, the panel used this as a measure of the potential large groups that exist currently in this country. A healthy space physics program should support about a third of these, in the estimation of the panel. Accordingly, it is recommended that the Coupling Complexity Initiative should support 10 groups, each at a funding level of between \$500,000 to \$1 million per year. This would require the commitment of \$7.5 million to \$10 million per year in funding. Many of the goals of the LWS program resonate very well with the Coupling Complexity Initiative, which might be its natural home within NASA. NSF has expressed considerable interest in the broad field of complexity, and the panel anticipates that the Coupling Complexity Initiative might also fit well into these programs. Similarly, DOE, NOAA, and possibly even the DOD have a fundamental interest in coupling complexity. Cross-agency programs have proved successful in the past. The panel recommends the formation of a cross-agency commission, with NASA possibly taking the lead though its LWS program, to examine the implementation of a cross-agency Coupling Complexity Initiative.

Challenge 4: Integration of Theory, Modeling, and Observations

Theoretical space physics develops models, which can be tested observationally and then refined. Ideally, it also defines research frontiers and then drives new mission and observation concepts. Theory is therefore both proactive and reactive, for besides demanding new observations, it must meet the challenges posed by existing data. The synergy that must exist between theory and experiment needs to be strengthened. The successful deployment of a scientific payload does not necessarily correspond to a successful scientific mission—it is the importance of the returned data to theoretical models and our ability to fully analyze, exploit, optimize, and refine these data that ultimately determine the success or failure of a mission. As important as instruments—space- or ground-based, in situ or remote—are to a scientifically successful mission, so too are data analysis, theory, and modeling. The panel strongly supports and endorses the current NASA Guest Investigator program and would like to see it strengthened, with similar programs created in other agencies.

Recommendation 2. The NASA Guest Investigator program should (1) be mandatory for all existing and new missions, (2) include both space- and ground-based missions, (3) be initiated some 3 to 5 years before launch, and (4) be peer reviewed and competed for annually, with grant durations of up to 3 years. Funding, at a minimum equivalent to 10 percent of the instrument cost, should be assigned to the Guest Investigator program and should explicitly support scientists working on mission-related theory and data analysis. Further, the Guest Investigator program for each mission should have the same status as a mission instrument. Other agencies should also consider guest investigator initiatives with their programs.

The implementation of this recommendation would address the very real concerns expressed by many experimentalists that too few theorists play an active role in exploring, interpreting, refining, and extending observations returned by expensive missions. The panel notes that in an era of fast missions, a cadre of theorists and data explorers needs to be in place and already active to take full advantage of a newly launched mission. A robust Guest Investigator initiative may also address the expressed concern that NASA expects PIs to submit proposals with extensive science goals but does not provide sufficient funding to support the science.

As set forth in recommendation 2, at least 10 percent of the instrument cost should be assigned to the Guest Investigator program, and this funding should be budgeted in the mission costs from the outset. The panel emphasizes that this is a bare minimum and that as much as 50 percent of the instrument cost would be more reasonable. The panel also recommends that a Guest Investigator program begin a few years prior to launch. For a $160 million MIDEX-class mission, 10 percent of the instrument cost may amount to approximately $3 million. Beginning 3 years before launch and assuming a 6-year lifetime for the mission after launch implies guest investigator funding of $1 million for every 3-year funding cycle, yielding a total of perhaps 12 guest investigators for the mission. The panel would regard 10 percent as a reasonable minimum since between 10 and 15 investigators are needed for a mission to provide a critical mass of researchers for data, theory, and modeling investigations. Although the panel would like to see considerably more funding for the Guest Investigator program, existing data analysis lines should not be cut to meet budgetary shortfalls elsewhere. It is essential that very expensive missions, ground- or space-based, should not be undercut by inadequate funding support for the data analysis. This is particularly true of LWS data analysis funds.

Challenge 5: Data Exploration and Assimilation

Government-supported ground- and space-based research projects are currently returning an unparalleled wealth of in situ and remote space physics observations. Systematic and correlative analyses of both these observations and archived observations help researchers test theories, identify new phenomena, define solar variability, and establish the relationships linking the Sun and interplanetary space to Earth's magnetosphere and ionosphere. The observations are essential to monitor and forecast space weather.

Research in space physics often requires the interpretation and intercomparison of large, complex, and diverse data sets derived from multiple instruments and locations. It already requires the assimilation of real-time data sets into numerical simulations and the direct comparison of observations with computer simulations (empirical and numerical). Successful exploitation of the increasingly large data sets expected from future projects and simulations will demand sharing and integration of archives on scales beyond current experience. The panel believes, and past experience has demonstrated, that observers, PIs, and modelers are best suited to maintaining, documenting, and providing their own real and synthetic data sets, including the simulation codes, whenever possible. On the other hand, data mining requires access to these data sets via centralized information trees and common standards. Only a distributed virtual data system can provide ready access to a wide variety of well-supported, high-quality data.

Recommendation 3. NASA should take the lead in convening a cross-agency consultative council that will assist in the creation of a cross-agency, distributed space physics information system (SPIS). The SPIS should link (but not duplicate) national and international data archives through a suite of simple protocols designed to encourage participation of all data repositories and investigator sites with minimal effort. The data environment should include both observations and model data sets and may include codes used to generate the model output. The panel's definition of data sets includes simulation output and supporting documentation.

Among other tasks, the system should do the following:

1. Maintain a comprehensive online catalog of both distributed and centralized data sets.
2. Generate key parameter (coarse resolution) data and develop interactive Web-based tools to access and display these data sets.
3. Provide higher-resolution data; error estimates; supporting platform-independent portrayal and analysis software; and appropriate documentation from distributed principal investigator sites.
4. Permanently archive validated high-resolution data sets and supporting documentation at designated sites and restore relevant offline data sets to online status as needed.
5. Develop and provide information concerning standard software, format, timing, coordinate system, and naming conventions.
6. Maintain a software tree containing analysis and translation tools.
7. Foster ongoing dialogues between users, data providers, program managers, and archivists, both within and across agency boundaries.
8. Maintain portals to astrophysics, planetary physics, and foreign data systems.
9. Survey innovations in private business (e.g., data mining) and introduce new data technologies into space physics.
10. Regularly review evolving archival standards.
11. Support program managers by maintaining a reference library of current and past data management plans, reviewing proposed data management plans, and monitoring subsequent adherence.

While a smoothly functioning data system will require enhanced funding, it is essential that the data system grow at a pace consistent with actual user needs, as measured by data set requests and community input. Solutions imposed by a central authority often fail to satisfy the requirements of working scientists. The SPIS must enable community researchers themselves to identify and prioritize problems, and then propose and implement practical solutions.

An appropriate and cost-effective management structure would include a small supervisory office (full-time project scientist, project manager, and administrative assistant) that reports to program managers at each funding agency, distributes funding to the system's nodes, and ensures cross-disciplinary integration. A panel composed of community members and representatives from the primary nodes should advise this supervisory office. Primary nodes within the SPIS should be organized by well-recognized scientific discipline (e.g., solar, heliospheric, magnetospheric, and ionospheric nodes). Each primary node should employ a half-time scientist, a quarter-time administrator, and a half-time programmer. The primary nodes would sponsor a set of smaller, dynamically evolving nodes tasked with accomplishing specific objectives, such as restoring and validating data sets or developing software and translators. To ensure responsiveness to evolving needs, all functions within the data system, including the central office, must be competed for periodically. A similar management structure, tasked with similar functions, was proposed at the Space Physics Data System Community Workshop held at Rice University in 1993[8] and has subsequently been endorsed at community forums and review teams.

Table 4.2 provides the suggested funding profile for an SPIS. For comparison, the NSF-NASA Virtual Solar Observatory initiative requested $4.3 million to support setup costs for 10 nodes and a core team during the first year, $2.9 million during the second year, $1.7 million each year thereafter, and between $115,000 and $350,000 for initial and subsequent costs for each node added. The ambitious goals of NASA's LWS and the NSF's Space Weather programs will require the implementation of a sophisticated data system like that envisaged in this recommendation, so part of the funding for

---

[8] NASA, Office of Space Science and Applications, Space Physics Division. 1993. *SPDS Concept Document on NASA's Space Physics Data System*. NASA, Washington, D.C.

TABLE 4.2 Funding Profile for a Space Physics Information System

Schedule	Tasks	Cost (million $)
Year 1	Establish central and four discipline nodes	1.5
Year 2	Fund competed tasks	3.5
Year 3+	Expand competition and services	5.5
Year 4	Continued expansion	7.5
Year 5+	System fully operational	10.0

SPIS should come from these programs. Pending the establishment of the SPIS, each agency might initiate or strengthen programs that devote funding to data management and access issues (e.g., NASA's AISRP). However, since SPIS will support research programs that cross agencies, disciplines, and methodology and will have, in addition, implications for data management in all the sciences, a broad new initiative for funding is clearly required that embraces the entire scope of the recommendation.

Challenge 6: Transition of Scientific Models to Operational Status

Despite the existence of many solar, heliospheric, and geospace models that can potentially be used for operational space weather forecasting, relatively few have so far been transitioned into operation at the NOAA and Air Force space weather centers. This is due to inadequate resources for transition efforts, in particular at the NOAA/Space Environment Center.

Recommendation 4. NOAA and the Air Force should initiate a program to support external research groups in the transitioning of their models to NOAA and Air Force rapid prototyping centers (RPCs). Program support should include funding for documentation, software standardizing, software support, adaptation of codes for operational use, and validation and should allow for the research group to assist in making the scientific codes operational. The RPC budgets of the NOAA Space Environment Center and the Air Force Space and Missiles Center/DSMP Technology Applications Division (SMC/CIT) should be augmented to facilitate the timely transition of models.

Competition within the transition initiative should be open to all potential model providers, and the grantees should be selected by the space weather centers on a peer-reviewed basis. As noted above, precise levels, particularly for the Air Force, are difficult to determine accurately. Costs at NOAA have been estimated based on prior costs to transition models that are currently operational at NOAA, including costs incurred directly by NOAA as well as by private-industry partners. The panel estimates that $1 million will be required to support the design and initial implementation of a software and database infrastructure; hardware acquisition; and initial validation, visualization, and hosting of two new models. Subsequently, $500,000 per year would be required to support at least three permanent NOAA staff and an additional $500,000 for operational support, real-time data links, software standardization, documentation, etc. Since the panel anticipates three to five codes per year being readied for transition at a cost of ~$200,000 each, an additional $1 million per year should be available for external researchers selected via open competition. A conservative annual budget is therefore between $2 million and $2.5 million.

The transition initiative should be funded through a partnership between the Air Force and NOAA, and the precise levels of funding should be determined by the needs of each. The NOAA/Space Environment Center and Air Force RPC budgets should be augmented, and more support from industry and business is necessary.

The successful implementation of this recommendation will recognize and strengthen the ability of space physics to contribute to economic, societal, and governmental needs, especially where space weather and space climatology impact human activities and technological systems.

5

Report of the Panel on Education and Society

SUMMARY

When considering the status and future of solar and space physics, we must also take into account the role of these disciplines in education at all levels. Solar and space physics is by no means unique in this—all areas of science have a responsibility to contribute to education. This responsibility is part of a new post-Cold War social contract between science and society, and it represents a considerable change for scientific communities that had not seen this broader responsibility as part of their core mission. In fact, in the past decade there has been a remarkable increase in educational activities by the solar and space physics community because of funding from the NSF and, especially, NASA, both of which have tried to involve the scientific community in science education.

However, efforts by the solar and space physics community to enhance science education do not take place in a vacuum. There is increasing recognition of our nation's need for a technically trained workforce and for a scientifically literate citizenry, and in particular for professionals trained in solar and space physics who will be capable of leading our efforts to understand, monitor, and respond to changes in Earth's space environment. Meeting this broad educational challenge requires all scientific communities to examine how they can contribute to meeting national goals in precollege, undergraduate, and graduate science education.

Large-scale efforts in K-12 science education reform have been and are being funded by the NSF. Moreover, there is a national movement to improve science education, the so-called "standards" movement, with which solar and space physics K-12 science education efforts must be aligned if we are to have an impact commensurate with the investment. Finally, there is a national need to recruit more students from populations that have historically not been a source of science students. Hispanics, African-Americans, and Native Americans all are becoming an increasing fraction of the undergraduate population, and we as a society need more of them to choose science careers.

Several national reports call for larger numbers of graduates in science and engineering fields, as well as for increased science literacy among nonscience majors. This requires changes in undergraduate science education. Solar and space physics can, and should, help improve general undergraduate science education, especially in gateway courses such as introductory physics or general education courses such as introductory astronomy. Moreover, by providing undergraduates increased opportunities to do meaningful and exciting research, solar and space physics can contribute to recruiting and retaining science and engineering majors. Solar and space physics also needs to recruit and retain excellent students for graduate study as well, since there is a need for more individuals with graduate science degrees in general, as well as a next generation of solar and space physicists, particularly as issues such as space weather become more important to our society.

Based on these considerations and on information gathered at several meetings with leaders in education, policy, and science, and with members of the solar and space physics community, the panel decided on four recommendations to help guide the community's next decade of education efforts. These recommendations, as well as the supporting arguments, are not necessarily unique to solar and space physics. In fact, much of what is contained in this document applies to other areas of science, since the problem that the panel is attempting to address here is systemic and of broad societal import. There are, of course, many areas of uniqueness, such as the tremendous effort made in the past decade by NASA's Office of Space Science (OSS) to significantly improve and expand the contribution of space science to general education. Where possible, the panel tries to point out the unique links to solar and space physics, or examples of how the community can contribute given its particular set of resources, one of the greatest being the enduring public fascination with space.

Recommendation 1. A program of "bridged positions" should be established that provides partial salary support, startup funding, and limited research support for four new faculty members per year for 5 years, yielding 20 new faculty lines in solar and space physics at U.S. universities over the next decade. This should be matched with an increased emphasis on solar and space physics research and hardware development at colleges and universities.

It is at the college and university level that research and teaching in solar and space physics can have the greatest and most direct impact over the next decade. In order to both increase the awareness of the importance of Earth's space environment among the next generation of the nation's leaders and foster a stronger national cadre of young and expert solar and space scientists, the panel recommends the establishment of a program of "bridged positions"—faculty positions that are partially supported by outside agencies for 5 years as an incentive for colleges and universities to strengthen (or initiate) programs in these fields. Moreover, agencies should seek ways to support the university research community, particularly those groups that build hardware, so as to maintain a strong link between the academic community, education, and research.

Recommendation 2. Federal agencies that fund solar and space physics should set aside funds to support undergraduate research in solar and space physics, either as a supplement to existing grants or as standalone programs.

Involving undergraduates in research has proven to be a positive factor in enhancing recruitment and retention of talented science students. Experiential education, which has its roots in the academic science laboratory, is now recognized to play a critical role in the development of both student expertise and confidence in nearly all academic fields. Such research experiences are available in solar and space physics, and resources to increase the ability of faculty to provide research opportunities for students are essential.

Recommendation 3. Three resource development groups should be funded over the next decade to develop educational resources (especially at the undergraduate level) needed by the solar and space physics community, to disseminate those resources, and to provide other services to the community.

Solar and space physics research projects already provide numerous images and informal educational opportunities for a wide audience in the media, in museums, via the World Wide Web, and to some extent at the K-12 level. As they are relatively new fields, however, relatively few applications or examples from solar and space physics currently appear in textbooks or in supplementary materials, particularly at the undergraduate level. But with sufficient support the popular fascination with space can be used to facilitate a nationwide advance in scientific literacy.

Recommendation 4. Current K-12 education and public outreach (EPO) efforts should be continued. However, there should be a careful evaluation of lessons learned over the past few years, particularly regarding the involvement of scientists in EPO activities, as well as increased coordination of NASA EPO efforts with other large projects in science education reform, especially NSF initiatives.

During the past few years NASA's OSS has begun to commit significant resources to education and public outreach. At the same time, efforts to revitalize and reform K-12 science education are well under way in several states, often with support from large federal programs, and particularly those funded by NSF. As the NASA efforts mature, the panel encourages closer cooperation and synergy with other existing programs.

5.1 INTRODUCTION

The relationship between science and society and the role of the scientific establishment in science education have undergone considerable evolution over the past 50 years, with a contribution to science education now being viewed as an essential commitment for the nation's science community. That period was also a time of huge growth in and the coming of age for the solar and space physics community. In charting its course for the next decade, it is important that the solar and space physics community consider how it can participate actively in education and outreach at all levels.

After World War II the importance of science was clear to all, since scientific and technical advances had been crucial to victory. There was also a broadly shared feeling that by advancing science, society as a whole would prosper. This idea was at the core of Vannevar Bush's seminal report *Science—The Endless Frontier*, which in essence laid out a social contract between science and society.[1] Science in general, and physics in particular, would expand and prosper, and the result of all the basic research would be the security of the nation abroad and increasing prosperity at home. This pivotal document did mention education in Section 4.0, "Renewal of Our Scientific Talent," although it did not call for the active involvement of the scientific research community in science education beyond training graduate students.

During the Cold War, and particularly after the launch of Sputnik on October 4, 1957—an event that looms large in the history of space physics—federal, state, and private support of basic research in universities and colleges, industry, and national laboratories flourished. The launch of Sputnik also precipitated a

---

[1] Vannevar Bush. 1945. *Science—The Endless Frontier, A Report to the President.* U.S. Government Printing Office, Washington, D.C.

national effort to improve science education. Scientists, especially physicists, played a seminal role in science education reform in the 1960s and early 1970s, but it was not then widely believed that scientists and the scientific community should be involved in precollege science education reform.[2] Solar and, especially, space physics were in their infancy in the 1960s, and there is no record of solar and space physicists or their community being significantly involved in science education at that time.

The end of the Cold War brought the recognition that the security and economic well-being of our nation are based on successful competition in the global economy, where technology, knowledge creation, and international cooperation are the engines of wealth. These issues were captured in the landmark book *Science for All Americans*, which laid out what a scientifically literate person should know and be able to do.[3] However, numerous reports, such as the Third International Mathematics and Science Study, have demonstrated that our educational system has not yet met the goal of imparting to our students an adequate level of scientific literacy.

With the growing importance of science education as a national priority, the scientific research community has been asked to play an important role in science education at all levels. The National Academy of Sciences and the American Association for the Advancement of Science have defined the features of a quality science education—widely referred to as "standards-based" (see Box 5.1)—in the reports *Benchmarks for Science Literacy*[4] and *National Science Education Standards*.[5] The standards have received support from scientific societies, including the American Geophysical Union, the scientific society with which most space scientists are associated.[6] It is now recognized that, in addition to performing basic research, practitioners in all areas of science must contribute to the scientific literacy of the nation.

Solar and space physics share the challenge of this new social contract between science and society. In fact, the solar and space physics community enjoys a natural advantage because of the keen interest in space shown by people of all ages. In recent years, in an attempt to use this advantage to help address national needs, significant resources have been invested in education and outreach programs by the agencies that fund solar and space physics research. These needs can be addressed at a wide range of scales, from the small scale of the individual scientist's contribution to the large scale of major agency-funded education and outreach projects (such as those associated with missions). The remainder of this report outlines those needs, reviews the current status of solar and space physics contributions to education and society, and presents recommendations for the future.

5.2 NATIONAL NEEDS

A TECHNICALLY TRAINED AND INCREASINGLY DIVERSE WORKFORCE

The United States is rapidly evolving into an economy in which the creation of wealth is tied to the creation and application of knowledge. In fact, the most dynamic sector of the U.S. economy in the past decade has been the technology sector. At the same time, it is increasingly difficult for technology companies to find, attract, and retain technically proficient employees. Organizations are taking steps to compete for technology talent and prepare for the long term. Companies have been developing internal and external resources to retrain and/or update the skills of current employees. They have also solicited technical talent overseas, and have successfully sought increases in the limits on H1-B visas for technically proficient employees. Yet the kinds of skills (computers and programming, electronics, optics, data analysis, etc.) that students gain by engaging in solar and space physics research at both the undergraduate and graduate level are precisely the skills that employers are seeking.

Despite various efforts, the number of undergraduates majoring in science and engineering has declined during the past decade. Many students who have an interest in science are discouraged by their early experi-

---

[2] R.E. Lopez and T. Schultz. 2001. Two revolutions in K-8 science education, *Physics Today,* September, pp. 44–49.

[3] F. James Rutherford and Andrew Ahlgren. 1989. *Science for All Americans*. American Association for the Advancement of Science, Project 2061. Oxford University Press, Cary, N.C.

[4] American Association for the Advancement of Science. 1993. *Benchmarks for Science Literacy, Project 2061*. Oxford University Press, Cary, N.C.

[5] NRC. 1995. *National Science Education Standards*. National Academy Press, Washington, D.C.

[6] American Geophysical Union (AGU). 2001. *Importance of the Earth and Space Sciences in Primary and Secondary Education: An Endorsement of the AAAS Benchmarks and NRC Standards*. Adopted by AGU Council December 2001. Available online at <http://www.agu.org/sci_soc/policy/earthspace_educ.html>.

BOX 5.1 THE STANDARDS-BASED MOVEMENT

One of the most significant developments in science education in recent years has been the emergence of science education standards at the national and state levels. The effort to establish science standards was spearheaded by scientific organizations, namely the National Research Council (NRC) and the American Association for the Advancement of Science (AAAS). The AAAS established Project 2061 to produce a series of documents, including Benchmarks for Science Literacy (1993),[1] which outlines what students should be studying at various grade levels in order to achieve science literacy. The NRC then published the National Science Education Standards (1995),[2] which outlines not only what students should know and be able to do (content standards), but also provides standards for teaching, assessment, schools, and systems.

These documents, collectively known as the "standards," form the basis of what is referred to as the standards-based movement, in which science education programs are aligned with the goals and methods set forth in the standards. States also have created individual state standards, which are often used as the basis for statewide testing of students. While many of the state standards are closely aligned with the national standards, some deviate significantly from the national recommendations. However, all state standards have been influenced by the national documents.

In general, the standards call for a more active, participatory approach to science education. The NRC report states that ". . . science is something students do, not something that is done to them." This is very much in line with the results of cognitive research.[3] Also implicit in the standards is the notion that reform is systemic; that it is necessary to take a global, systems view of science education if the promise of science literacy for all is to be achieved.

---

[1] American Association for the Advancement of Science. 1993. *Benchmarks for Science Literacy, Project 2061*. Oxford University Press, Cary, N.C.

[2] NRC. 1995. *National Science Education Standards.* National Academy Press, Washington, D.C.

[3] For example, NRC, 2000, *How People Learn: Brain, Mind, Experience, and School,* Expanded Edition, National Academy Press, Washington, D.C.

ences in introductory science classes.[7] Approximately 300,000 students take introductory physics courses each year. The topics taught in those classes (mechanics, electricity, and magnetism) include topics that relate directly to many aspects of solar and space physics. However, traditional presentation of these subjects has not drawn on solar and space physics phenomena to provide a real-world context for the physics.

Graduate enrollments in science and engineering decreased during the early 1990s, but increased somewhat toward the end of the decade. However, the National Science Foundation pointed out as follows: "With current retirement patterns, the total number of retirements among science and engineering degreed workers will dramatically increase over the next 10–15 years. This will be particularly true for Ph.D. holders because of the steepness of their age profile."[8] The projections are that unless we significantly increase the numbers of science graduates, the shortfalls already experienced by industry will only worsen.

When considering the need to increase enrollments in science we must also be concerned with the need to increase diversity in science. Total enrollments in all postsecondary education institutions rose from 10,985,000 in 1976 to 14,345,000 in 1997, and a disproportionate part of this growth came from increases in minority groups going to college. While white, non-Hispanic enrollment went from 9,076,000 in 1976 to 10,161,000 in 1997, enrollments of Hispanics, African-Americans, and Native Americans went from 1,493,000 to 2,872,000.[9]

Given that an increasing fraction of students is coming from groups that historically have been underrepresented in science, any attempt to increase the number of technically trained professionals must grapple with the issue of fostering greater diversity in science. This is especially true in states such as Texas and California, in which there will soon be no majority ethnic group. Recent diversity initiatives by NASA and the NSF have focused on using solar and space physics to enhance science education in minority-serving institutions and to recruit underrepresented students for science careers (see Box 5.2).

---

[7] E. Seymour and N.M. Hewitt. 1997. *Talking About Leaving: Why Undergraduates Leave the Sciences.* Westview Press, Boulder, Colo.

[8] National Science Board, National Science Foundation. 2000. *Science and Engineering Indicators—2000,* NSB-00-1. U.S. Government Printing Office, Washington, D.C.

[9] National Center for Education Statistics. 2001. *Digest of Education Statistics—2000,* NCES-2001-034. U.S. Department of Education, Washington, D.C.

BOX 5.2 DIVERSITY INITIATIVES IN SOLAR AND SPACE PHYSICS

The need to increase diversity in science has become increasingly recognized as a priority for federal agencies. Those agencies responsible for solar and space physics are no exception, and NSF, NASA, and NOAA have in recent years launched diversity initiatives that use solar and space physics to attract students into science.

In 1996, NSF held a workshop to examine the issue of diversity in the geosciences in order to make recommendations for a diversity strategy. This led to the creation of a Diversity Initiative program and grants to a variety of universities and organizations such the Society for the Advancement of Chicanos and Native Americans in Science. While the NSF effort is aimed at increasing diversity in the geosciences broadly, it does include space physics, which is in the geosciences directorate.

At the same time NASA was developing a diversity strategy, and in the summer of 2000, OSS launched its Minority University Initiative (MUI). The MUI made funds available to minority-serving institutions for a wide range of programs, such as new space science courses and degree programs and public education and outreach efforts. While both efforts are in their early phases, both are quite promising and may soon serve as models for increasing diversity in solar and space physics.

NOAA has also created a diversity initiative aimed at supporting NOAA-related science research, and it issued a request for proposals in 2002. Like the NASA initiative, the funds are targeted at minority-serving universities. If solar and space physics does not have a presence in such institutions, then it will not be able to contribute to diversity programs like those of NOAA and the NSF, which target a wide range of science disciplines.

A SCIENTIFICALLY LITERATE CITIZENRY

In addition to the need for technically trained professionals, there is also increasing recognition of the value of a scientifically literate citizenry. Making informed political and economic decisions, and even consumer choices, in a world permeated by science and technology requires increasingly knowledgeable citizens. *Science for All Americans* arrived at a consensus on scientific literacy by examining the technological society around us and determining what a scientifically literate citizen should know and be able to do.[10]

At the university level, introductory astronomy (Astronomy 101) enrolls more students nationwide than any other science class for nonscience majors. Solar and space physics topics and resources have not been fully utilized in introductory astronomy, in part because of the relative youth of the field. However, as technology advances, so too must the knowledge base of our citizens. In the future it will become increasingly apparent that some knowledge of the space environment and how it can affect humans and our technology is part of science literacy. As another case in point, solar physics is likely to play an increasingly important role in debates about global climate change.

Education in science begins in elementary school, and for many of our nation's current and future leaders, it continues through the college and university level. As pointed out above, science literacy also must include some understanding of our society's current reliance on space-based monitoring and communications. In this area, solar and space physics can make substantial contributions. The well-known fascination that space exploration holds for most people provides one avenue for beginning to educate future citizens in a variety of science topics. The challenge is to find the points in the educational system where the resources and talents of the solar and space physics community can contribute most effectively to national goals.

It is important to recognize the key role that undergraduate education plays in both educating citizens and preparing future teachers and scientists, including future solar and space physicists: "It is in college where future scientists and college faculty are recruited and prepared for graduate study; where our nation's elementary and secondary teachers, educators of America's youth, are equipped; and where tomorrow's leaders gain the background with which to make critical decisions in a world permeated by vital issues of science and technology."[11]

Although undergraduate education has not historically been a focus in the solar and space physics community, the panel believes that the community has much to offer in this area. Given the comments of many solar

---

[10] F. James Rutherford and Andrew Ahlgren. 1991. *Science for All Americans*. Oxford University Press, New York, N.Y.

[11] Project Kaleidoscope. 1991. *What Works: Building Natural Science Communities,* Vol. 1. Project Kaleidoscope, Washington, D.C.

and space physicists during the preparation of this report, it also believes that the community will respond enthusiastically to a call for greater involvement in undergraduate education, for both science majors and nonscience majors.

THE NEED FOR SPACE SCIENCES

Stepping away from the larger issue of a technically trained workforce and general science literacy, there is also the need to ensure that there are sufficient technical professionals in solar and space physics to meet the growing national need to understand and monitor the space environment, which is of critical importance to our nation's assets, both those in space and those on the ground. We are becoming increasingly dependent on orbiting satellites for applications such as communication networks, global positioning for ship and airline navigation and for military operations, and monitoring the Earth system for climate change and weather forecasting. Astronauts travel into space on a space shuttle and now permanently inhabit the International Space Station. As our dependence on space for economic and national security increases, so must our public awareness and understanding of the space environment grow. This includes our ability to monitor and predict conditions in space and to better characterize and understand possible space weather impacts both in space and on the ground.

The near-Earth regions of space are driven by the Sun and vary from minute to minute and day to day within the 11-year cycle of solar activity. Like seasonal variations in the terrestrial weather, each stage of the solar cycle is characterized by its own set of conditions that affect different sectors of human and technological activity. During the period of minimum solar activity, effects such as spacecraft charging and the resultant abrupt electrical discharges due to energetic electrons can seriously damage our assets in space. During solar maxima severe disturbances degrade satellite power systems, enhance the atmospheric drag on orbiting satellites, damage satellite instrumentation, disrupt electric power distribution on the ground, interfere with telecommunications, and pose radiation hazards to astronauts.

Unlike terrestrial weather, which is monitored routinely at thousands of locations around the world, the conditions in space are monitored by only a handful of space- and ground-based facilities. Space weather forecasters are required to specify and to predict conditions in space with very little guidance from actual measurements. Given this extreme undersampling of the diverse, coupled regions of space, extending all the way from the Sun to Earth, computer models that provide continuous quantitative assessment and prediction of the geospace environment are required.

Extensive scientific research, modeling, and monitoring efforts directed at understanding the space environment have produced a broad spectrum of data and modeling resources. In the United States, various aspects of this effort have been funded by NASA, the National Science Foundation, NOAA, the Department of Defense, the Department of Energy, and the Department of the Interior. An interagency effort has been initiated to fund research targeted specifically at understanding and predicting the space environment. This initiative, the National Space Weather Program (NSWP), stems from the broad interest in space shared by commercial, educational, and governmental organizations. A primary goal of the NSWP is to focus and to build on our existing resources to produce quantitative predictive models of the space environment. Research relevant to the NSWP also benefits from strong international collaboration among scientists and space weather forecast centers.

Space weather monitoring and prediction will be an area of growth in the future, and the community must make certain that professionals are being trained in the field. As discussed below, there is reason to be concerned about the status of solar and space physics in colleges and universities, where students are recruited and trained, and about means for increasing diversity in solar and space physics, since the field has not traditionally had a presence in minority-serving institutions.

5.3 THE ROLE OF SOLAR AND SPACE PHYSICS

A MOTIVATOR FOR THE GENERAL PUBLIC

Our society maintains an abiding interest in space science and exploration. Built upon decades of thrilling exploration of near-Earth space and the solar system, public interest in events or activities associated with space continues at a high level. Indeed, solar and space physics is one of the few areas of scientific research that combines the rigors of science with awesome and inspirational natural phenomena (aurorae, sunspots, eclipses) that can sometimes be viewed directly by people on Earth. It is no wonder, therefore, that the public's fascination with solar and space physics has continued.

Over the past decade, the solar and space physics community has become more aware of the importance of sharing the excitement of space science research with the public. Numerous scientists have participated in efforts to bring solar and space physics to the public through various venues and media. Solar and space physics has in recent years been highlighted by the IMAX film *SolarMax* and museum exhibits like *Electric Space*. Sun-Earth Day 2002 allowed solar and space physicists across the country to engage in a variety of public outreach events. And recurring news stories, such as the attention paid to the January 1997 magnetic storm or the erupting prominence photographed by SOHO on July 1, 2002, testify to ongoing public interest in the field. Bringing the wonder of science to the public is an important part of the solar and space community's contribution to society.

PROVIDING EDUCATIONAL RESOURCES FOR K-16 EDUCATION

Solar and space physics must also address scientific literacy through formal science education. Teachers need the active involvement of the scientific community to support high-quality science education. They also need contact with the current science in order to communicate relevance to their students. Solar and space physics can provide an exciting context for science education. And while precollege science education is very important, the community must think beyond K-12, which is what most scientists think of when speaking of "education." Specifically, solar and space physics can contribute substantially to undergraduate education by providing the context for instructional programs in many different disciplines.

PROVIDING OPPORTUNITIES FOR UNDERGRADUATE RESEARCH

Undergraduate research experiences are widely recognized as having a significant impact on the recruitment and retention of science and engineering majors. The solar and space physics community never viewed undergraduate research as a priority, in part because the community has been historically more involved in government laboratories and research centers. Yet many research topics in solar and space physics are quite amenable to undergraduate participation, and the lure of space research is strong for many students.

5.4 SOLAR AND SPACE PHYSICS IN COLLEGES AND UNIVERSITIES

HISTORICAL BACKGROUND

Solar and space physics may be described either as a collection of interdisciplinary fields or as parts of a newly emerging discipline. Although historical records reveal curiosity about the Sun and the heavens in many societies, their scientific study has emerged only since the dawn of the space age, when satellites and rocket-borne probes could observe beyond the confines of Earth's atmosphere.

Recognition of the importance of solar and space physics is increasing but is still limited. In a sense, the study of space has come full circle. In ancient Egypt, the need to predict the date of the annual flood of the Nile River, which brought the annual supply of precious water for the land's crops, led Egyptian astronomer-priests to study the skies to find a way to give advance warning of the water's arrival. Their search was successful; they learned to associate the rising of the Nile with the time each year when the star Sirius first appeared in the eastern sky. Now, 4,000 years later, with our technological society increasingly affected by streams of particles and radiation from the Sun, we again must study what is "above" us in order to protect and benefit our society.

Space physics has roots in several different scientific fields. These include geophysics (from the study of Earth's magnetic field, the upper atmosphere and ionosphere, and the aurora), elementary particle and cosmic ray physics (from the study of energetic particles and radiation originating beyond Earth), electrical engineering (from the study of radio emissions and propagation above Earth's surface), and, of course, astronomy (the study of the Sun, planets, asteroids, and other solid bodies; the solar wind and interplanetary medium; and the heliosphere and its interaction with local interstellar gas).

The discipline of solar physics is similarly young, having come out of the larger field of astronomy. Advances in instrumentation made possible detailed study of the physics of our nearest star. Those studies have led, among other things, to the recent discovery of neutrino flavor oscillation. The increasing recognition that Earth and all other objects in the solar system are bathed in the solar wind (essentially an extension of the Sun's atmosphere) and that dynamical and at times explosive processes originate on and/or inside the Sun has fueled continuing research.

The fields of solar physics and space physics are now closely coupled. Research satellites and ground-based instruments monitor the complex trail of variations in solar energy from their origin within the Sun, outward through the corona and solar wind, past the inner planets and Earth, outward throughout the solar system. The solar system has become a natural laboratory for understanding a number of fundamental astrophysical processes. Furthermore, beyond the long-range benefits of this fundamental research, substantial interest is directed toward understanding the impact of these highly variable processes on Earth's increasingly technological society.

Many of the programs in solar and space physics at U.S. universities and colleges were founded with substantial external support from NASA and other federal agencies. NASA even built space science centers on many university campuses in the 1960s. Over the past two decades, however, solar and space physics has experienced decreasing visibility and support on university campuses despite still-ample funding for specific research projects. The long development time lines for missions have also had a negative impact on graduate education.

In part because of their relatively short history, and in part because of the great commercial interest in other high-tech areas in the past two decades, solar and space physics as disciplines now have little visibility in either the K-12 educational system or higher education. Graduate education in solar or space physics is scattered among a variety of departments, variously within physics, geophysics, astronomy, and electrical engineering programs. These disciplines have no presence in any department at a number of major universities. At the undergraduate level, only a handful of institutions offer specific courses, much less minors or majors in these areas. For example, a survey of solar physics groups as identified by the Solar Physics Division of the American Astronomical Society reveals that of the 37 institutions that host solar physics groups, only 13 have groups of three or more solar physicists. Only one solar physicist is found at 15 of these institutions. Space physics has a similar, less-than-robust presence in universities.

Such an absence from the academic world ensures that students will have little exposure to solar and space physics and that the community will not be able to contribute as well as it could to the national needs discussed at the outset of this report. Moreover, the rise of separate, narrowly focused scientific journals and scientific societies for solar and space physics, typical signs of a field's maturity, have ironically also led to a loss of interaction between the fields and the wider communities of physicists, astronomers/astrophysicists, and engineers. Thus, faculty at colleges and universities do not even realize that they could establish an effort in solar and space physics and that faculty in these areas could be a significant asset to the educational aims of the institution.

To be sure, introductory college-level astronomy textbooks now often include material on the solar wind, the aurora, and the ionosphere. Similarly, many introductory college-level physics textbooks make reference to the occurrence of fundamental electromagnetic interactions in space, though such references are cursory. In addition, although textbooks in electrical engineering have for many years discussed the impact of the ionosphere and its variations on the propagation of radio waves, more comprehensive treatment is now required. Our society's accelerating use of satellite-based communications systems in both low Earth orbit and high geosynchronous orbit have led to a significant emphasis within the communications industry on space-based communications and hence to a need for understanding the environment in which these systems function.

In the past decades, the importance of solar and space physics within our nation's system of higher education has grown, not shrunk. Continued support from NASA, NSF, and other interested federal agencies is needed to sustain the vitality and, especially, the visibility of this field on our campuses. This is especially true for minority-serving institutions that historically have not been part of the solar and space physics enterprise. The field has much to offer such institutions in terms of outstanding opportunities for students and the excitement that space science generates.

SOME ISSUES IN UNDERGRADUATE AND GRADUATE SCIENCE EDUCATION

Access to undergraduate education has grown explosively in the past 50 years, and access to at least introductory college-level science courses has increased at a nearly comparable rate. However, the fraction of undergraduates completing majors in science, mathematics, or engineering has not at all kept pace. Increasingly, positions in our graduate schools of science and engineering, as well as in our industries and research laboratories, are being filled by students from other nations, if they are being filled at all. When in addition one considers the ongoing shortage of K-12 teachers with a science background and the large number of current science teachers scheduled to retire in the next decade, one cannot but recognize that a renewed focus on undergraduate science education is long overdue.

Today there are fewer physics majors in the United States than at any time in the past 40 years. At the same time, there is a larger pool of high school students studying physics than ever before, and the total number of bachelor's degrees awarded in the United States has gone up. During the mid-1950s, 10 of every 1,000 bachelor's degrees awarded in the United States were in physics, whereas today the number is only 3 of every 1,000. The decline is not unique to physics in relative or absolute terms. In 1998, engineering departments reported that they awarded their lowest number of bachelor's degrees in 17 years. Similarly, the bachelor's degree class in computer science in 1997 was the smallest in 11 years.[12] Overall, the number of engineering bachelor's degrees granted decreased by 16 percent from 1983 to 1996.[13]

Just as retention is an issue for undergraduate physics and other science programs, recruitment is an issue for graduate programs. The number of first-year graduate students in physics and astronomy is roughly the same this year as it was in the late 1970s, but the proportion who graduated from U.S. and non-U.S. undergraduate programs has changed dramatically. In the late 1970s, more than 2,200 came from U.S. colleges and universities, while about 800 came from foreign institutions. In 2000 fewer than 1,500 came from U.S. colleges and universities and slightly more from foreign institutions. Low graduate enrollments lead to further erosion of undergraduate programs in some departments, because many universities rely on ever-scarcer graduate teaching assistants to direct laboratory and tutorial sessions. Recent moves to offer more generous graduate stipends and fellowships do not address the larger problem of there being simply too few students who have maintained an interest in advanced scientific research through their undergraduate years. Investing additional funds only at the graduate level will not adequately address the shortage of graduate students, because in many cases decisions to not continue in a scientific field are made long before specific graduate school offers are under consideration.

Other panels, including the Boyer Commission on Educating Undergraduates, and a recent study by the National Research Council[14] have discussed these issues. For example, although descriptive astronomy courses continue to play a vital role in imparting general science literacy to undergraduates nationwide, there is evidence that many introductory undergraduate science courses, especially introductory physics, continue to present a daunting and often unattractive perspective on science.[15]

Clearly, there is a strong national interest in reversing these trends. And fortunately, there are success stories in recruiting and retaining students. One bright spot of undergraduate science education—undergraduate research—is a proven success in motivating and retaining students in science. For example, University of Texas system science and engineering students involved in the Louis Stokes Alliance for Minority Participation, a program funded by the NSF that supports undergraduate research, have a 90 percent graduation/retention rate. Solar and space physics can actually be quite amenable to undergraduate research. For example, undergraduates are quite capable of assisting in software development and data analysis for current and past spacecraft or ground-based data sets. Below the panel suggests several steps that can be taken by the community to increase the extent of such research in solar and space physics and thus contribute to a critical national goal.

Undergraduate Research

The Boyer Commission on Educating Undergraduates has identified a major reason for the crisis in undergraduate science education as a destructive lack of connection at many colleges and universities between undergraduate study and the creation of future research faculty.[16] In many universities undergraduates are isolated from the challenge and excitement their professors find in research. However well presented undergraduate courses may be, they do not and essentially cannot expose students to the character of the research world. Indeed, ". . . many studies have shown that the undergraduate programs most successful at producing scien-

---

[12] Kate Kirby, Roman Czujiko, and Patrick Mulvey. 2001. The physics job market: From bear to bull in a decade. *Physics Today,* April, Vol. 54, p. 40.

[13] National Science Board, National Science Foundation. 2000. *Science and Engineering Indicators.* U.S. Government Printing Office, Washington, D.C.

[14] National Research Council. 2001. *Physics in a New Era.* National Academy Press, Washington, D.C.

[15] E. Seymour and N.M. Hewitt. 1997. *Talking About Leaving: Why Undergraduates Leave the Sciences.* Westview Press, Boulder, Colo.

[16] The Boyer Commission on Educating Undergraduates in the Research Community. 1998. *Reinventing Undergraduate Education: A Blueprint for America's Research Universities.* State University of New York at Stony Brook, Stony Brook, N.Y.

tists are those that include research and publication in refereed journals. . . . Students who have the opportunity for research complete their science programs in greater numbers than those who do not."[17]

Although undergraduate research involvement may have been overlooked at many universities, those departments at which it is prominent have frequently reported steady or increased enrollments in recent years. A survey of the undergraduate enrollment trends in 750 U.S. physics programs from 1990 to 1997 indicated that those departments with increasing numbers of graduates had taken steps in at least one of the following areas: instituting double majors or programs linking to engineering degrees; increasing opportunities for student research; doing more student mentoring; and offering practical career skills.[18] The panel thus believes that one significant step that the United States can take to increase the number of college and university students in science majors is to expand opportunities for undergraduate research experiences (see Figure 5.1).

FIGURE 5.1 Andy Sackreiter, an Augsburg College undergraduate, installing a magnetometer in the Canadian Arctic. Courtesy of Mark Engebretson, Augsburg College.

On-campus undergraduate research and off-campus research internships are valued not only because of the hands-on experience they provide. They also offer students the opportunity to work in a team or group environment, make decisions that affect the success or failure of the project, and feel the responsibility and satisfaction of being creative. Employers and graduate schools like to see this kind of experience in applicants.[19]

Recent studies continue to show the high correlation between undergraduate research experience and continuing on to scientific careers. A recent survey by the American Institute of Physics indicated that undergraduate research is already nearly universal among those heading for graduate work in physics (90 percent), whether as part of a senior thesis project, participation in a faculty-directed research project, or in a cooperative work arrangement with an outside employer. Such research experience was also quite common among those heading for graduate study in other fields (70 percent) or going directly into the workforce (63 percent), but was less prevalent (45 percent) among those going into high school teaching. The report noted that the last-mentioned percentage was "doubly unfortunate, first because such teachers will have a responsibility to introduce their own future students to a hands-on laboratory experience, and second, because such teachers typically have little opportunity to work in physics research after completing their undergraduate studies."[20]

Although research efforts in solar and space physics at American colleges and universities typically involve faculty, full-time research associates, and graduate students, they are not limited to those levels. Solar and space physics also provides attractive research projects for undergraduates in a variety of related disciplines, including electrical and mechanical engineering, computer science, and business. Opportunities can originate from diverse projects: the design and execution of a sounding rocket payload, the utilization of sophisticated computer systems, the management and operation of a small satellite program, or the design and verification of high-technology instrumentation. Much of the analysis of data obtained by high-tech instruments, as well as by research satellites, is also appropriate for undergraduate involvement. Waves and wave propagation are central topics in the undergraduate curriculum, as are electromagnetic fields, charged particles, and particle distribu-

---

[17] Robert Gavin. 2000. The role of research at undergraduate institutions: Why is it necessary to defend it? Chapter 1 in *Academic Excellence: The Role of Research in the Physical Sciences at Undergraduate Institutions,* M.P. Doyle, ed. Research Corporation, Tucson, Ariz.

[18] Robert Ehrlich. 1999. *The Physics Teacher* 3, 142.

[19] Barrett Ripin. 2001. Preparing physicists for life's work, *Physics Today,* April, p. 47.

[20] Michael Neuschatz, Patrick J. Mulvey, and Starr Nicholson. 1999. *Physics and Astronomy Senior Report: Class of 1998,* American Institute of Physics (AIP) Report R-211.30. AIP Statistical Research Center, College Park, Md., December.

tions. The time scales of these projects make it possible for undergraduates to participate in meaningful ways, and by so doing they not only learn about research but also develop their computer skills, learn about instrument development and testing, and gain practical experience they can apply to their subsequent course work.

Undergraduate research in solar and space physics provides benefits to students regardless of whether they continue their studies in one of these areas. First, because of the fundamental interconnection of all physical sciences, many of the concepts and techniques introduced in one research field will be applicable in others. Second, science education research has shown that higher-level learning is enhanced when students obtain research experience in more than one field. Indeed, many individuals currently active in solar and space physics began their research careers as undergraduates, often in other fields. For these reasons, the geosciences directorate of the NSF, which includes space physics, has called for additional resources to be made available to undergraduate institutions to support research activities.[21] Solar and space physics should take advantage of such initiatives to expand undergraduate participation in research.

The panel commends the NSF's Research Experiences for Undergraduates (REU) program, which provides support for undergraduate research involvement both within existing research grants to faculty and in stand-alone (mainly summer) programs. NSF's Research at Undergraduate Institutions program and the earlier Keck Geology Consortium have also provided excellent funding opportunities nationwide, and locally supported undergraduate research programs have continued at scattered institutions nationwide, not all of them wealthy or prestigious. There are other effective federal programs for the support of undergraduate research, such as the Department of Education's McNair Scholars Program (which is not limited to the natural sciences) and the Significant Opportunities in Atmospheric Sciences program at the National Center for Atmospheric Research, which has had considerable success in exposing students from underrepresented groups to the geosciences (including, on occasion, solar and space physics). Grassroots efforts, such as the Council on Undergraduate Research and Project Kaleidoscope of the Independent Colleges Office, have also succeeded in focusing attention on, and supporting, undergraduate research programs.

Although there are some discipline-specific summer programs for selected undergraduates at individual NASA centers, as well as several state-level space grant programs funded ultimately by NASA that support undergraduate research by students at their home colleges and universities, such activity does not appear to be an explicit part of NASA's educational strategy. The panel recommends that NASA provide specific funding for undergraduate research as part of proposals for both instrument development and data analysis, as it does now for other types of educational and public outreach activities, perhaps following the model of current NSF grants, or possibly even via a system of block grants. NASA may also want to consider other funding models, including those listed above, that target specific populations of students but not specific categories of colleges or universities. That is, such support for undergraduate research should focus not only on large institutions, but also on small ones, minority institutions, and those serving minorities in regions where there are no historically minority institutions. The panel also recommends that programs to support undergraduate research include individual young scientists just beginning their faculty careers as well as active groups and individuals already active in research.

One common objection to a greater emphasis on undergraduate research is that faculty time is already stretched thin among courses, advising, and, possibly, research. Undergraduate involvement in research can be promoted by funding for faculty for this purpose. Some smaller departments, despite their relative success in educating undergraduate physics students, already have difficulty offering their students more than a minimal set of advanced courses. One solution might be to develop consortia or partnerships between educational institutions to extend the offerings of intermediate and advanced courses for physics and astronomy majors. Such an approach would have both economic and pedagogical advantages: economic in that costs could be shared, and pedagogical in that small departments could offer a greater variety of advanced courses to their students and faculty would have more time to lead undergraduates in research. As a side benefit, such partnerships can emphasize the cooperative nature of science and engineering as faculty members themselves serve as models for the ability to work collaboratively with others in their field.

A related means of improving the undergraduate education of physics majors is to provide a set of short, targeted summer schools to introduce students to the research enterprise and the complex, cooperative na-

---

[21] National Science Foundation (NSF). 1997. *Geoscience Education: A Recommended Strategy,* NSF 97-171. Arlington, Va.: NSF.

ture of scientific research. One notable example is the Polar Aeronomy and Radio Science school at the University of Alaska, Fairbanks, which includes graduate students and some undergraduates in a 2-week late summer program combining focused classroom sessions with immersion experiences in large laboratories. Similar programs, perhaps purely for undergraduates, could be targeted at those who have completed their sophomore or even their freshman year.

Instructional Materials for Undergraduate Education

The community of solar and space physicists is presented with several educational opportunities. The public's fascination with outer space (part of its fascination with the larger field of astronomy) provides an avenue to achieve the broad goal of science literacy. There is, as well, a practical need to promote understanding of the science of the space environment as an integral part of global science and technology. The solar and space physics community must ensure that the next generation of scientists, engineers, and civic and corporate leaders have a working knowledge of Earth's space environment.

All students in technical majors study physics. In addition, approximately 10 percent of all college students take introductory astronomy,[22] and it is often their last experience with formal science education. This juncture presents an excellent opportunity for solar and space physics to contribute to general science education and to increase knowledge of the space environment. Instructional materials can be developed that use solar and space physics as a context. In the case of astronomy courses, the connection is natural. In the case of introductory physics courses, phenomena drawn from solar and space physics can be used to illustrate basic concepts. Many connections can be made between solar and space physics and topics in mechanics or electricity and magnetism. However, to contribute in this way, the solar and space physics community needs to develop and disseminate teaching resources and have a greater presence in college and universities.

Some Issues in Graduate Science Education

The solar and space physics community has the obligation to train the next generation of solar and space physicists. Some of the problems outlined earlier that have led to smaller numbers of adequately trained or inspired bachelor's-level graduates feeding into graduate programs must be addressed at the undergraduate level or even earlier. Other problems facing the community, however, must be addressed at the graduate level. For example, the shrinking number of graduate programs and the shrinking size of existing groups in solar and space physics at universities in the United States make it difficult to cover the breadth of the field in graduate courses.

Another issue in training graduate students is the increasing need for students trained in hardware and instrumentation. The large number of instrumentation programs at NASA centers and other government labs exacerbates this widely recognized problem. A special effort should be made to maintain and enlarge university groups that specialize in ground-based or space-based instrumentation and also to provide funding opportunities for graduate students being trained in instrumentation. By making it a priority to maintain university hardware groups, or even to help seed new ones, agencies will be able to preserve a critical link between research and education. Moreover, agencies should give particular consideration to establishing and supporting such a group or groups at minority-serving institutions. By creating the conditions that would allow research groups at such institutions to thrive, agencies with an interest in solar and space physics will also help to increase diversity in science.

5.5 K-12 SCIENCE EDUCATION AND PUBLIC OUTREACH

SCIENCE EDUCATION REFORM, THE NATIONAL STANDARDS, AND SOLAR AND SPACE PHYSICS

The National Science Education Standards[23] provide guidance on specific content understandings and process skills that students should achieve by certain grade levels. These standards allow using solar and space physics basic content and research to enrich student learning. Solar and space physics is rich in fascinat-

---

[22] National Research Council. 2001. *Astronomy and Astrophysics in the New Millennium*. National Academy Press, Washington, D.C.

[23] National Research Council, National Committee on Science Education Standards and Assessment. 1996. *National Science Education Standards*. National Academy Press, Washington, D.C.

ing examples of applied science that can satisfy content standards set for physical science, life science, Earth and space science, and the history of science and bring learning to life for students. Similarly, the standards stress the importance of students gaining an understanding of systems, cycles, and scales—all richly exemplified in this field of research.

School systems around the country are working to align their science education programs with standards, be they the national standards or (more likely) state standards. For solar and space physics to play a role in science education it is necessary that our science content support system goals. Unfortunately, neither the National Science Education Standards nor state standards explicitly include much of solar and space physics. This does not, however, mean the solar and space physics community cannot contribute to general science education in schools. When considering content issues, it just has to find the connections.

In some cases, the connections between solar and space physics research and the National Science Education Standards are explicit. For example, through the standards for Earth and space science, high school students are expected to learn that the Sun is a star and to understand how the solar system evolved, and how the Sun will evolve with time. They are not expected to gain an understanding of the magnetospheres of the planets and how they behave. However, through standards identified for the physical sciences, we can, for example, use Earth's magnetic field as an example of a planetary-scale magnet. As magnetic lines of force are explored in satisfaction of the physical science standards, Earth's interesting magnetic field geometry can be introduced, leading to interesting questions that can then be connected to the overarching concepts of coupled systems, scale, and the history of science.

In summary, although the standards do not specifically call for some topics that are central to solar and space physics (for example, the plasma state of matter is not mentioned), their focus on broad scientific knowledge yields many opportunities for the community to contribute. Many current education efforts in the community, including the NASA Office of Space Science program, have a component that supports standards-based instruction.

Magnetism, electricity, and the nature of the solar system are topics that every student is expected to master. Collaboration between the NASA-funded Sun-Earth Connections Education Forum and the Lawrence Hall of Science in Berkeley produced *The Real Reasons for the Seasons*, a middle-school curriculum package that teaches about seasons and the phases of the Moon. This excellent product is now in use by many school systems around the country, since all state middle-school standards include the solar system. Materials produced by the Imager for Magnetopause-to-Aurora Global Experiment (IMAGE) education and public outreach project have also been developed with standards in mind. As more mission EPO projects and individual solar and space physics scientists become aware of the need to carefully align their products with various standards, we can expect that the solar and space physics community will have an increasing impact in schools.

It must also be remembered that science is more than content. The standards point out that teachers need opportunities to understand how scientists think and approach problems. By providing teachers with opportunities to build long-term relationships with solar and space physicists, the community can contribute significantly to their professional development. Individual scientists can make a tremendous contribution by volunteering their time to work with a single teacher or group of teachers.[24] And, on a larger scale, professional development programs for teachers set up by the solar and space physics community should take advantage of the research base[25] on how best to incorporate scientists into school programs.

NASA EDUCATION AND PUBLIC OUTREACH AND THE CONNECTION TO NSF EDUCATION INITIATIVES

The current EPO activities of the solar and space physics community are driven heavily by NASA policies and funding. NASA has always made a point of communicating the U.S. effort in space to the public as a means of securing public support for the space program. Thus, much of the NASA effort in the past was directed essentially to public relations as opposed to serious contributions to science education. However, in the mid-1990s, the Office of Space Science initiated an ambitious program to mobilize space science in support of science education.

The NASA/OSS plan called for the creation of a set of "forums" and "brokers." Together, the forums and brokers constitute the "support network." The forums

---

[24] National Research Council. 1996. *The Role of Scientists in the Professional Development of Science Teachers.* National Academy Press, Washington, D.C.

[25] See, for example, S. Loucks-Horsley, P.W. Hewson, N. Love, and K.E. Stiles, 1998, *Designing Professional Development for Teachers of Science and Mathematics,* Corwin Press, Thousand Oaks, Calif.

TABLE 5.1 Forums

Organization	Topic
Space Telescope Science Institute	Astronomical Searches for Origins
Goddard Space Flight Center and Berkeley Space Science Laboratory	Sun-Earth Connection
Jet Propulsion Laboratory	Solar System Exploration
Smithsonian Astrophysical Observatory	Structure and Evolution of the Universe

TABLE 5.2 Brokers[a]

Organization	Region
Lunar and Planetary Institute	Southwest and Southern Plains
Southeast Regional Clearinghouse	South and Southeast
Center for Educational Technologies	Mid-Atlantic region
New England Space Science Initiative in Education	New England
DePaul University	Upper Midwest
Space Science Institute	West and Northern Plains
Space Science Network Northwest	Northwest

[a] As of June 2002.

were to take the results of space science research and translate them into materials and resources useful to educators. They were discipline-based, with one forum for each of the four science themes in OSS (see Table 5.1), and they were awarded without competition. The brokers were to facilitate relationships between scientists and educators, using the products of the forums. As a result, they are geographically based (see Table 5.2 and Figure 5.2). In contrast to the forums, there was an open competition for the brokers. In 2001 the brokers were recompeted, leading to the set of brokers listed in Table 5.2.

This network has had several successes, such as funding workshops for scientists, making deeper connections with minority professional societies, supporting innovative projects in the informal realm, and producing quality educational products.

A recent evaluation based on extensive interviews with a wide range of OSS EPO providers, customers, and others documents areas where real progress has been made.[26] The support network has established strong working relationships with informal science centers around the country, leading to the development of successful programs and museum exhibits. It has also produced instructional materials that incorporate recognition of national standards, and it is attempting to review the quality of existing space science educational materials. Along these lines the support network has developed a Space Science Education Resource Directory that allows educators to browse a wide range of electronic resources. In general, NASA education efforts are improving because of the increasing sophistication of people involved in the EPO efforts of the Space Science Enterprise and the Earth Science Enterprise.

At the same time, there are some issues that clearly have to be addressed. In particular, current efforts by the support network to engage the scientific community need to be expanded. The active participation by scientists in science education in schools is viewed by many as crucial to the long-term improvement of science education.[27] The OSS evaluation report points out that the culture of science continues to impede the involvement of scientists in EPO activities and that more needs to be done to bridge the gap between science and science education. The workshops for scientists on science education developed by the Space Science Institute (one of the brokers) are a good start, but the problem of how to effectively engage scientists and their institutions in science education is still an open issue.

Other activities can also bring scientists and educators together, building trust and forming the basis for long-term partnerships.[28] For example, regional planning meetings sponsored by the support network could

---

[26] S.B. Cohen and J. Gutbezal. 2001. *Office of Space Science, Education, and Public Outreach January 2000–May 2001 Final Report.* NASA, Washington, D.C. Available online at <http://spacescience.nasa.gov/education/resources/evaluation/index.htm>.

[27] See, for example, NSF, 1997, *Foundations: The Challenge and Promise of Science Education Reform,* NSF 97-76. NSF, Arlington, Va.; NRC, 2000, *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning,* National Academy Press, Washington, D.C.; R.E. Lopez and T. Schultz, 2001, Two revolutions in K-8 science education, *Physics Today,* September, pp. 44–49.

[28] NRC, 1997, *Science for All Children,* National Academy Press, Washington, D.C.; NRC, 1996, *The Role of Scientists in the Professional Development of Science Teachers,* National Academy Press, Washington, D.C.

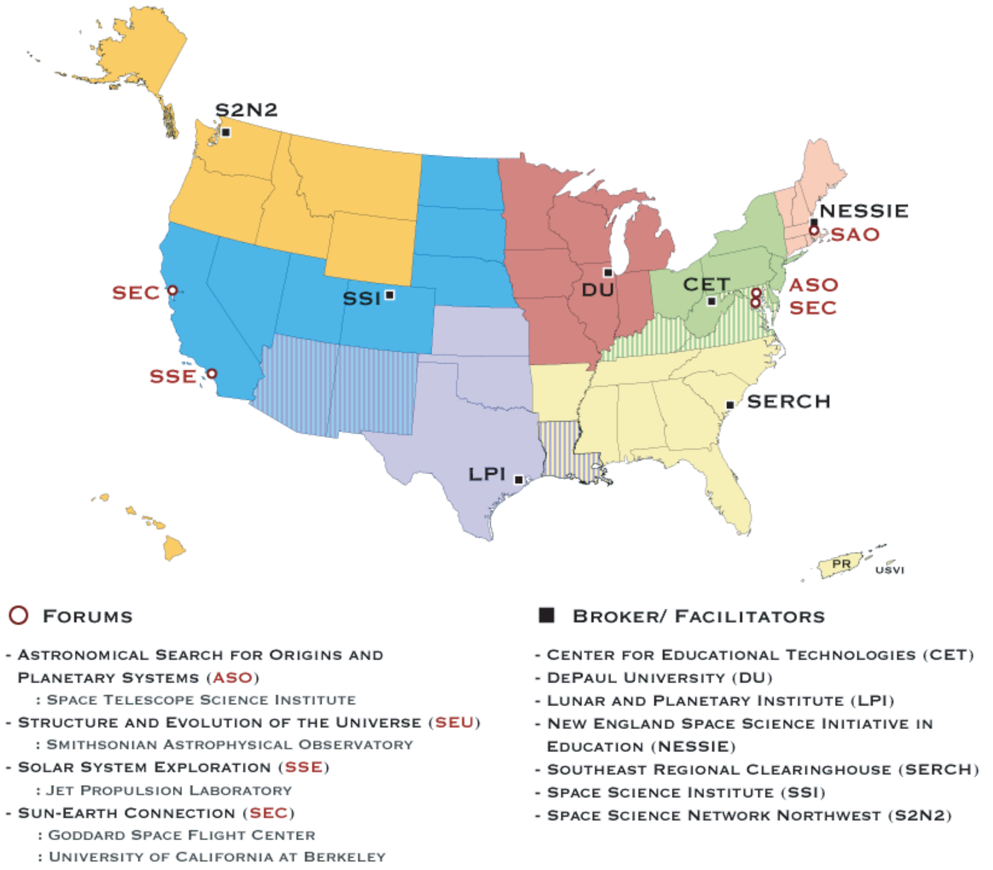


FIGURE 5.2 Geographical distribution of brokers and locations of forums. Courtesy of NASA Office of Space Science Education and Public Outreach.

develop a local set of clearly defined activities that would require limited time commitments by scientists, such as reviewing instructional materials for scientific accuracy. Such activities do not require scientists to become education experts but do allow valuable scientific input into the education process.

Another issue that should be considered is the need to better connect the solar and space physics EPO effort to other large-scale education efforts. There has been little connection between the NASA support network (or other solar and space physics education projects) and the NSF systemic initiatives (as reported to this panel by NSF program officers). Currently, hundreds of school systems are engaging in systemic change, as described in *Science for all Children*.[29] Many of these school systems have received local systemic change grants from the NSF or are involved with the various NSF state, urban, or rural systemic initiatives (see Box 5.3).

These efforts in systemic change are much broader than the relatively narrow focus of the solar and space physics community. Yet new NSF programs, such as Math and Science Partnerships and Centers for Learning and Teaching, are major initiatives with which the developing solar and space physics education infrastruc-

---

[29] NRC. 1997. *Science for All Children*. National Academy Press, Washington, D.C.

BOX 5.3 SYSTEMIC REFORM

The premise of systemic reform is that the system as a whole must be examined and addressed in order to achieve a lasting change in science education. Systemic reform, in turn, has been the driver of some of the biggest education initiatives undertaken by the National Science Foundation. While many people argue about the term "systemic reform," the consensus is that true systemic reform impacts every student and teacher in a school system.

In 1991, the NSF launched a new set of programs: the State, Rural, and Urban Systemic Initiatives (referred to as SSI, RSI, and USI, respectively). These provided large blocks of funds to the appropriate entities (depending on the program) to undertake major reform efforts in K-12 science and mathematics education. The reform efforts varied considerably, but many have had lasting effects, creating infrastructure that could benefit solar and space physics efforts in science education.[1]

Another NSF effort was the local systemic change (LSC) program. LSC programs received large grants from the NSF to pay the professional development costs of introducing standards-based instructional materials, with the rest of the expenses to be paid for by the school districts. Many LSC programs utilized a well-documented model for systemic reform described in the NRC publication *Science for All Children*.[2] Such projects created districtwide science programs that could be valuable partners for solar and space physics efforts.

The newest initiatives in large-scale science and mathematics reform are the NSF-funded Centers for Teaching and Learning and the Math and Science Partnerships. These large collaborations represent additional opportunities for the solar and space physics community to provide resources and expertise for science education throughout the country. In fact, NSF's request for proposals for the Math and Science Partnerships emphasizes that mathematicians, scientists, and engineers accept vital roles in this effort to impact the teacher workforce and to work with teachers and administrators to substantially improve student achievement.

As with the previous systemic reform efforts, these programs will create infrastructure that can be used by the solar and space physics community, which should make an effort to connect its large projects (like mission EPO efforts) to broader efforts, as well as help to encourage and connect individual solar and space physicists who want to participate in these new programs.

---

[1] W. Clune. 1999. *Toward a Theory of Systemic Reform: The Case of Nine NSF Statewide Systemic Initiatives*. National Institute for Science Education, Madison, Wisc.

[2] NRC. 1997. *Science for All Children*. National Academy Press, Washington, D.C.

ture should collaborate. The support network and mission EPO projects should contact these new centers and develop collaborations with them. In fact, the newly established Math and Science Partnerships program at the NSF specifically calls for involving science faculty in K-12 science education. The support network should make contact with groups preparing such proposals and offer them solar and space physics resources and expertise.

Another area of EPO support is small grants or supplements to individual investigators. These education supplements give individual scientists the resources to get involved in local educational activities. However, many scientists do not have contacts in the education and outreach arena, and many who would like to become more involved in education are not certain how to go about it. Also, the proposal process and evaluation requirements represent significant obstacles for scientists who are not education experts. These issues were clearly identified in the OSS evaluation report.

The NASA EPO program should encourage NASA-funded scientists to participate in existing education and public outreach programs. Further, such participation should not be limited to sharing details of specific research programs. As discussed in Box 5.4, there are many valuable educational programs in which a PI could be involved as a participant, but not necessarily in an oversight capacity.

An excellent example of an existing program is Project Astro, which was developed by the Astronomical Society of the Pacific with a start-up grant from NSF. The purpose of the program is to pair astronomers and teachers at a number of sites around the country. These sites require a modest level of funding to support a part-time coordinator and materials for workshops to train the astronomer and teacher partners. While the NSF provided start-up funds for Project Astro, at present all Project Astro sites must secure their own funding. Funding sources now include private donations, grants from

BOX 5.4 LARGE SCALE VERSUS SMALL SCALE

The solar and space physics community can interact with, and support, the science education community on many levels. At the large-scale end of the spectrum are the big-mission education and public outreach efforts, NASA's Office of Space Science support network of forums and brokers, and National Science Foundation-funded projects like the Centers for Learning and Teaching and the Math and Science Partnerships. These projects can have the greatest impact by aligning themselves to national and state standards and producing resources that can be widely disseminated and used by others.

At the other end of the spectrum, small-scale efforts, often by individual scientists, also can be very valuable. Individual scientist-teacher partnerships can result in very rich experiences for both sides, and the nationally recognized Project Astro has made such individual partnerships the foundation of its effort.

Better support is needed, though, for individual scientists who want to contribute but don't know where to begin. Principal investigators who are applying for education supplements should be able to choose from a menu of ready-made items complete with how-to information.

Another important point to consider is that some scientists who become involved in small-scale efforts may decide to make a career change and become educators, thus bringing to the profession a solid understanding of the science. Providing scientists with the opportunity to make local contributions to education will help to renew talent within the solar and space physics community as well as provide real support to teachers and students across the country.

---

education/outreach programs in NSF or NASA, or state education funds.

Contributing a small amount of funding to solar and space physicists to be used for the enrichment of their own communities is an excellent way to encourage participation in and success of such programs. The supplemental funding could be administered in much the same way as the NSF's successful REU program, where a simple letter to the program officer specifies the way the money will be spent, and the decision to fund the request is made by the program officer.

The intent of the forum/broker structure of the OSS education program was to provide a connections service, with forums connecting science to educational resources and brokers connecting those resources and individual scientists to schools, with funding from small EPO grants. However, the structure has been less than ideal owing to confusion about roles and appropriate activities. The end result, then, is that the EPO supplements are typically small and not highly leveraged—a prime goal of the overall program. Through larger EPO proposal opportunities ($100,000 to $200,000 per year)—smaller than mission scale but larger than research grant supplements—a more efficient, integrated, and leveraged suite of activities could be developed that takes advantage of the EPO expertise that exists in a number of locations across the country.

Over the past decade, the OSS took a major step forward when it allocated significant resources to be used for public outreach and education. By providing budgets that amount to 1 or 2 percent of total funding for mission proposals, OSS has shown a tangible commitment to the importance of supporting highly leveraged science education and literacy. Several of these mission-related EPO efforts, among them the ISTP and IMAGE efforts, have had considerable success. And in general there has been a good degree of interaction between the support network and the mission EPO efforts; they really can be viewed as complementary parts of the overall NASA EPO effort (see Box 5.5).

Support for NASA EPO efforts should be continued and at the same time improved. One area of improvement could be the forum and broker structure. As mentioned above, the structure in place has led to confusion about roles both within the OSS education program and with other well-developed educational networks in the space sciences (for example, the NASA Space Grant program). The success of the current approach varies widely across the country, depending on the activity of the regional brokers and proximity to forums. The panel recommends an open review of this structure, from which lessons can be learned that will lead to an improved education program for OSS.

The last area of NASA EPO the panel wishes to address is a new initiative begun in 2000 within OSS—the Minority University Initiative (MUI). This program solicited a broad range of proposals from minority-serving universities to expand space science education

BOX 5.5 THE ISTP EDUCATION AND PUBLIC OUTREACH PROGRAM

The International Solar-Terrestrial Physics (ISTP) program was for many years the flagship effort of the solar and space physics community. ISTP also had a very successful education and public outreach (EPO) program, and it took advantage of space weather events like the January 1997 magnetic storm to get the message out to the public.

A key aspect of this success was that ISTP involved communications professionals at the core of its EPO effort. This led, for example, to an increasing number of press conferences at American Geophysical Union meetings. Numerous products like the Dynamic Sun CD were produced utilizing ISTP images from SOHO and Polar. ISTP in collaboration with the Space Science Institute (in Boulder, Colorado) even produced a small traveling museum exhibit, the Space Weather Center, about the space environment.

But resources are not useful if they do not get into the hands of teachers. The ISTP program ran several teacher workshops, basing the activities on an understanding of the nature of good professional development,[1] and teachers responded very positively to the workshops, which provided them with materials and strategies that they could actually use in their classrooms.

---

[1] Described in S. Loucks-Horsley, P. Hewson, N. Love, and K. Stiles, 1998, *Designing Professional Development for Teachers of Science and Mathematics*, Corwin Press, Thousand Oaks, Calif.

and outreach in traditionally underserved communities. The kinds of projects funded range from professional development for teachers in space science topics, undergraduate research programs, and even the creation of new space science degree programs. As described in the introduction to this report, Hispanics, African-Americans, and Native Americans will provide the greatest growth in college enrollments (as well as the greatest growth in the general population), and it is imperative that increasing numbers of students from these groups go into science.

While it is too early to tell what the impact of this program will be, NASA's OSS should be commended for launching it. This targeted initiative promises to identify effective mechanisms for diversifying the space science enterprise, as well as to directly impact those participating in the program. Many aspects of the OSS initiative served as a model for a similar diversity initiative in the Geosciences Directorate of the NSF. While the geosciences are broader than solar and space physics, the solar and space physics community should take advantage of this opportunity to partner with the NSF to increase diversity in the field.

To its credit, the support network has also been proactive in building contacts with minority professional societies like the Society for the Advancement of Chicanos and Native Americans in Science. And there are new collaborations between the support network and MUI programs, such as a multifaceted set of activities in El Paso, Texas, to take advantage of the presence of the Space Weather Center at the Insights El Paso Science Museum. That partnership includes the Sun-Earth Connection Education Forum, the Space Science Institute broker, Insights, and the MUI project at the University of Texas at El Paso. The panel strongly encourages the support network to continue to find ways to partner with the MUI projects. It also urges the solar and space physics programs at NSF to support similar initiatives to broaden participation in solar and space physics.

MUSEUMS, THE WEB, NEWSPAPERS, AND OTHER OUTREACH

Solar and space physics events and discoveries lend themselves exceptionally well to use of the news media to educate the public and disseminate information (see Box 5.6). Television news is one of the most important avenues of communication for the average American. It demands video action segments, not just commentary and still photos. NASA-related missions yield action sequences of solar phenomena as natural by-products of scientific investigations. NASA has done a commendable job of rendering products such as time sequence "movies" collected by instruments on the SOHO and TRACE spacecraft into forms suitable for television broadcasts. These sequences appear regularly on TV news programs produced both by the networks and by local news stations. Animation sequences derived from space physics modeling computations, which show the response of the terrestrial magnetosphere to solar-generated interplanetary disturbances, can also have a strong impact, although they are still less available. Visualiza-

BOX 5.6 SOLAR PHYSICS: CASE STUDY FOR PUBLIC OUTREACH

An important aspect of improving science education is improving public awareness of and appreciation for current science exploration via the popular media and informal education.

Solar physics has gained widespread attention in the popular media over the past 6 years through publicized science results, and through efforts of the scientific community to communicate to the public. Solar physics is well suited to television, print, and Web media because of the striking visual appeal of the data and the visibility of the target. However, successful public outreach through the media requires coordinated activity.

Even though organizations such as NSF and NASA have very capable press offices, press officers themselves are not in a good position to identify the newest results, and individual scientists in turn are typically not good at identifying which stories will easily capture the public eye. NASA's Office of Space Science now funds a scientist at the 25 percent level as a press liaison who surveys current developments in solar and heliospheric physics and identifies interesting new results for press attention; this approach has yielded significant media coverage, and it has enhanced public awareness of solar physics phenomena.

Education programs (both formal and informal) have been shown to be successful when properly executed and coordinated with other aspects of a large program. The Yohkoh Public Outreach Project (YPOP) was funded for several years at the one-full-time-equivalent level and developed a superior Web site, a set of teacher workshops, and lesson plans for use in the classroom.

The Web site includes accessible images, student activities, and a guestbook that closes the feedback loop and allows fine-tuning of the site. The YPOP Web site (at http://www.montana.edu/YPOP) is widely used for K-12 homework, as an extracurricular resource, and as a home-schooling aid. Two important lessons demonstrated by YPOP's success are that verification and feedback are essential to the success of outreach projects, and that significant outreach projects require significant resources. YPOP achieved substantial leveraging of existing resources and was also very well funded compared with many NASA mission-level education and public outreach projects. Moreover, the resources developed continue to have an impact even though the project is completed.

tions produced from data gathered on space physics missions such as Polar and IMAGE vividly illustrate the impact of space physics phenomena on the terrestrial environment.

The dramatic effects of space physics phenomena on Earth and on artificial systems ensure that news of space physics appears regularly in the mass media, helping to teach the public about the importance of basic research on the Earth-Sun system in which we live. Obvious examples are the storm of January 10–11, 1997, during which Telstar 401 was lost (this event received considerable press coverage); the May 1998 failure of the Galaxy 4 satellite during a magnetic storm, when personal pagers became inoperative over much of North America; the brilliant aurora that extended to the southern states in March 2001, since which time inhabitants of that latitude have never again seen such an aurora; and the often-cited failure of the HydroQuebec electrical power grid during the geomagnetic storm of March 1989.

Pioneering investigations in the basic physics of Earth's magnetic and ionospheric environment are funded by NSF, which should develop a cost-effective capability for bringing news—including video visualizations—of its accomplishments in this area to the news media and the public. Currently the news media turn almost automatically to NASA for information and illustrations when solar events occur, thanks to the agency's work in developing and distributing imagery from recent Sun-watching spacecraft. Yet some of the most advanced imagery of the Sun is regularly obtained at NSF-sponsored observatories, where new technology such as adaptive optics is being introduced. And many computer visualizations of the space environment (such as the movies of the magnetosphere responding to the January 1997 coronal mass ejection, which were shown on CNN and CBS) have been produced at NSF-funded supercomputing centers. NSF needs to capitalize on its achievements in this area by developing techniques to more rapidly disseminate suitable illustrations from its national observatories and grantees to the news media.

The success of recent projects, such as NSF-funded museum exhibits like Electric Space and the Space Weather Center (which were also supported by NASA and other groups), demonstrates that the public finds such science compelling. The panel encourages the NSF science programs engaged in solar and space physics to broaden their ties to the NSF education programs, especially informal science and teacher enhancement. In fact, such recommendations have come from within the NSF itself.[30] Efforts in this direction should generate support for joint projects that bring the results of scientific research to the public that pays for the research.

The exceptional portrayals of solar and space physics phenomena in the recent IMAX film *SolarMax* make clear how much can be done for public information and education through the IMAX medium. The funding agencies should carefully follow trends in the IMAX industry, where digital IMAX is expected to be introduced and to sharply reduce production costs, just as the number of IMAX-capable theaters in science museums and other venues expands worldwide. The subject matter of solar and space physics lends itself well to this medium.

A total eclipse of the Sun is one of the most dramatic phenomena in nature. Such eclipses were once seen only by the fortunate few who lived along the paths of totality or could afford to travel to suitable viewing places. Now, satellite television and Internet Webcasts make an eclipse readily accessible once a year or so to audiences worldwide. Science museums, agencies, and amateur astronomy organizations actively use these events to introduce scientific research to students and the general public. Often, hundreds of school children convene in a museum, while thousands of others watch on computer screens in their schools as the eclipse phenomenon unfolds. The agencies should examine how they can bring these programs to an even greater proportion of the school age population, including students in communities that lack modern museum facilities.

Over the past 10 years, scientists with a particular interest in science education and literacy efforts have made significant contributions to outreach ventures aimed at the general public. Through musuem exhibits such as Electric Space and kiosks at the Houston Museum of Science, hundreds of thousands of people have had the opportunity to explore the role of plasmas in the space environment and the beautiful phenomena they display in the upper atmosphere. Through the Windows to the Universe Web site, millions of users—70 percent of them precollege students—have explored the Earth and space sciences in directed study as well as individually. At the Adler Planetarium in Chicago space scientists have contributed to a Sun-Earth Connections sky show and provided scientific animations to kiosks in public galleries. These informal science education efforts should continue to be supported (see Box 5.7).

5.6 ADDRESSING THE NEEDS: MAJOR RECOMMENDATIONS AND DISCUSSION

Throughout this report, the panel has highlighted a number of issues that it considers to be important. Having considered the various priorities involved, it has developed a set of critical recommendations for the next decade. Implementing the panel's recommendations will require that approximately $23.5 million be spent over the next decade, with no more than $3.3 million to be spent in any given year. The panel considers this to be a relatively modest investment in the future health of solar and space physics. It is confident that the recommendations outlined below, if implemented, will have a very significant effect not only on the field but also on the science education infrastructure at all levels.

Recommendation 1. A program of "bridged positions" should be established that provides partial salary support, startup funding, and limited research support for four new faculty members per year for 5 years, yielding 20 new faculty lines in solar and space physics at U.S. universities over the next decade. This should be matched with an increased emphasis on solar and space physics research and hardware development at colleges and universities.

Despite the natural interest that students have in space, solar and space physics cannot fulfill its potential to contribute to national educational goals unless the field is more widely represented in colleges and universities. It is essential that the long-term trend of fewer solar and space physics faculty at universities be reversed and that smaller institutions, or institutions that serve populations that are underrepresented in science, have access to faculty who can inspire and motivate students. The panel therefore calls for the creation of a set of "bridged positions," where agencies will provide partial support for new faculty lines at academic institutions.

---

[30] NSF. 1997. *Geoscience Education: A Recommended Strategy,* NSF 97-171. NSF, Arlington, Va.

BOX 5.7 SOME SOLAR AND SOLAR PHYSICS EDUCATION-RELATED WEB SITES

Formal Education Materials

XSPACE UCLA—http://www-ssc.igpp.ucla.edu/ssc/software/xspace.html
SOHO Lesson Plans—http://sohowww.nascom.nasa.gov/explore/lessons/
Standford Solar Activities—http://solar-center.stanford.edu/activities.html
MIT's Teal Project—http://web.mit.edu/jbelcher/www/anim.html
List of OnLine Texts—http://www.oulu.fi/~spaceweb/lib/education.html

Public Outreach Sites

The Lion Roars—http://science.nasa.gov/ssl/pad/sppb/index_Edu.html
Our Dynamic Sun—http://www-istp.gsfc.nasa.gov/exhibit/dynamic.html
From Stonehenge to Satellites—http://www-istp.gsfc.nasa.gov/exhibit/stonehenge.html
Living in the Atmosphere of the Sun—http://www-istp.gsfc.nasa.gov/exhibit/main.html
Space Update—http://earth.rice.edu/connected/space_weather.html
Space Weather Center—http://www.spacescience.org/SWOP/Exhibits/Mini_Exhibit/
Windows on the Universe—http://www.windows.ucar.edu/spaceweather/
Mission to Geospace—http://www-istp.gsfc.nasa.gov/istp/outreach/
Today's Space Weather—http://www.spaceweather.com/

Animations/Movies/Applets

Polar Aurora—http://www.gsfc.nasa.gov/topstory/20011025aurora.html
Comet Hitting the Sun—http://www.gsfc.nasa.gov/gsfc/spacesci/pictures/soho/plungefasts.mov
MIT's Teal Project—http://web.mit.edu/jbelcher/www/anim.html

Collections of Links

Space Weather Resources—http://space.rice.edu/ISTP/
Glossary of Terms—http://www-ssg.sr.unh.edu/index.html
List of Educational Resources—http://www.oulu.fi/~spaceweb/lib/education.html

The panel believes that support for these positions needs to provide 5 years of half-time salary, a small annual research support fund for travel and undergraduate researchers, and access to NASA and/or NSF project resources (such as guest investigator status on a mission). The academic institution would provide the remaining salary, and the agreement would be outlined in a memorandum of understanding. The 5-year time scale would take the individual to tenure, which the panel believes is crucial, so a 3-year program would not be effective.

It is important that these bridged positions not be restricted to particular types of colleges or universities. Either they should be open to both graduate and undergraduate institutions, public and private, minority-serving and otherwise, or a set number of positions should be allocated to each kind of institution. They should not be restricted to, say, increasing the size of already large programs or to starting programs at institutions where no solar and space physics programs currently exist. In fact, because solar and space physics is becoming a distributed enterprise, with many investigators having access to facilities and laboratories (spacecraft data sets, etc.), solar and space physics research programs can produce world-class results without the need for enormous local investments to build laboratories. Accordingly, the science might be very appealing to smaller institutions, especially those serving minority populations.

In addition, federal agencies should prepare to allocate a greater fraction of solar and space physics resources to colleges and universities. They should help support new university-based groups by offering more opportunities for break-in projects such as rocket and balloon projects, which lower the barrier for participation. They should ensure that university groups are major participants in ground-based instrument projects, with responsibilities for some aspects of the hardware (perhaps subsystems). Only in this way will the decline in university-based groups (especially hardware groups) be arrested.

It is expected that the new bridged faculty positions will gain broader visibility for solar and space physics in introductory courses for all students (as well as for science majors) and will bring more opportunities for undergraduate research, which will increase interest in scientific careers and boost the number of undergraduate science majors. Examples of such positions actually exist: At Utah State University and Montana State University, federal funds have been leveraged to create new tenure-track positions; the Thomas Jefferson Accelerator Facility, facing a similar issue in experimental nuclear physics, created a formal set of bridged positions that have proved attractive to many universities.

Recommendation 2. Federal agencies that fund solar and space physics should set aside funds to support undergraduate research in solar and space physics, either as a supplement to existing grants or as stand-alone programs.

The natural corollary to an increased presence in colleges and universities is an increase in the support provided to undergraduate research. A $5,000 grant can support an undergraduate for a summer or give partial support over the entire school year. The panel calls for $200,000 per year to be set aside to support undergraduate students. The 40 students supported would become a valuable source of graduate students. Over a 10-year period, as many as 400 future technical professionals could be fostered. Often such funds can generate additional matching funds from institutions that, when they gain external support, come to recognize the value of undergraduate research and support it with their own funds.

Recommendation 3. Three resource development groups should be funded over the next decade to develop educational resources (especially at the undergraduate level) needed by the solar and space physics community, to disseminate those resources, and to provide other services to the community.

Solar and space physics provides many opportunities for the demonstration and application of fundamental physical processes like electromagnetic radiation, charged particle motion, electromagnetic induction, and wave propagation. But expansion of the field into education will require innovative approaches. The general interest in space missions allows linking solar and space physics phenomena to the underlying physics, much of which is a component of curriculum standards (see Box 5.8). Advances in information technology provide opportunities to develop learning tools that combine data, images, animations, and interactive applets. Such tools not only bring the subject alive but also make learning an active experience for students, and using them in mission EPO activities would enhance their impact.

The decline in science education, particularly at the undergraduate level and in physics, is well documented. Solar and space physics could play an important role in revitalizing physics and astronomy education, particularly at college, by using people's natural interest in space to reach a wide range of students. Two arenas where SSP could have a national impact are introductory physics (a requirement for many majors) and introductory astronomy (one of the most popular general education courses for nonscience majors).

If we wish to improve introductory physics and astronomy courses that reach large numbers of students, we must have high-quality instruction materials that show how basic physics and astronomy are applied to concepts from space physics. Funding groups to collect, develop, and adapt materials for a national audience would enhance the availability and effectiveness of these materials nationwide. Introductory astronomy in particular could benefit from the availability of these materials since only 20 percent of those teaching Astronomy 101 have an astronomy degree, and most do not consider themselves astronomers.[31] Such faculty already are using materials developed by others.

The panel has focused on these two introductory college courses to maximize the national impact. Nevertheless, similar tools could be usefully adapted either for more advanced undergraduate courses (e.g., physics for majors) or for high school. For example, materials developed for introductory astronomy for nonscientists are often appropriate for science-gifted middle school-

---

[31] A. Fraknoi. 1996. *Astronomy Education: Current Developments, Future Coordination.* Astronomical Society of the Pacific Conference Series, Vol. 89, J. Percy, ed. Astronomical Society of the Pacific, San Francisco, Calif.

BOX 5.8 SOLAR AND SPACE PHYSICS CURRICULAR CONNECTIONS

Introductory Physics

Magnetic and electric fields
Charged-particle motions, currents
Plasmas
Atomic physics: ionization, excitation, radiation, recombination

Introductory Astronomy

The Sun and stars (interior, atmosphere, corona, solar wind)
Planetary magnetic fields (implications for interiors, surfaces, and atmospheres)
Terrestrial space weather

ers. Furthermore, with a little repackaging, Web-based material developed for formal classes can be valuable in informal education arenas (e.g., lifelong learning via the Web, museums, and planetariums). For such tools to be of educational value, however, space scientists will need to team with experienced educators to ensure that the tools are effective and aligned with appropriate curriculum standards.

At the graduate level, solar and space physics is taught at only a dozen universities, and often coursework covers only part of the field. Advances in information technology could allow distance learning to link students and postdoctoral students to the expertise that resides at different universities. Courses could be offered by distance learning (either over the academic year or during the summer), with several faculty from a variety of institutions, and would allow small research groups (or start-up programs with bridged positions) to offer solar and space physics courses to their students.

To achieve the above goals, the panel calls for the establishment of two or three competitively funded resource development groups (RDGs), the exact size and structure of which remain to be defined. (To have a substantial impact they probably should be much larger than typical single-investigator research grants.) The RDGs would develop instructional materials, do research on teaching and learning (including guiding graduate students to space science education research), and disseminate teaching resources. In this sense what is envisaged is similar to recent NSF Centers for Learning and Teaching. The RDGs also would develop and provide services to the solar and space physics community (such as workshops on a variety of subjects, special summer graduate and undergraduate schools in the field, and coordination/development of shared academic-year graduate and undergraduate courses) and could provide professional development for scientists involved in education issues at all levels. It is expected that the RDGs would be funded at approximately $500,000 per year.

Recommendation 4. Current K-12 education and public outreach (EPO) efforts should be continued. However, there should be a careful evaluation of lessons learned over the past few years, particularly regarding the involvement of scientists in EPO activities, as well as increased coordination of NASA EPO efforts with other large projects in science education reform, especially NSF initiatives.

Although considerable progress has been made as a result of solar and space physics K-12 education and outreach efforts, aspects of the current system at times prove unwieldy. The panel believes this is the inevitable result of embarking on a new venture. It commends the agencies, particularly NASA OSS, for their commitment to education and outreach and the institutionalization of EPO efforts as part of the mission of science. Given the experience of the past few years, the agencies, principally NASA OSS, should be able to evaluate the impact of solar and space physics on science education and identify successes and barriers in the quest for a meaningful contribution. Those lessons learned should be widely disseminated, and all evaluation reports should continue to be made public, perhaps with better information on their availability.

It seems quite clear that improvements can be made in some areas. First, engaging the scientific community in science education continues to be difficult. Much of this difficulty stems from the nature of science and from a reward system that discourages scientists from participating in EPO activities. The NASA EPO initiative has made strides toward countering this trend, but more needs to be done. Additional mechanisms should be developed that allow scientists to contribute to science education without becoming science education experts themselves. Examples of proven activities that individuals can conduct should be developed, along with mechanisms that can better link scientists to their local science education community.

The connection between NASA EPO efforts and NSF-funded efforts directed to systemic change could also be strengthened. Projects such as the National Science Resources Center LASER initiative and the work of Project Impact in New England have essentially no connection to the solar and space physics infrastructure. NASA EPO efforts seem largely unknown to most systemic reform projects in the country, although there are notable exceptions. In June 2002, NASA's Support Network project held a very successful education conference in Chicago; nonetheless, that conference was not attended by leaders of the major NSF-funded science education efforts.

Urban, rural, and local and state systemic change initiatives represent a tremendous opportunity to leverage resources. If NASA EPO efforts connect with a single, moderate-size project with 100 middle-school science teachers serving 15,000 students, and that project incorporates solar and space physics content and the appropriate professional development into the core of instruction, the impact will be enormous relative to the investment. Building links to such efforts does not require additional funds, but it does take a commitment to reach out beyond solar and space physics to the general science education community, especially to leverage NSF efforts, where hundreds of millions of dollars are being invested.

New initiatives from the NSF, particularly the Centers for Learning and Teaching and the Math and Science Partnerships, should provide fertile soil for NASA's Support Network and other solar and space physics EPO projects. These new initiatives will be looking for partners. The Support Network and other EPO projects have created excellent resources that can prove very valuable to such initiatives. Moreover, the evaluation of the impact of the Support Network and the lessons learned should also prove extremely valuable to others who are trying to engage diverse scientific communities in support of science education.

Finally, the panel commends efforts by NASA, NSF, and NOAA to increase diversity in solar and space physics. While this is not strictly a K-12 issue, leadership in this area has emerged from the K-12 education and outreach effort. The panel urges that activities such as the NASA Minority University Initiative and the NSF Diversity in Geosciences program continue to be funded, and that those projects that succeed in engaging students in the solar and space physics enterprise be replicated in other communities.

Appendixes

A

Statement of Task

Background: The last integrated strategy for solar and space physics was released by the NRC in 1995. Since that time, there have been dramatic scientific developments and a significant evolution in relevant federal programs. In the space arena these developments stem from the launches and successful operation of the Wind, Geotail, SOHO, Polar, FAST, ACE, TRACE, IMAGE, and Cluster-II missions. These missions have helped revolutionize solar physics, provide a new level of understanding of important processes in space plasma physics, and create a new basis for characterizing and predicting space weather. Over the same period, the relevant federal agencies have taken steps to build on the new level of scientific progress by embarking on new efforts such as the National Space Weather Program, the Relocatable Radar (formerly the Polar Cap Observatory), and Living With a Star. Furthermore, the NSF Geospace Environment Modeling (GEM) program has initiated its second and third campaigns, the international Super Dual Auroral Radar Network (SuperDARN) has established effective collaboration among a large number of high frequency radar programs; and the community-wide Solar, Heliospheric, and Interplanetary Environment (SHINE) initiative has spawned a number of important activities related to the National Space Weather Program. As a consequence of all these developments, the preparation of a comprehensive scientific assessment and strategy for the field of solar and space physics that looks across the interests of all agencies, both ground- and space-based, is especially timely.

Plan: The study will be organized in a manner similar to the decadal survey that is regularly conducted by the astronomy and astrophysics community. The Committee on Solar and Space Physics (CSSP) will establish a 14-person survey committee to carry out the study with input from five panels, each of which will have approximately 10 members. Most CSSP members will serve either on the survey committee or the panels, with additional membership drawn from the relevant research communities.

The study will generate consensus recommendations from the solar and space physics community regarding a systems approach to theoretical, ground-based and space-based research that encompasses the flight programs and focused campaigns of NASA, the ground-based and basic research programs of NSF, and the complementary operational programs of other agencies such as NOAA, DOD, and DOE. During this study, the community will survey solar and space physics and recommend priorities for the decade 2003–2013. Attention will be given to effective implementation of proposed and existing programs and to the human resource aspects of the field involving education, career opportunities, and public outreach. Promising areas for the development of new technologies will be suggested. A minor but important part of the study will be the review of complementary initiatives of other nations in order to identify potential cooperative programs.

An important aspect of the study's consideration of operational programs will be an assessment of how the research programs of NASA and NSF can serve both to provide the operational tools of agencies such as NOAA and DOD and to provide training for future expert staff for those agencies. The study will consider how the science of solar and space physics can lead to new forecast tools and products that have the potential of making the space weather program more operational, and it will identify appropriate next steps to accomplish the transition from research to operations.

Three of the five panels will be organized around interdisciplinary science themes:

- magnetosphere-ionosphere-atmosphere interactions,
- solar-wind magnetosphere interactions, and
- solar and heliospheric physics.

Each of these panels will consider theory and computation as well as ground-based and space-based research. The first two panels will cover both terrestrial and planetary objectives. The three science panels will be complemented by two cross-disciplinary panels:

- theory, computation, and data exploration and
- education and society.

The survey committee will be responsible for preparing a summary report. The reports of the study panels along with the summary report will be published by the National Research Council. One important goal of these reports is to address the scientific foundation and priorities for the implementation of major NASA programs such as Living With a Star, Solar-Terrestrial Probes, Solar Probe, and Interstellar Probe and major NSF facilities such as the Relocatable Radar.

In conducting its work, the CSSP would draw on an extensive history of prior studies performed by the Space Studies Board, including

- *Astronomy and Astrophysics in the New Millennium* (Astronomy and Astrophysics Survey Report) and *Astronomy and Astrophysics in the New Millennium: Panel Reports* (2000). (Survey and panel reports are joint projects of the SSB and the NRC Board on Astronomy and Astrophysics.)
- *Readiness for the Upcoming Solar Maximum* (1998).
- *Ground-Based Solar Research: An Assessment and Strategy for the Future* (1998).
- *Scientific Assessment of NASA's SMEX and MIDEX Space Physics Mission Selections* (1997).
- *An Assessment of the Solar and Space Physics Aspects of NASA's Space Science Enterprise Strategic Plan* (1997).
- *Space Weather: A Research Briefing* (Web report, 1997).
- *A Science Strategy for Space Physics* (1995).

B

Acronyms and Abbreviations

AAS	American Astronomical Society
AC	Auroral Cluster
ACE	Advanced Composition Explorer
ACR	anomalous cosmic ray
ACRIM	Active Cavity Radiometer Irradiance Monitor
AFOSR	Air Force Office of Scientific Research
AGU	American Geophysical Union
AIA	Atmospheric Imaging Assembly
AIM	Aeronomy of Ice in the Mesosphere
A-I-M	atmosphere-ionosphere-magnetosphere
AISRP	Applied Information Systems Research Program
AMISR	Advanced Modular Incoherent Scatter Radar
AMR	adaptive mesh refinement
ASIC	application-specific integrated circuit
AST	Division of Astronomical Sciences (NSF)
ATM	Division of Atmospheric Sciences (NSF)
ATST	Advanced Technology Solar Telescope
AU	astronomical unit (~150,000,000 km)
BBF	bursty bulk flow
BBSO	Big Bear Solar Observatory
C-A-I-M	comparative atmospheres, ionospheres, and magnetospheres
CASI	Cornell All-Sky Imager
CCD	charge-coupled device
CCMC	Community Coordinated Modeling Center
CDH	command and data handling
CEDAR	Coupling, Energetics, and Dynamics of Atmospheric Regions
CGRO	Compton Gamma Ray Observatory
CHSSI	Common High-Performance Computing Software Support Initiative
CIR	co-rotating interaction region
CME	coronal mass ejection
CMOS	complementary metal oxide substrate
C/NOFS	Communications/Navigation Outage Forecast System

COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
COTS	commercial off-the-shelf
CRRES	Combined Release and Radiation Effects Satellite
DBC	Dayside Boundary Constellation
DMSP	Defense Meteorological Satellite Program
DOD	Department of Defense
DOE	Department of Energy
DQE	detective quantum efficiency
DS1	Deep Space 1
DSN	Deep Space Network
EIT	Extreme Ultraviolet Imaging Telescope
ELF	extremely low frequency
ENA	energetic neutral atom
EPO	education and public outreach
ESA	European Space Agency
EUV	extreme ultraviolet
eV	electron volt
FASR	Frequency-Agile Solar Radiotelescope
FAST	Fast Auroral Snapshot Explorer
FLR	field line resonance
FOV	field of view
FPGA	field programmable gate array
FUV	far ultraviolet
GCR	galactic cosmic ray
GEC	Geospace Electrodynamic Connections
GEM	Geospace Environment Modeling
GEO	Directorate of Geosciences (NSF)
GGCM	Geospace General Circulation Model
GIC	geomagnetically induced current
GOES	Geostationary Operational Environmental Satellite
GONG	Global Oscillations Network Group
GProbes	Geospace Probes
GPS	Global Positioning System
GRB	gamma-ray burst
GSFC	Goddard Space Flight Center
GSRI	Geospace System Response Imagers
GUVI	Global Ultraviolet Imager
G-Z	Giacobini-Zinner
HAO	High Altitude Observatory
HESSI	High Energy Solar Spectroscopic Imager
HF	high-frequency
HMI	Helioseismic and Magnetic Imager
HPCC	High Performance Computing and Communications
HST	Hubble Space Telescope
HT	de Hoffman-Teller

IACG	Inter Agency Consultative Group
ICE	International Cometary Explorer
IE	Io Electrodynamics
IGY	International Geophysical Year
IMAGE	Imager for Magnetopause-to-Aurora Global Exploration
IMC	interplanetary magnetic cloud
IMEX	Inner Magnetosphere Explorer
IMF	interplanetary magnetic field
IMP	Interplanetary Monitoring Platform
IRAS	Interspacecraft Ranging and Alarm System
IRI	International Reference Ionosphere
ISAS	Institute of Space and Astronautical Science (Japan)
ISEE	International Sun-Earth Explorer
ISM	interstellar medium; also, Interstellar Sampler Mission
ISOON	International Solar Optical Observing Network
ISP	In-Space Propulsion
ISR	incoherent scatter radar
ISS	International Space Station
ISTP	International Solar-Terrestrial Physics program
ITAR	International Traffic in Arms Regulations
ITR	Information Technology Research
JAO	Jupiter Auroral Orbiter
JPO	Jupiter Polar Orbiter
KBO	Kuiper Belt object
keV	kiloelectron volt
LANL	Los Alamos National Laboratory
LAPD	Large Plasma Research Device at UCLA
LCAS	low-cost access to space
LEO	low Earth orbit
LISM	local insterstellar medium
LORAN	Long Range Navigation
LUF	lowest usable frequency
LWS	Living With a Star (NASA)
LWS-GID	Living With a Star Geospace Instrument Development
MagCon	Magnetospheric Constellation
MagTom	magnetospheric tomography
MAP	Mars Aeronomy Probe
MCO	Mars Climate Orbiter
MEMS	microelectromechanical system
Messenger	Mercury Surface, Space Environment, Geochemistry, and Ranging
MGS	Mars Global Surveyor
MHD	magnetohydrodynamic(s)
MIDEX	Medium-Class Explorer
MLSO	Mauna Loa Solar Observatory
MMS	Magnetospheric Multiscale
MO&DA	mission operations and data analysis
MOO	mission of opportunity

MPS	Directorate of Mathematical and Physical Sciences (NSF)
MRX	Magnetic Reconnection Experiment
MSA	Magneto-Seismology Array
MSFC	Marshall Space Flight Center
MSM	Magnetospheric Specification Model
MSO	Mees Solar Observatory
MTIDO	Mars Thermosphere Ionosphere Dynamics Orbiter
MUI	Minority University Initiative
MUF	maximum usable frequency
MURI	Multidisciplinary University Research Initiative
MWSO	Mount Wilson Solar Observatory
NASA	National Aeronautics and Space Administration
NASA/AISRP	NASA Applied Information Systems Research Program
NASDA	National Space Development Agency of Japan
NATO	North Atlantic Treaty Organization
NCAR	National Center for Atmospheric Research
NEAR	Near Earth Asteroid Rendezvous
NEP	nuclear electric propulsion
NGDC	National Geophysical Data Center
NLC	noctilucent clouds
NMP	New Millennium Program
NO	Neptune Orbiter
NOAA	National Oceanic and Atmospheric Administration
NORAD	North American Aerospace Defense Command
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NRAO	National Radio Astronomy Observatory
NRC	National Research Council
NSF	National Science Foundation
NSO	National Solar Observatory
NSWP	National Space Weather Program
ONR	Office of Naval Research
OSS	Office of Space Science (NASA)
OTH	over the horizon
OVRO	Owens Valley Radio Observatory
PASO	particle acceleration solar orbiter
PDS	Planetary Data System
PI	principal investigator
PIDDP	Planetary Instrument Definition and Development Program
PMC	polar mesospheric cloud
PPARC	Particle Physics and Astronomy Research Council
PSBL	plasma sheet boundary layer
PSPT	Precision Solar Photometric Telescope
PVO	Pioneer Venus Orbiter
R_S	solar radius
RAM	Reconnection and Microscale probe
RAO	Relocatable Atmospheric Observatory
RBM	Radiation Belt Mapper
RDG	resource development group

REU	Research Experiences for Undergraduates
RHESSI	Reuven Ramaty High-Energy Solar Spectroscopic Imager
RISE	research internships in science and engineering
ROSAT	Röntgen Satellite
RPC	rapid prototyping center
RTG	radioisotope thermoelectric generator
SAA	South Atlantic Anomaly
SAMPEX	Solar Anomalous and Magnetospheric Particle Explorer
SAT	science architecture team
SCOSTEP S-RAMP	Scientific Committee on Solar-Terrestrial Physics—STEP–Results, Applications, and Modeling Phase
SDO	Solar Dynamics Observatory
SDSC	San Diego Supercomputing Center
SEC	Sun-Earth Connection
SECID	Sun-Earth Connection Instrument Development program
SECTP	Sun-Earth Connection Theory Program
SEE	Solar Extreme Ultraviolet Experiment
SEP	solar energetic particle
SFSO	San Fernando Solar Observatory
SHINE	Solar, Heliospheric, and Interplanetary Environment
SIE	spectrometer for irradiance in the EUV
SMC	steady magnetospheric convection
SMC/CIT	Space and Missile Systems Center/DMSP Technology Applications Division (Air Force)
SMEI	Solar Mass Ejection Imager
SMEX	Small Explorer
SMI	Stereo Magnetospheric Imager
SMM	Solar Maximum Mission
SNOE	Student Nitric Oxide Explorer
SOHO	Solar and Heliospheric Observatory
SOLIS	Synoptic Optical Long-term Investigation of the Sun
SOON	Solar Optical Observing Network
SORCE	Solar Radiation and Climate Experiment
SPIS	space physics information system
SR&T	Supporting Research and Technology
STDT	Science and Technology Definition Team
STEDI	Student Explorer Demonstration Initiative
STEP	Solar-Terrestrial Energy Program
STEREO	Solar Terrestrial Relations Observatory
STP	Solar Terrestrial Probe(s) (NASA)
STSP	Solar Terrestrial Science Program
SuperDARN	Super Dual Auroral Radar Network
SWS	Solar Wind Sentinels
SXI	Solar X-ray Imager
TEC	total electron content
TES	transition edge sensors
TIGCM	Thermosphere-Ionosphere General Circulation Model
TIMED	Thermosphere Ionosphere Mesosphere Energetics and Dynamics
TRACE	Transition Region and Coronal Explorer
TWINS	Two Wide-angle Imaging Neutral-atom Spectrometers

UARS	University Atmospheric Research Satellite
UHF	ultrahigh frequency
UFL	ultralow frequency
UNEX	University-Class Explorer
UPOS	University Partnership for Operational Support
USAF	United States Air Force
USRA	Universities Space Research Association
UV	ultraviolet
VAP	Venus Aeronomy Probe
VCO	Venus Climate Orbiter
VHF	very high frequency
VLF	very low frequency
VSO	Virtual Solar Observatory
WAAS	Wide Area Augmentation System
WCI	White-light Coronagraphic Imager
WSO	Wilcox Solar Observatory