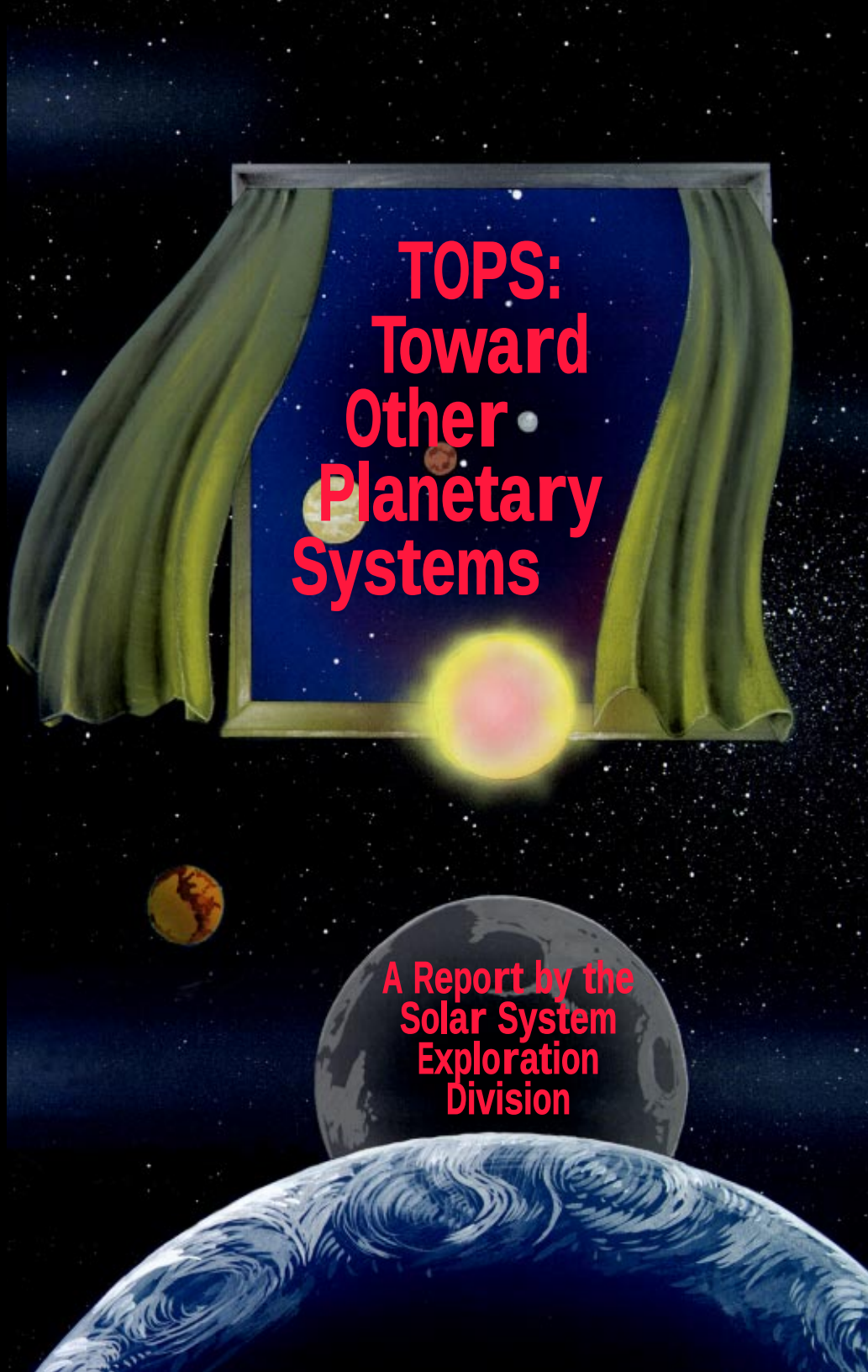
TOPS: Toward Other Planetary Systems

A Report by the Solar System Exploration Division

NASA



TOPS: Toward Other Planetary Systems

A Report by the Solar System Exploration Division

TOPS: Toward Other Planetary Systems

A Report by the
Solar System Exploration
Division

TOPS SCIENCE WORKING GROUP MEMBERS

Bernard F. Burke (Chair), Massachusetts Institute of Technology
Jurgen H. Rahe (Executive Secretary), NASA/Headquarters

Reta F. Beebe*, New Mexico State University
Michael J.S. Belton, Kitt Peak National Observatory
David C. Black, Lunar and Planetary Institute
Alan P. Boss, Carnegie Institution of Washington
Robert A. Brown, Space Telescope Science Institute
James A. Cutts, Jet Propulsion Laboratory
Edward W. Dunham, NASA/Ames Research Center
Charles Elachi, Jet Propulsion Laboratory
George D. Gatewood, University of Pittsburgh
Garth D. Illingworth, University of California, Santa Cruz
William D. Langer, Jet Propulsion Laboratory
Eugene H. Levy, University of Arizona
H. Warren Moos, Johns Hopkins University
David Morrison*, NASA/Ames Research Center
Michael J. Mumma, NASA/Goddard Space Flight Center
Joseph A. Nuth, NASA/Goddard Space Flight Center
Stanton J. Peale, University of California, Santa Barbara
Robert D. Reasenberg, Smithsonian Astrophysical Observatory
Anneila I. Sargent, California Institute of Technology
Bonny L. Schumaker, Jet Propulsion Laboratory
Michael Shao, Jet Propulsion Laboratory
Bradford A. Smith, University of Hawaii
Harlan J. Smith (deceased)*, University of Texas
Stephen E. Strom*, University of Massachusetts
Stephen P. Synnott*, Jet Propulsion Laboratory
Richard J. Terrile, Jet Propulsion Laboratory

*former members who contributed to formulation of the TOPS Program

INVITED ATTENDEES AND CONSULTANTS

Roger J. Angel, University of Arizona
John F. Appleby, NASA/Headquarters
Christopher J. Burrows, Space Telescope Science Institute
Richard W. Capps, Jet Propulsion Laboratory
William D. Cochran, University of Texas
David J. Diner, Jet Propulsion Laboratory
Christ Ftaclas, Hughes Danbury Optical Systems
William B. Gray, Jet Propulsion Laboratory
Donald N.B. Hall, University of Hawaii
Wendy Hunter, Massachusetts Institute of Technology
David W. Latham, Harvard-Smithsonian Center for Astrophysics
Douglas N.C. Lin, University of California, Santa Cruz
Robert S. McMillan, University of Arizona
Aden B. Meinel, Jet Propulsion Laboratory
Marjorie P. Meinel, Jet Propulsion Laboratory
Thomas H. Morgan, NASA/Headquarters
Kenji Nishioka, Jet Propulsion Laboratory
Bohdan Paczyñski, Princeton University Observatory
Vernon Pankonin, National Science Foundation
Alfred W. Pappano, Jet Propulsion Laboratory
Steven H. Pravdo, Jet Propulsion Laboratory
Terrence H. Reilly, Jet Propulsion Laboratory
Rex W. Ridenoure, Jet Propulsion Laboratory
Donald E. Rockey, Jet Propulsion Laboratory
Fred E. Vescelus, Jet Propulsion Laboratory
Richard W. Vorder Bruegge, Science Applications International Corporation

Stars and planetary systems form from dense interstellar clouds of molecular gas and dust grains.

These clouds contract in size because of their self-gravity, and then undergo a rapid phase of collapse, leading to the formation of a central protostar and a surrounding nebula or disk.

Protostars emit a strong wind in the form of a bipolar outflow perpendicular to the plane of their nebula, even as gas continues to fall onto the nebula.

Dust grains in the nebula accumulate into planetesimals, which form the terrestrial planets and the cores of the giant planets, while residual gas forms the envelopes of the giant planets.

Preface

In 1988, NASA's Solar System Exploration Division established a Science Working Group to formulate a strategy for the discovery and study of other planetary systems, considering both short-term and long-term aspects. The charter for this group included defining the interface between the study of planet formation processes and star formation. The range of consideration, therefore, included the search for planets, the study of their properties, and the more general study of circumstellar and protostellar material. Astrometry and imaging techniques were identified for immediate consideration, but other approaches were to be evaluated as well. Recommendations were sought for instrumental and non-instrumental research to encourage and link initiatives in observations, laboratory investigations, and theory. Finally, the Science Working Group was to identify and recommend technologies necessary to carry out the strategy.

The Science Working Group was formed in response to the gradually developing realization that a program could be formulated to study the planetary systems of other stars. The steady growth in technological development was responsible in part for this awareness. NASA sponsored a series of workshops that examined the possibilities in some detail.[1] At the same time, NASA had been supporting studies on the search for extraterrestrial intelligence,[2] and in that context, it was clear that the knowledge of where and how planetary systems formed would have a significant influence on search strategies.

In 1985, two independent efforts carried the issue of searching for other planets from the study phase to a level in which the formulation of a program could be contemplated. NASA's Solar System Exploration Division established the Planetary Astronomy Committee (PAC) to provide advice on the status and future of planetary astronomy, both for continuing the study of the solar system and for initiating the search for and characterization of other planetary systems. In the same year, the Space Science Board formally directed its Committee on Planetary and Lunar Exploration (COMPLEX) to extend its exploration strategy to planetary systems outside the solar system. In their assessment, COMPLEX was to include both the implications for our understanding of the origin and evolution of the solar system, and the status of and prospects for the development of planetary detection techniques.

The PAC report[3] was published in 1989, followed in 1990 by the COMPLEX report.[4] By 1988, however, the recommendations of both committees had already taken shape. NASA was urged to develop an interdisciplinary program for this new field of planetary science, and to convene a science working group to formulate a program for the detection and study of other planetary systems.

The present Science Working Group was able to immediately start defining the scope and goals of its deliberation. At its first meeting in April 1988, the Group concluded that studies of planet formation, planetary evolution, and circumstellar material in general were appropriately included, to prevent valuable research efforts from "falling into the crack" between astrophysical and planetary studies. Several meetings were held in 1988 and 1989 to review the state of knowledge and the prospects for future work, leading to a 1-week workshop in January 1990. The theme "Toward Other Planetary Systems" (TOPS) became the name of the workshop.

The outcome of the first TOPS Workshop was the formulation of the three-phase TOPS Program for NASA's consideration. The first phase, TOPS-0, would concentrate on ground-based approaches, combined with preparatory work for the next phase. TOPS-1 would include the development and launch of a space-based instrument. Finally, in the third phase, TOPS-2, a major instrument that could directly detect Earth-like planets would be built. In a larger context, the new initiative anticipated the recommendations that were subsequently made by the Astronomy and Astrophysics Survey Committee of the National Research Council, which identified the search for other planetary systems, including the study of their origins and evolution, as a key field of scientific opportunity for the 1990s.[5]

Subsequently, the opportunity arose for NASA to participate in building and using a second 10-meter telescope at the Keck Observatory on Mauna Kea in Hawaii. An ad hoc committee was convened in July 1990 to consider that proposal. The committee recommended unanimously that NASA seek a partnership with the Keck Observatory. Review of the implications of the partnership for the TOPS Program led to the conclusion that it would be a natural and powerful component of the TOPS-0 phase.

It became clear that an intensive study was required to work out a full rationale and detailed plan for the TOPS Program, and that a site

visit to the Keck Observatory should take place. The second TOPS Workshop, therefore, took place on the island of Hawaii in January 1992. Building on the first TOPS Workshop and on the accumulated experience of the group, the second TOPS Workshop completed the formulation of the long-range program described in this report.

REFERENCES

1. David C. Black, Ed., *Project Orion: A Design Study of a System for Detecting Extrasolar Planets*, NASA SP-436. D.C. Black and W.E. Brunk, Eds., 1980, *An Assessment of Ground-Based Techniques for Detecting Other Planetary Systems, Vols. I & II*, NASA CP-2124.
2. The US/USSR 1971 Symposium in Byurakhan, Armenia, *Communication with Extraterrestrial Intelligence*, C. Sagan (Ed.) (MIT Press, Cambridge, 1973) gives a good introduction to early work. The NASA study *The Search for Extraterrestrial Intelligence SETI*, P. Morrison, J. Billingham, and J. Wolfe (Eds.) (NASA Publication SP-419, 1977) summarized the results of a series of NASA workshops. IAU, Montreal 1979 (Papagiannis).
3. Planetary Astronomy Committee of the Solar System Exploration Division, *Other Worlds From Earth: The Future of Planetary Astronomy*, National Aeronautics and Space Administration, Washington, D.C., 1989.
4. Committee on Planetary and Lunar Exploration, Space Studies Board, Commission on Physical Sciences, Mathematics, and Applications, National Research Council, *Strategy for the Detection and Study of Other Planetary Systems and Extrasolar Planetary Materials: 1990-2000*, National Academy Press, Washington, D.C., 1990.
5. Astronomy and Astrophysics Survey Committee, National Academy of Sciences, National Research Council, *Astronomy and Astrophysics for the 1990s*, 1991.

A view of the 10-meter-diameter Keck Telescope in Hawaii, at the summit of Mauna Kea. A second 10-meter telescope, a twin of the first, will be placed to the right of the existing structure.

Executive Summary

IN RECENT DECADES, SEVERAL LINES OF SCIENTIFIC INVESTIGATION have converged to bring into sharp focus our understanding of the solar system. Intensive observations by spacecraft and other means, combined with extraordinary laboratory analytical methods and theoretical investigations, are beginning to paint an intelligible picture of our solar system's history, the mechanisms of its development, and the relationship between the formation of our Sun and its associated planets. Astronomical observations are providing important new information about the processes that give birth to stars and about the conditions in star-forming regions and around very young stars that might be conducive to establishing planetary systems. This progress leads naturally to a new line of inquiry: the discovery and characterization of planetary systems around other stars.

This report describes a general plan and the pertinent technological requirements for TOPS (Toward Other Planetary Systems), a staged program to ascertain the prevalence and character of other planetary systems and to construct a definitive picture of the formation of stars and their planets. The first stages focus on discovering and studying a significant number of fully formed planetary systems, as well as expanding current studies of protoplanetary systems. As the TOPS Program evolves, emphasis will shift toward intensive study of the discovered systems and of individual planets. Early stages of the TOPS Program can be undertaken with ground-based observations and space missions comparable in scale to those now being performed. In the long term, however, TOPS will become an ambitious program that challenges our capabilities and provides impetus for major space initiatives and new technologies, which will be accomplishments of historical significance.

RECOMMENDATIONS OF THE TOPS SCIENCE WORKING GROUP

This report follows an earlier report of the Space Studies Board Committee on Planetary and Lunar Exploration (COMPLEX), published by the National Academy of Sciences' National Research Council. The COMPLEX report set forth the objectives and overall strategy for the discovery and study of other planetary systems and for the investigation of planet formation. The TOPS Science Working Group concurs with COMPLEX's conclusions about the importance and timeliness of

this area of science. The program summarized in the pages that follow presents a set of scientific objectives that address fundamental issues: the discovery of planetary systems, and the investigation of their origin and evolution.

Several important ground-based planet-discovery programs are currently under way. These are presently supported at a low level, and this effort deserves consideration for expansion. Moreover, recent developments suggest that significant improvements may be possible in the sensitivity and range of ground-based searches for planets orbiting other stars. The TOPS-0 phase is aimed at extracting the best possible level of scientific advancement from ground-based methods. Optimistic projections for such ground-based indirect searches suggest the possibility that Jupiter-mass planets may be detectable around 100 or more nearby stars. In addition, much remains to be accomplished with ground-based telescopes using direct-detection techniques (filled-aperture and interferometric imaging) at a variety of wavelengths to investigate star- and planet-forming regions.

Recommendation

The first phase of the TOPS Program, TOPS-0, should begin as soon as possible. This phase is centered on a ground-based observing program, coupled with elements that will prepare for future space missions.

The mid-term goals of the TOPS Program require the capability to discover and study planets with masses as low as those of Uranus and Neptune around 100 or more stars in the solar neighborhood. In addition, the TOPS science objectives require direct imaging studies of some anticipated discovered planets as well as diffuse circumstellar material. These objectives will require long-term observations from specially designed instruments flown in space above the disturbing influence of Earth's atmosphere.

Recommendation

The second phase of the program, TOPS-1, is centered on a space-based mission. Mission definition should be implemented in a timely fashion, anticipating a project new start toward the end of this decade.

Observational research is but one crucial component of a program to

understand the formation, prevalence, and character of planetary systems. To advance and succeed, such a program must also be tied to vigorous and sustained conceptual, theoretical, and modelling studies of the pertinent physical processes and phenomena.

Recommendation

A continuing and vigorous program of scientific research, including theoretical and modelling studies, must be established. Such support is crucial to the success of the program.

The preceding recommendations pertain to the near- and mid-term evolution of the TOPS Program. In the long term, detailed studies of other planetary systems and star-forming regions will require conceptually new generations of scientific instruments, involving scales and technologies not yet realized, and located in space or on the Moon. These developments will take place over a sufficiently long period of time that consideration of their actual implementation lies outside the scope of this report. Both the technologies and the in-space infrastructure needed to support this level of investigation lie beyond the horizon of current program planning. Nonetheless, the Science Working Group believes that long-term advanced technology efforts should be developed with such future programs in mind.

Recommendation

Advanced technology development efforts should be established to support the objectives of TOPS-0 and TOPS-1. This activity should also include early planning for a third, much more ambitious phase, TOPS-2. This phase, aimed at intensive study of discovered planetary systems, and at very high-resolution investigations of star-forming objects, will require large, complex instruments in space or on the Moon.

MOTIVATION FOR THE TOPS PROGRAM

Throughout history, human beings have felt a deep drive to understand their relationship with the universe. Much of religion and philosophy has focused on attempts to understand how we and our world came to be. Part of that understanding is tied to knowledge about the sequence of events and conditions that led to the formation of our solar system and our planet Earth. Closely related questions that also have long occu-

pied human thought ask about the prevalence of planetary systems throughout the universe, the nature of those systems, and the likelihood that other planets have given birth to life, even advanced and intelligent forms of life. We live in a remarkable time, when human beings, after thousands of years of wondering, have attained the possibility of finding the first real answers to some of these most meaningful questions.

Current thinking suggests that our planetary system is not a cosmic accident. We believe that a star and its associated planetary system form more or less contemporaneously through a sequence of related and almost deterministic events, as the interior of a spinning interstellar cloud collapses under the influence of its own gravity. The spin of the collapsing matter forces some of the material to whirl about the center in a thin, disk-shaped nebula. Through processes that only now are beginning to be understood, the solid matter and then the gas in this protoplanetary disk accumulate and organize into a set of planets orbiting the central star. The nature of the overall system and the gross characteristics and distribution of the individual planets are determined by the physical and chemical processes and thermal conditions in the protoplanetary nebula. The observation that stars seem frequently to form in similar disks suggests that stars throughout the universe are commonly accompanied by planets. The gross characteristics of our planetary system appear to have resulted more or less inevitably from conditions associated with the formation of our Sun, which implies not only that planetary systems are likely to exist in large numbers, but also that many will have overall characteristics similar to our own solar system.

It is, of course, conceivable that our planetary system is unique. Thus far, no planet beyond our solar system has ever been confidently discovered, although recent observations hint at planets or planet-like objects around a few stars. Current observing techniques, using ground-based or space-based instruments, can barely detect even massive planets around the nearest stars, and only a few planet-discovery programs are under way. Many nearby planets are likely to lie beyond the reach of current technologies.

STAGES OF DISCOVERY

Roughly speaking, our own solar system consists of three kinds of planets. The terrestrial planets, Mercury, Venus, Earth, and Mars, are so called because of their similarity to Earth. The jovian planets, Jupiter and

Saturn, are respectively 317 and 95 times more massive than Earth. Uranus and Neptune are both approximately 15 times more massive than Earth. The planets revolve around the Sun in roughly the same plane—near that of the Sun's rotational equator. The small, rock- and metal-dominated terrestrial planets formed close to the Sun, near the center of the precursor nebula, where temperatures were so high that only rock and metal could condense to a solid and accumulate to form planets. Farther from the Sun, where the nebular temperatures were lower, water also solidified, and the resulting planets formed with much higher masses, owing in part to the high abundance of water as a cosmic substance and in part to the greater number of planetesimals. The high mass of condensible solids in the more distant planets—approximately 15 Earth masses each—apparently induced even greater growth as the strong planetary gravity captured gas from the surrounding nebula. In this view, the differences between the Jupiter-Saturn pair, with their massive envelopes of gas, and the relatively gas-poor Uranus-Neptune pair probably arose because of the timing of their accumulation. Uranus and Neptune formed later, after much of the gas had escaped from the system. The similarity of the condensible-solid components of the four outer planets is important both in understanding the formation of planetary systems and in formulating search strategies and programs.

Of the planets in our solar system, Earth is unique in its combination of temperature, atmosphere, and presence of liquid water. Earth presently supports widespread and varied life, something that, as far as we know, is also unique. Long-term interest focuses on discovering and studying Earth-like planets orbiting other stars. However, the discovery of such small planets poses a difficult challenge, largely beyond the reach of current technology and programs. In this respect, the systematic aspects of our solar system and the expected similarities in other planetary systems present a remarkable opportunity to search for planets in a phased manner, with staged advances in technology and scientific understanding.

The largest and most massive planets of any system will naturally be the most readily discovered. This fact, together with the high degree of organized regularity expected of planetary systems, provides the starting point for the TOPS strategy. Jupiter and Saturn, with masses ranging from about one hundred to several hundred times that of Earth, set a reasonable expectation of possible planetary masses. Planets of these masses

can be discovered around a small number of the nearest stars using current ground-based astronomical techniques. With more sophisticated ground-based observing methods, this level of sensitivity can be extended to a significantly larger population of stars. The fact that the cores of the outer planets are equivalent to 15 Earth masses is the basis for defining the objectives of a first definitive search for other planetary systems. Current theories predict that planetary systems are prevalent around other stars, and that such a search, targeted at 100 or more single, Sun-like stars, would result in the discovery of numerous planetary systems. However, the demonstrated absence of Uranus-Neptune-mass planets around this number of stars would challenge accepted theories of the formation of planetary systems, particularly those of the origin of our solar system.

The organized regularity that we observe in the solar system, and expect in other planetary systems, plays another key role in interpreting the significance of early discoveries of very massive planets: the presence of one or more massive planets orbiting at the distances expected of such objects—beyond the boundary of water condensation in the precursor nebula—would provide strong presumptive evidence for the existence of a system displaying regularities much like our own solar system, including the presence of Earth-like planets closer to the central star.

METHODS OF DISCOVERY AND INVESTIGATION

The primary observational and technological thrusts of the TOPS Program center around several approaches and methods. These can be divided into indirect and direct detection techniques. Indirect methods of investigation capitalize on the fact that although a planet itself is difficult to "see," the star around which it orbits is easily visible. Therefore, these methods seek to discover the effect on a star of the planet(s) orbiting it. Direct detection involves the sensing of photons actually emitted or reflected from the target object.

Two indirect approaches have reached a high degree of development and are currently in routine use in a few small, highly focused ground-based observing programs. Both techniques—involving either spectroscopic radial velocity measurements or astrometric positional measurements of the central star—detect the effect of a planet's gravitational pull on the star. By virtue of a planet's gravitational tug, the central star will be seen to move back and forth across the sky. The spectroscopic radial

velocity approach measures the oscillating velocity of that motion along the line of sight. The astrometric positional approach measures the oscillating changes in a star's actual position on the sky. All current searches for planets of other stars are based on one of these two complementary techniques, which are sensitive to planets in different kinds of orbits. Both approaches have much to be gained by using larger telescopes and/or better observing sites. Although most of the capability of the spectroscopic radial velocity approach will be exhausted with ground-based systems, the astrometric technique promises much higher levels of sensitivity by deployment in space above the disturbing effects of Earth's atmosphere.

Direct detection of planets is expected to play a secondary role in the early stages of planet searches; with foreseeable systems, fewer planets are likely to be detectable through direct sensing than with indirect methods. However, direct detection will play a critical role in the intensive study of those discovered individual planets for which images can be formed. Moreover, direct detection plays the primary role in the investigation of diffuse matter associated with planetary system formation; this matter includes both protostellar nebulae and the debris disks expected to exist around stars during, and some time after, the accumulation of planets. Direct detection techniques can be applied at a wide variety of wavelengths ranging from the visible to the far infrared and millimeter. The longer wavelengths are especially useful for seeing deeply into star-forming regions, which typically are densely shrouded in gas and dust.

Four related measures of direct-detection effectiveness are the number of photons that can be collected in a given time (associated with the collecting area of the telescope); the spatial resolution of the constructed images (associated with the linear dimension of the telescope or interferometer array); the spectral range and resolution of the observations; and the usable dynamic range of the measurement. Improving all four of these is critical to advancing our understanding of other planets and of planet formation. Future improvements in collecting area, spatial resolution, and spectral range and resolution lie along the mainstream of development for ground-based astronomy. The remaining measure of direct-detection effectiveness—dynamic range—takes on special significance in the study of distant planets, disks, and circumstellar material. Seen from afar, a planet like Jupiter provides to the telescope only one photon for each 10^8 to 10^9 photons from the central star. Although

the relatively small number of planetary photons is not itself an insurmountable obstacle, the overwhelming wash of stellar photons presents special problems as it splashes over the entire near-star scene. Advanced techniques for separating planetary and disk light from this stellar blaze will be critical in enabling studies of other planetary systems.

INSTRUMENTATION

A number of rapidly developing observational techniques will contribute to meeting the goals of TOPS. Large-aperture telescopes, made possible by the technology of active figure control, will boost the collecting area for planetary photons. Adaptive optics technology, newly declassified, provides higher resolution images from telescopes within Earth's atmosphere. Optical and infrared interferometer technology, drawing on large-aperture technology and adaptive optics, will enable even higher-resolution imaging and positional measurements. By carrying observations into space, outside Earth's atmosphere, major advances can be made in the fidelity and accuracy of these observations. Ultimately, the stable platform of the lunar surface may provide the ideal site to detect and characterize Earth-like planets around nearby stars.

Discovery and study of planetary systems are inherently long-term endeavors. Extracting crucial information from the observations requires the application of both indirect and direct detection methods, which provide complementary information. Moreover, the critically important dynamical information requires studies carried out over times comparable to the orbital periods of the main planetary components. Therefore, in the initial discovery phases, sustained, consistent, and reliable measurements must be carried out over 1 to 2 decades.

CONCLUSION

The TOPS Program addresses questions of high scientific and human significance. It is a program of inquiry and discovery; the ultimate results cannot be predicted. Whatever the outcome—whether planetary systems and Earth-like planets are found to be common or rare—the results of TOPS investigations cannot fail to inform the human spirit and self-concept in a deep and fundamental way.

Contents

I. Overview 1
II. Scientific Overview and Measurement Requirements 7
2.1 Formation of Planetary Systems 7
2.2 Discovery and Characterization of Other Planetary Systems 14
2.3 Characterization of Extrasolar Planets 20
III. Indirect Detection 23
3.1 Introduction 23
Search Space 24
Discovery Space 29
3.2 Astrometry 35
Descriptions of Techniques 35
Ground-Based Astrometry: Existing Programs and Projected Improvements 43
Proposed TOPS-1 Space-Based Programs 49
Proposed TOPS-2 Indirect Detection from the Moon 59
3.3 Radial-Velocity Measurements 59
State of the Art 60
Role of Radial-Velocity Measurements in TOPS 66
3.4 Photometry 66
IV. Direct Detection 73
4.1 Introduction 73
4.2 Planet Detection 74
Planetary Radiation 74
Planetary Spectra 77
Planetary Properties 79
4.3 Protoplanetary and Debris Disks 80
Protoplanetary Disks 81
Debris Disks 83
4.4 Instrument Strategies 84
Angular Resolution 84
Signals and Noise 85
Diffracted and Scattered Light 86
Fourier Transforms 86
Coronagraphs 89
Interferometers 91
Adaptive Optics 96
4.5 TOPS-0: The Keck II Telescope 99
TOPS-0 Instrumentation 100
The Keck I and II Telescopes 104
Outrigger-Enhanced Interferometry 108
4.6 TOPS-1: Astrometric Imaging Telescope 111
4.7 TOPS-2: Characterization of Extrasolar Planets 113
Infrared Detection of Earth-Like Planets 114
Cold Optics 114
Planet Detection 115
The Future 118

Page xx — Blank page

I. Overview

THE SCIENTIFIC METHOD WAS ESTABLISHED IN THE DEBATE that followed Galileo's 1610 discovery of the phases of Venus and the satellites of Jupiter using the first astronomical telescope. As science and technology advanced through the 18th, 19th, and early 20th centuries, Herschel, Galle, and Tombaugh discovered Uranus, Neptune, and Pluto, the most distant planets in our solar system. With the advent of the space program, cameras in space revealed the detailed appearances of the planets, including the face of Earth itself. Today, we can finally address and resolve a further question of the ancient issue of planets: Do planetary systems exist around other stars? The TOPS Program—Toward Other Planetary Systems—is a plan for NASA to take up this challenge and answer that question. This report lays out that plan.

The TOPS Program has three facets: to discover planetary systems, to understand their formation, and to lay the groundwork for a greater challenge, the ultimate question engendered by the Copernican revolution: Does life exist on planets around other stars?

Because of the wide interest in this topic and its newness as a research field, the authors have taken care that both our report and our plan are intellectually complete. That is, as scientists, we recognize that the question of whether planets exist around other stars is as formidable a research topic as the answer. A simple yes or no response to the question of the existence of extrasolar planets is not sufficient to satisfy humanity's curiosity about other worlds.

The question of the existence of extrasolar planets draws upon all we know of Earth, the solar system, the life of stars, the remarkable phenomenon of terrestrial life, and the vantage point—unique or otherwise—of the human mind on the cosmos. A dependent relationship exists between such questions and the intellectual context that gives rise to them and gains impetus from their answers. The search for extrasolar planets has a framework in the physical sciences of two rich fields of related research, planet formation theory and observational studies of possibly planet-related phenomena around other stars. Therefore, this report presents—and our recommended program pursues—the theory of how planets form and observations related to planet formation.

Modern thought about the formation of the solar system can be traced back almost 200 years to the writings of I. Kant and

P.-S. Laplace, both of whom hypothesized that the Sun and planets formed from a single rotating, contracting cloud of gas. Other hypotheses for planet formation have been proposed and discarded, but this hypothesis constitutes the most basic paradigm for how planets like Earth formed.[1,2]

Our current understanding of planet formation and evolution indicates that the process begins with the gravitational collapse of a dense interstellar cloud of gas and dust. The protostar is the central condensation, which, although it is not at this stage visible from the outside, accretes further material, shrinks, and later becomes a "sun." Material that does not fall on the protostar because of its rotational motion forms a disk in the equatorial plane. A few tens of thousands of years after the initial collapse, the dust sinks to the midplane and forms a dense sheet. Within this sheet, the dust begins to clump together and agglomerate into planetesimals. The planetesimals accumulate to form terrestrial planets and the cores of giant planets. Next, the cores accrete atmospheres. Finally, the residual gas and dust dissipate, either blown away by stellar winds or accreted onto the planets that remain.

This hypothesis may soon be confirmed by detailed observations of the environs of young stars still in the process of formation. Several independent lines of evidence strongly suggest the presence of disks of gas and dust around nearby young stars.[3,4] These disks are likely to be capable of producing planetary systems, perhaps similar to our own. Theoretical studies of the physical mechanisms through which planets may form have yet to uncover any reasons that the formation of a planetary system should be a rare or unique event. Planetary systems are thought to be common outcomes of the ongoing process of star formation throughout our galaxy.[5,6] The possible existence of many planetary systems demands a serious effort to detect and eventually characterize these extrasolar planetary systems, for the intrinsic excitement of finding and studying other worlds, and, through comparative studies, for increased understanding of how our own solar system came into being.

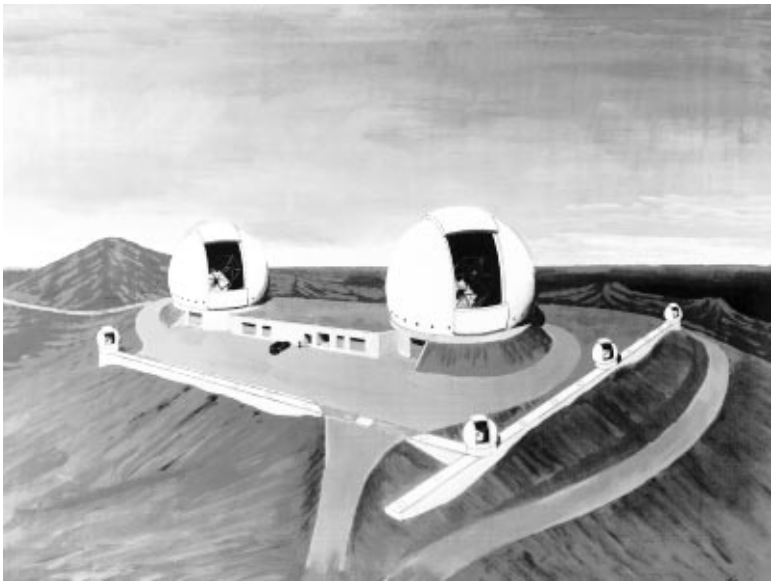
The observational techniques to detect evidence of other planetary systems are developed sufficiently to begin a promising search now, using "indirect" methods at first, and progressing to "direct" detection. A program can be laid out with some confidence, building on knowledge that already exists. Indirect searches entail the study of the parent star rather than its orbiting planets. The gravitational influence of giant plan-

ets like Jupiter, Saturn, and Uranus induces a substantial wobbling motion of the star. This motion, which should be measurable, is being looked for now with modest instrumentation, and the TOPS Program can provide major improvements in sensitivity and accuracy. Direct methods, by which the radiation from the planet itself is detected, present a more challenging, but also realistic, goal. Expected technological advances will provide unprecedented abilities to detect and study other planetary systems, despite the glare of the parent star. Confirmed detection of extrasolar planetary systems will further stimulate advances in observational techniques that will eventually allow us to determine increasingly detailed properties of these other planetary systems and ultimately to confirm or disprove the existence of other Earth-like planets. If it can be shown statistically that planets form only rarely, that information will undermine our current assumption that planetary system formation is a natural consequence of star formation. If, as expected, other planets are found, and a significantly large sample of data has been gathered, the spectrum of the planetary masses and their distances from the central star can be determined. This knowledge would greatly influence our theories of solar system origin.

Improving our understanding of the origin of planetary systems will also require thorough examination of preplanetary disks and more advanced disks containing protoplanets. These disks must be probed at much higher angular resolution than is currently possible, which is challenging, but well within current technology projections. The result of this major effort will be a stronger knowledge of how our solar system came to be, and a census telling us how common protoplanets are in star-forming regions.

Just as technical advances have opened the astronomical record and revealed evidence relating to planet formation, even more recent advances now permit our first definitive searches for planets using indirect techniques that measure the stellar motion induced by an orbiting planet. Such techniques can rule out the presence of a planet (in a specific range of mass and orbit size) if no perturbation is detected. If it can be shown statistically that planets form only rarely, that information will undermine our current assumption that planetary system formation is a natural consequence of star formation.

The TOPS program takes advantage of recent technological advances and builds on them to achieve specific objectives. These are orga-



The Keck Observatory after the development of the second 10-meter telescope, in which NASA intends to participate.

nized into three programmatic phases: TOPS-0, centered on the Keck II telescope; TOPS-1, centered on an orbiting instrument; and TOPS-2, centered on a technical development program for interferometry. Interferometry is an extension to optical wavelengths of a technique commonly used in radio astronomy that combines several separate telescopes to build, in effect, a larger and more powerful telescope. This astronomical technique will be required to image and study Earth-like planets around other stars. All phases of the TOPS Program assume certain continuous elements, including ground-based observing projects and a theory program involving modelling, interpretation, and laboratory investigations. These elements are necessary for the vitality of any science program, but are particularly important for an emerging field of research.

The defining element of TOPS-0 is the Keck II telescope, which will build upon the experience and accomplishments of NASA's Infrared Telescope Facility (IRTF). The Keck program will require (1) telescope construction, (2) instrumentation to enhance image sharpness, (3) focal-plane instrumentation for studying planetary phenomena and materials around other stars, and (4) an interferometer project using the Keck tele-

scopes and additional outrigger telescopes. The direct detection component of TOPS-1 is an imaging instrument that is part of one of the candidate concepts for the TOPS-1 indirect search. The TOPS-2 goal of ultimately studying "earths" around other stars requires a far more ambitious initiative. Its specific form depends on the exploitation of imaging interferometry on a large scale in space or on the Moon. The preparation can start now as a technology program.

Present funding available for research on these topics is insufficient to support theoretical and laboratory efforts on a scale proportional to the ambitious observational programs to be discussed in the following chapters. The highly interdisciplinary nature of planetary formation studies will require larger research groups than the single investigator efforts that currently characterize the field. NASA's Origins of Solar Systems Program already supports related research, but its practice of providing relatively small augmentations to a wider range of investigations means that the Program will not meet the requirements for supporting scientific research. The major observational thrust necessary to meet TOPS science goals must be coupled with strongly enhanced support for related laboratory and theoretical studies in new TOPS research and analysis programs to expedite the development of our combined understanding of stellar and planetary system formation.

Chapter II of this report describes the current theoretical understanding of planet formation: the basis for the TOPS theory program and the motivation for the goals of the projected TOPS observations. Chapter III describes the indirect search techniques—astrometry and radial velocity—that will be used to carry out complementary, definitive searches for solar systems with massive planets like Jupiter, Saturn, Uranus, and Neptune. Support for a variety of ground-based indirect search programs will also be an ongoing element of TOPS. The space-based astrometry project, which will likely use one of three candidate instruments, is the distinguishing feature of TOPS-1.

Chapter IV describes the wide range of direct detection techniques that are critical to starting and sustaining the TOPS Program, and realizing its ultimate goal: finding and studying Earth-like planets around other stars.

There is a further chapter that does not appear in this report, but that will certainly be addressed in the future: the coupling of the TOPS Program with education and the public. In the last 30 years, NASA's plan-

etary program, through its dramatic visits to the planets of our solar system, has forged a process that has transformed the excitement of modern-day exploration. Its success has been due in large part to the solidity and resilience of the program's framework. The astronomers and planetary scientists writing this report wish to continue this tradition of excellence by maximizing the public's opportunity to participate in the search for planets around other stars. Our search for other planets is, after all, a Nation's quest. The great public curiosity about the existence of extrasolar planets should thus be encouraged by the TOPS research program. Whereas *Mariner, Pioneer,* and *Voyager* returned a deluge of magnificent pictures, the special nature of the TOPS investigations will benefit from a concerted educational program to enhance the public's understanding of the research and its results.

REFERENCES

1. Safronov, V. S. 1969, *Evolution of the protoplanetary cloud and formation of the Earth and the planets* (Moscow: Nauka).

2. Cameron, A. G. W. 1988, *Ann. Rev. Astron. Astrophy.*, 26, 441.

3. Beckwith, S., and Sargent, A. I. 1992, in *Protostars & Planets III*, ed. E. H. Levy, J. I. Lunine, and M. S. Matthews (Tucson: Univ. Arizona Press), in press.

4. Strom, S. E., et al. 1992, in *Protostars & Planets III*, ed. E. H. Levy, J. I. Lunine, and M. S. Matthews (Tucson: Univ. Arizona Press), in press.

5. Shu, F. H., Adams, F. C. and Lizano, S. 1987, *Ann. Rev. Astron. Astrophy.*, 25, 23.

6. Boss, A. P. 1989, *Pub. Astron. Soc. Pac.*, 101, 767.

II. Scientific Overview and Measurement Requirements

2.1 FORMATION OF PLANETARY SYSTEMS

DURING THE LAST SEVERAL DECADES, SCIENTISTS HAVE USED theoretical modelling to gain a preliminary understanding of how planetary systems form. However, only a long-term observational program of the type envisioned by TOPS will ultimately confirm or disprove this general paradigm.

We believe our planetary system formed from the collapse of a rotating cloud of interstellar gas and dust. (Likely candidates for an analogous dense molecular cloud have been detected and studied in the numerous star-formation regions in the disk of our galaxy.) The result of this self-gravitational collapse was the formation of the protosun and a rotationally flattened disk, the solar nebula, in which the plane of the nebula was more or less perpendicular to the total angular momentum of the solar system. We have strong evidence for this formation mechanism—the near coplanarity and corevolution of the planets, the close coincidence of the plane of the planets to the plane of the rotational equator of the Sun, and the near circularity of the planetary orbits. These structural regularities indicate that the protostellar cloud had collapsed and settled into a flattened, dissipative disk, with circularized motions, before the planets formed.

After any initial turbulence associated with the collapse died down, and the solar nebula became quiescent enough, dust grains of silicates, iron, and ices from the protostellar cloud settled to form a denser region in the nebula midplane.[1] Random motions produced by interactions of the dust grains with gas molecules caused collisions between the dust grains. Collisions provided the opportunity for dust grain coagulation caused by the stickiness associated with intermolecular forces between the surfaces of the dust grains. Dust grain coagulation initiated the process of solid body growth from the very small 0.1-micron sizes characteristic of interstellar dust grains to more or less solid particles perhaps centimeters in size, over thousands of years. After kilometer-radius bodies formed, gravitational forces were sufficient to account for continued growth. In the intermediate size range—1 cm to 1 km—we are uncertain how the particle agglomeration and growth proceeded.

In the terrestrial planet region (inside about 4 astronomical units (AU): 1 astronomical unit is the distance from Earth to the Sun, about 150 million km), the km-sized bodies grew for about 100,000 years through further random collisions with their neighboring planetesimals.[2,3] Once roughly lunar-sized (1000 km) bodies formed, however, all the neighboring planetesimals were already accreted, and large changes in orbital eccentricity were required to collide with planetesimals at even greater distances.[4] Gravitational forces produced these increases in eccentricity over a much longer time scale. For example, the lunar-sized bodies grew to Earth-sized protoplanets over 10^7 to 10^8 years. This phase was apparently characterized by catastrophic, giant impacts between nearly planet-sized bodies.[5]

For the outer planet region (outside 4 AU), theoretical evidence indicates that planet formation also proceeded initially through a phase in which solid bodies accumulated following collisions. However, the cooler temperatures found in the outer solar system allowed the condensation and preservation of ice grains composed of water, carbon dioxide, methane, ammonia, and other volatile species. When the resulting ice and rock planetesimals and their gaseous atmospheres grew to masses on the order of ten times that of Earth, a new phenomenon occurred: the dynamical collapse and accretion of gas from the solar nebula onto the solid protoplanetary core.[6] This fairly rapid addition of a hydrogen and helium envelope to the rocky core must have occurred while the gaseous component of the solar nebula was still present. Astronomical observations of possible protoplanetary disks imply the disappearance of at least the small dust components of the disks over hundreds of thousands to millions of years,[7] forcing a severe time constraint on the planet-building process in the outer solar nebula. Various hypotheses have been advanced for speeding up accumulation in the outer nebula,[8,9] but this time constraint remains one of the fundamental difficulties in understanding giant planet formation and assembling a consistent and complete theory of planet formation.

The nearest star-forming molecular clouds are about 150 parsecs (1 parsec (pc) is approximately 3 light-years; more precisely, 1 parsec is 206,265 AU = 3.1×10^{13} km) away. Many T Tauri stars (a phase of evolution our Sun passed through early in its life) in these clouds appear to be associated with protoplanetary disk-like structures.[10,11] Interferometric arrays of millimeter-wave telescopes have uncovered evidence for

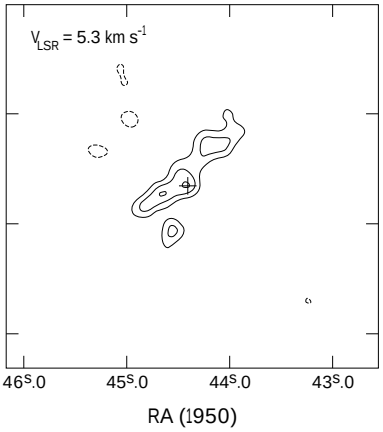


Figure 2-1. The gaseous disk surrounding HL Tau as seen by mm-wave interferometry.[33] The map of ^{13}CO emission at 2.7 arcsec angular resolution and centered at a velocity of 5.3 km s^{-1} suggests the presence of a nearly edge-on disk around the pre-main-sequence star HL Tau (cross). Such disks may lead to the formation of extrasolar planetary systems.

large-scale gaseous disks in Keplerian rotation about low-mass pre-main-sequence stars.[12] Figure 2-1 depicts structure suggestive of a gaseous disk in orbit around the T Tauri star HL Tau. Circumstellar dust disks have been inferred around numerous young stellar objects on the basis of observations showing excess emission in the far-infrared.[7] These tantalizing observations certainly support the general concept of protoplanetary disks, but the structure of gaseous disks in the planet-forming regions has not been observed, and so we are unable to constrain the theory of how planets form.

Other questions about the origin of the planets remain. For example, when were the terrestrial planets formed with respect to the phase of nebular evolution—was the gaseous component of the nebula still present when the final bodies formed? Answering this question will require detailed observations of the gaseous content of protoplanetary disks. Also unknown is the interaction of the solar nebula with the growing protosun. Young stellar objects show evidence of energetic mass loss even at the earliest phases, and this mass loss may be driven by ongoing accretion of gas from the protostellar cloud and the protoplanetary disk.[13] How does the interaction between disk accretion and wind expulsion work? Does the mass loss in bipolar flows from the central regions of a protostar evolve into T Tauri winds that may ultimately disperse any unaccreted components of the nebula?

One of the most challenging observational requirements for improving our understanding of planetary formation is to obtain images with increasingly detailed spatial resolution, ultimately reaching about 0.1

AU at the distance of the nearest star-forming molecular clouds, nominally 150 pc, i.e., sub-milliarcsec angular resolution. Telescopes operating with wavelengths ranging from the near-infrared (2 μm) to millimeter (mm) wavelengths (1000 μm) are necessary to explore these optically thick protoplanetary disks. Near-infrared telescopes should be able to detect emission from warm gas and dust in late-phase protoplanetary nebulae surrounding T Tauri stars, whereas mm-wave telescopes will be needed to detect cold dust and molecular absorption and emission lines in nebulae both before and after the appearance of a pre-main-sequence star.

Although the ultimate goal of the TOPS Program is to achieve spatial resolution on the order of 0.1 AU, such an ambitious goal will most likely only be reached through incremental improvements. The current limiting resolution for mm-wave interferometer observations of nearby star-forming regions is about 100 AU. An intermediate resolution of 10 AU would enable mapping of the outer regions of protoplanetary disks, whereas a resolution of 1 AU would yield valuable information about the gradients of density and temperature in the inner regions of such disks.

Sub-milliarcsec angular resolution at optical wavelengths (0.5 μm) requires a minimum telescope aperture (or baseline for an interferometer) of about 200 m; at the longer wavelengths necessary to penetrate optically thick nebulae (5 μm to 1000 μm), this size increases proportionately. Even when this criterion is satisfied, space-based instruments seem to be required in general because the deleterious effects of atmospheric seeing can smear out ground-based images to much lower angular resolutions (1 arcsec).

Spatial resolution of 0.1 AU would allow us to map protoplanetary nebulae on a fine enough scale to determine key physical properties. For example, the nebular temperature determines which elements remain in the gas phase and which condense into solids and hence are available for accumulation into planetesimals. The surface density determines the total amount of matter available for planetary formation and ultimately the type of planet produced at any given radius. Both the temperature and density need to be determined to a precision of about 10% to identify features expected in planet-forming disks, such as hot young protoplanets still embedded in the thermal bath of the nebula[6] and possible gaps in the surface density caused by gravitational interactions with em-

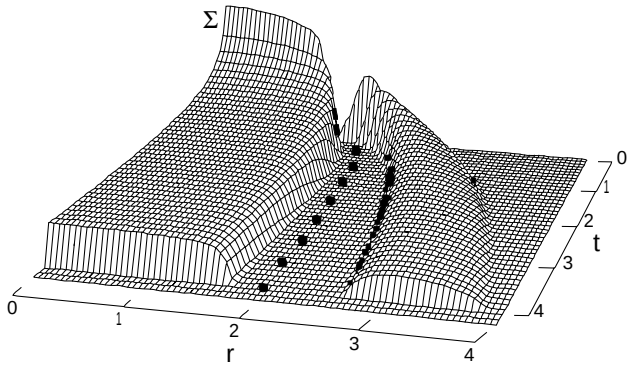


Figure 2-2. Gas surface density as a function of radius (r) and time (t) in a theoretical model of the solar nebula.[14] At t = 0, a massive planet is introduced, which interacts with the gaseous nebula through gravitational forces and clears a gap around the planet (solid dots). Depending on the initial nebular gas distribution, the planet may remain fixed in radius, or move inward or outward (as shown). Whether gaps actually form depends on the planet mass and the viscosity of the nebula.[15]

bedded protoplanets[14,15] (Figure 2-2). Midplane (maximum) temperatures[16,17] are expected to range from around 1500 K at 1 AU to about 100 K at 5 AU, falling to about 10 K in the outermost regions (beyond 100 AU). The effective radiation temperatures characteristic of the surface layers of the nebulae will be considerably cooler in the inner nebula, with values around 100 to 200 K at 1 AU.[18] The latter value is the nebula temperature inferred from observations limited to optically thin surface layers. Total gas surface densities will vary[17] from about 10^4 to 10^2 g cm^{-2} between 1 and 5 AU respectively. Dust surface densities will be about a factor of 100 lower in protoplanetary disks with a gas-to-dust ratio (i.e., chemical composition) similar to that inferred for the solar nebula.

A spatial resolution of about 0.1 AU would also allow study of the vertical structure of nearly edge-on nebulae, and reveal the details of the interaction of early stellar winds with the nebula surface, an interaction that may play a role in the removal of nebular material. Do stellar winds escape preferentially through bipolar holes, avoiding the nebula, or do they sweep over the nebular surface, entraining and removing matter[19] to infinity? Figure 2-3 illustrates the region surrounding the young stellar object responsible for Herbig-Haro object 34, where disk accretion and energetic stellar winds may be occurring simultaneously.

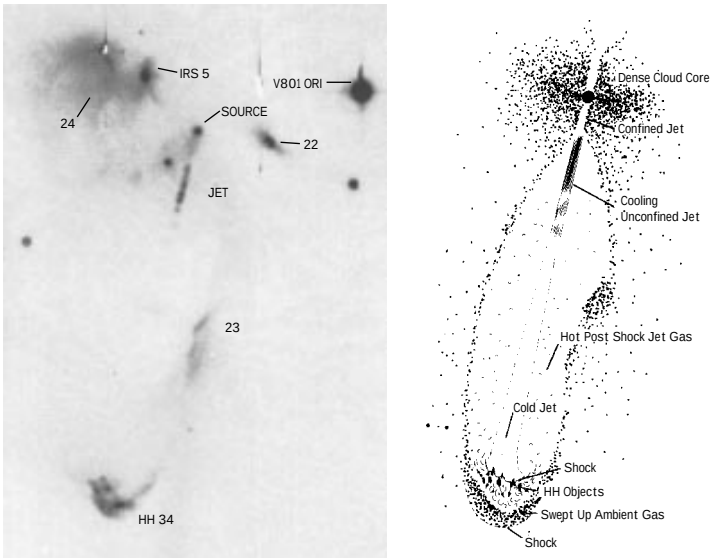


Figure 2-3. Optical CCD image of the region around Herbig-Haro object 34 (left) and an interpretation (right) of the activity.[34] A low-mass, young stellar object (labelled SOURCE in the image on the left), still deeply embedded in its placental cloud core, is undergoing energetic mass loss. The mass loss is in the form of oppositely directed jets that sweep out ovoidal cavities and produce Herbig-Haro objects at the bow shock fronts. One jet is visible here; the object produced by the jet on the opposite side has also been detected.[35]

A protoplanetary nebula must evolve so that most of the nebula's mass is transported inward to be accreted by the central protostar. The physical mechanism that leads to this major transport process is unknown. Massive disks may be particularly subject to the growth of nonaxisymmetric (with respect to the rotation axis) structures, such as bars and spiral density waves, similar in appearance to those observed in barred and spiral galaxies. Determination of the level of nonaxisymmetric structure in protoplanetary nebulae will help identify the mechanism chiefly responsible for nebular evolution. A nonaxisymmetric nebula implies that gravitational torques dominate angular momentum transport,[17] whereas an axisymmetric nebula must evolve through viscous[20,21] or magnetic torques.[22]

One of the more interesting aspects of the predicted time scale for evolution of protoplanetary disks is that it is short enough to be observable if we can detect a system at this early phase. With high spatial resolution observations spanning a period of several years, we may find evi-

dence for the rapid accretion of nebular gas by growing giant planets.[6] Time-variability is an intrinsic property of pre-main-sequence stars,[23] and as we probe deeper into their environs, we can expect to uncover evidence for disk evolution and orbital motion.[24]

Finally, imaging at high spatial resolution will help us understand the differences between planet formation around single stars like our Sun and around the binary star systems that contain the majority of the stars in our galactic disk. A binary system with a 10 AU stellar separation at the center of a protoplanetary nebula would be easily resolvable (Figure 2-4). We may be able to observe the rapid evolution (perhaps over decades) expected if some close binary protostellar systems merge into single protostars through loss of orbital angular momentum.[25]

In addition to the sub-milliarcsec spatial resolution requirement for imaging protoplanetary systems, the TOPS observations must be made with moderately high spectral resolution (at least 0.1 km s^{1}) of the molecular lines that probe protoplanetary disks. Achieving these goals will require large collecting areas to split a weak signal into multiple velocity channels. High-resolution spectroscopy will be useful for viewing the velocity features in the nebula and its envelope and for separating specific nebular features from foreground or background matter at slightly different velocities. This velocity resolution would also be sufficient to pinpoint the rotational motion of the gas and dust in the nebula. Significant rotational motion is the primary reason nebulae should exist and form planets; in the absence of angular momentum, all matter would fall onto

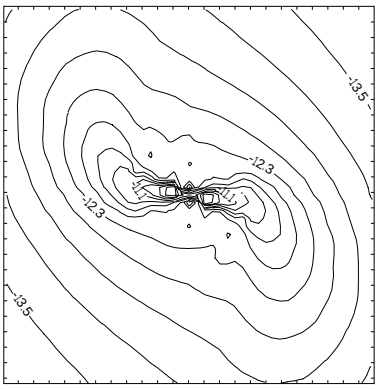


Figure 2-4. Density distribution during a theoretical calculation of the formation of a protostellar disk.[25] A dense interstellar cloud core has collapsed and fragmented into the two protostars visible at the center, separated by about 10 AU. Trailing spiral arms, which can transfer angular momentum outward, merge the binary into a single star, and force the protoplanetary disk to evolve. Very high spatial resolution is necessary to search for these and other hypothesized processes in nearby star-forming regions.

the central star. Combining spectroscopy and high-resolution imaging, for example, by mapping molecular transitions over a wide wavelength range, will make it possible to isolate different nebular regions for investigation. The signature of the bipolar regions, swept by winds flowing from the central protostar, may deviate from that of the equatorial regions, where gas and dust from the parent molecular cloud continue to accrete onto the nebular disk and afterward onto the protostar. Such observations should provide direct evidence of accretion by low-mass protostellar objects.

One of the most important applications of spectroscopy is that it allows us to identify the chemical constituents of protoplanetary disks, and implicitly to search for evidence of chemical evolution in our solar system. Spectroscopy of molecular species such as water (H_2O), methane (CH_4), and carbon dioxide (CO_2) is necessary to ascertain radial abundance profiles that will locate the water ice condensation/vaporization boundary (crucial for giant planet formation),[8,9] and to provide insight into the conversion of interstellar CO into the CH_4 predominantly found in our planetary system.[26] Similarly, spectroscopy could uncover variations in the gas-to-dust ratio, abundances of refractory species, and isotopic fractionation ratios, very important information for interpreting our knowledge of the chemistry of asteroidal (meteoritical) and cometary bodies.[27] In the solar system, these studies provide important clues to the origin and chemical evolution of Earth and the other planets.

Pre-main-sequence stars and their accretion disks are subjects of much current observational and theoretical interest and will likely become even better understood in coming years. Meeting the science requirements outlined in this Chapter will ensure that our understanding of protoplanetary nebulae keeps pace with these efforts.

2.2 Discovery and Characterization of Other Planetary Systems

Inasmuch as our current ideas about planet formation are controlled almost entirely by what we know of our own planetary system—illuminated and embellished by studies of observable protostars, young stellar objects and disks, and possibly planetary-mass bodies orbiting pulsars—knowledge of the solar system plays a critical role in establishing our measurement objectives.

The planets seem to have formed through a process controlled by the

accumulation of the solid-matter fraction of protostellar material.[2,4] The gross variation of planetary masses and compositions with distance from the Sun reflects this process. On this basis, the major contrast between the small, rock- and metal-rich terrestrial planets and the large, gaseous, ice-rich giant planets resulted because the giant planets formed far from the Sun, where cosmically abundant water could persist as ice. Uranus-mass planets formed farther out, beyond the more massive Jupiter-like planets, because of the slower rate of formation at greater orbital radii and the decrease in condensed and gaseous matter with distance from the Sun. Figure 2-5 demonstrates that, although we may expect to find certain regularities in extrasolar planetary systems, a stochastic compo-

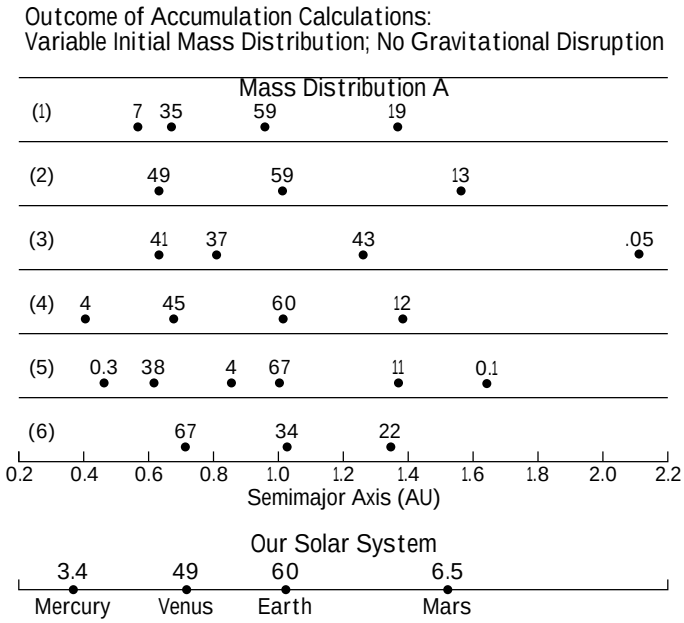


Figure 2-5. Masses, semimajor axes, and eccentricities for six theoretical simulations of the formation of the terrestrial planets.[5] Our solar system is shown at the bottom for comparison; the numbers associated with each planet give its mass in units, where 60 is the mass of Earth. The calculations follow the accumulation of the planets through collisions within an initial swarm of lunar-sized bodies. Even though each of these simulations started with the same energy and angular momentum, because of the random nature of planet formation by accumulation, the result is not always four planets. Evidently we can expect a certain degree of variety in extrasolar planetary systems, even though Earth-like planets may nearly always be present.[36]

nent arising naturally from the process of formation through collisional accumulation should result in a variety of possible configurations for extrasolar planetary systems.

The search for planetary companions of nearby stars has long been the focus of a small number of astronomers, but the difficulties attending the search have precluded a confirmed detection. Early evidence for planetary companions to Barnard's star has not been confirmed. A number of ground-based efforts currently under way may yet detect Jupiter-sized planets. Candidate companions with masses intermediate between the lowest-mass main-sequence stars (0.08 solar mass) and Jupiter's mass (0.001 solar mass), depending on the orientation of the orbit, have been identified through measurements of the Doppler shift of the central star's orbit about the common center of mass.[28,29] Objects with a similar intermediate mass have been inferred in orbit about white dwarf stars,[30] and two bodies of at least 3 Earth masses have been claimed recently to be in circular orbit around a neutron star (pulsar),[31] but the resemblance of such objects to planets like those of our solar system is quite unknown. Nevertheless, the unexpected evidence for planetary-mass companions to pulsars suggests that accumulation processes in circumstellar disks may lead to sizeable bodies in very different phases of stellar evolution. This evidence greatly strengthens the case for the existence of planetary systems similar to the solar system.

A major program for the discovery and study of planetary systems should have several stages, each a major step toward accomplishing a definitive search. These stages should be undertaken on the basis of ambitious, but reasonable, expected increments of knowledge, and organized on the basis of technological, programmatic, scheduling, and measurement-capability considerations. The measurement program should be designed to be sensitive to the general systematic regularities expected to be found—nearly circular, coplanar orbits. No short-term program utilizing a single measurement technique will accomplish all the measurement objectives. Whatever the technique, measurements may be needed over a period longer than the longest orbit period of interest to ascertain the planetary orbits and masses. Thus, to discover and study the expected giant planets, the initial investigation program must be planned for at least 1 to 3 decades. It is important that such a long-term observing program be designed to gather, from the beginning, the quality of data needed to achieve as complete an understanding of extrasolar sys-

tems as technically possible.

Although the discovery and study of Earth-like planets is perhaps the most exciting aspect of the TOPS program and one of its most important goals, advancing our knowledge of star and planetary system formation requires discovering and studying planets in both the giant and terrestrial classes. Because of their shorter expected orbital periods, an "indirect" (i.e., astrometry, photometry, or Doppler spectroscopy; see Chapter III) program to investigate the existence and properties of terrestrial planets could be carried out over a period of a year or two. A "direct" program seeking to image the light from Earth-like planets would give results in a fraction of an orbital period. However, generally speaking—because of their much smaller masses and sizes, as well as their proximity to the central star—the discovery and study of terrestrial planets pose more severe technical challenges than the discovery of giant planets. Based on the orbital regularities observed in our solar system and our conceptual picture of planetary system formation, discovery of giant planets arrayed with these systematic regularities would provide strong presumptive evidence for the existence of terrestrial planets in the same systems, and would motivate the strenuous technical efforts needed to discover and study terrestrial planets in other systems.

A progression of specific objectives must correspond to successively more demanding requirements on instrument sensitivity and measurement accuracy. The four general techniques—astrometry, Doppler spectroscopy, imaging, and photometry—are described in detail in the following chapters. In this section, we outline the performance required for each technique to meet the objectives of finding Jupiter-like, Uranus-like, and Earth-like planets around other stars. Inherent in this discussion is the assumption that the target stars of greatest interest in the search for other planetary systems are isolated, nearby, main-sequence stars of F, G, K, and M spectral types with known parallaxes. These stars range in mass from about 0.2 to 1.8 solar masses, and we loosely refer to them below as being of solar type. Secondary emphasis would be given to F, G, K, and M stars in wide binary systems, where we may also expect to find planetary systems.

On the basis of these considerations, we envision that for the discovery and study of planetary systems, the following stages constitute a logical approach. The first objective (TOPS-0) is to search for planets at least as massive as Jupiter (radius about 0.1 solar radii, mass about 300

Earth masses, and effective temperature about 130 K) 0.5 to 8 AU from solar-type stars. This objective should be achievable with ground-based instruments for at least 100 stars.

The second objective (TOPS-1) is to search for planets at least as massive as Uranus 0.5 to 8 AU from solar-type stars. This objective should be the next step beyond current ground-based programs, and should serve as the primary design criterion for the first generation of ground- or space-based instruments to search for extrasolar planets.

The third objective (TOPS-2) is to search for and study Earth-sized planets 0.2 to 3 AU from solar-type stars. This search will be the ultimate challenge, requiring technology not yet developed, and possibly the operation of interferometric telescopes on the lunar surface. Identification and development of the critical technologies for this third objective should begin as early as possible.

These objectives should be understood as describing nominal measurement capabilities, based on the considerations given above. In addition to these primary objectives, the search plan should be constructed to be definitive. A possible negative result, however unanticipated, should be interpretable as a fundamental advance in our knowledge, comparable to actual discoveries. Therefore, the search should encompass a sufficient population of target stars, ultimately exceeding 100 to 1000 stars of solar type, and the search should extend to distances between 15 and 30 pc.

The detailed investigation of discovered systems should be planned to characterize the systems at a level sufficient to ascertain the strengths or weaknesses in our current ideas and to establish the gross variations that may occur among planetary systems. Thus, the prevalence and characteristics of planetary systems should be determined as functions of stellar spectral type, and the possible existence of planets in binary or multiple star systems should be explored.

The scientific objectives also require determining the degree to which the planetary orbits are circular and coplanar and the variation of planetary masses with distance from the central star. Thus, for individual systems, it is important to determine the number of planetary objects (at least the major objects in both giant and terrestrial planets), the radii of the orbits to about 30% accuracy, the orbital eccentricities in intervals of about 0.2, the orbital inclinations to within about 20 degrees, and the masses of the planets to within a factor of about two.

Determination of the mass spectrum of planets is critical to resolving the question of what distinguishes a planet from a star. If the planetary mass spectrum does not extend far above Jupiter's mass, a clear gap would exist between the maximum planetary mass and the theoretical minimum protostellar mass[25] of about 10 Jupiters. Stellar objects of 10 to 80 Jupiter masses ("brown dwarfs") are already the subject of intense observational search, and proof of their existence would populate the lower end of the stellar mass spectrum, which otherwise would terminate at about 80 Jupiter masses. The confirmation of a suspected gap between the planetary and stellar mass distributions would also support the hypothesis that planets and stars differ essentially in their mechanisms of formation, with stars forming as a result of hydrodynamic collapse of interstellar clouds or cloud fragments, and planets forming as a result of collisional accumulation of solid matter in accretion disks surrounding protostars.

A significant additional area of discovery and study for main-sequence stars is the more or less diffuse solid matter that may be observable in circumstellar disks and halos, involving dust as well as objects of cometary or asteroidal type. The discovery[32] of the dust disk surrounding the main-sequence star Beta Pictoris (Figure 2-6) was a spectacular example of the exciting science that can result from using a specialized instrument. The scientific measurement objectives for mature circumstellar material are identical to those for the near-infrared observations re-

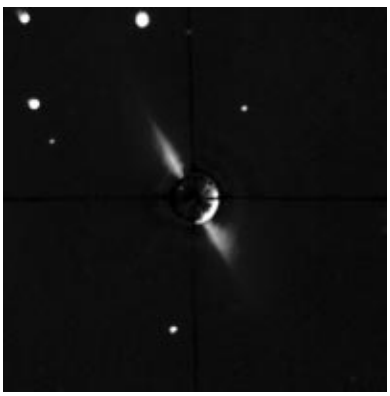


Figure 2-6. Optical image of the dust disk around the nearby main-sequence star Beta Pictoris.[32] The presence of a dusty circumstellar component around this mature star was suspected on the basis of excess infrared emission detected by the Infrared Astronomical Satellite. The use of a stellar coronagraph (the four dark "cross hairs" are the supports) on a telescope at the Las Campanas Observatory allowed the light from Beta Pictoris to be largely blocked, revealing a flattened disk of small dust grains. The continued presence of small dust grains with limited lifetimes in such a circumstellar environment implies a population of much larger bodies in orbit about Beta Pictoris.

garding protoplanetary formation; i.e., sub-milliarcsec angular resolution and 0.1 km s^{-1} spectral resolution.

2.3 CHARACTERIZATION OF EXTRASOLAR PLANETS

If planets are detected around other stars, attempts should be made to observe the spectra of the individual planets. The spectra of planets in our solar system reveal characteristic bands that identify many of the atmospheric constituents and enable estimates of their relative abundances. Some fraction of this information about the atmospheres of planets in a remote system might be collected with instrumentation affording sufficient spatial and spectral resolution and sensitivity (e.g., large collecting areas and interferometric configurations), at least for those systems discovered within 5 to 10 parsecs. Molecules identifiable in solar system planets in spectra taken from outside the planet's atmosphere include diatomic oxygen (O_2), ozone (O_3), carbon dioxide (CO_2), and methane (CH_4).

It is too soon to be certain what constraints on planet history or current evolutionary state can be implied by the observation of these molecules in the atmospheres of planets in a remote system, with the probable exception of oxygen. A very strong O_3 absorption would indicate the presence of an atmosphere with a major oxygen fraction. Free oxygen exists on Earth only because of constant replenishment by ubiquitous plant life, and it would rapidly vanish from the atmosphere if all plants were removed. If explicit disequilibrium were demonstrated by the simultaneous presence of CH_4, a product of both plant and animal processes, the implication that a particular planet harbored carbon-based life would be very strong. The possible existence of a potentially oxygen-rich planet might be indicated by a transition from CO_2- to CH_4-rich atmospheres in a given planetary system, where the transition occurred over the radial distance from the star where water could remain liquid. Such a transition would motivate an intensive search for an O_3 or O_2 signature.

The possibility of determining the constituents of atmospheres of planets in remote planetary systems through spectroscopy is the only means of constraining the constituents of these individual planets. The correlation of these constituents with other properties of the planet, such as its mass and distance from its central star, will lead to important comparisons with our own solar system, and inevitably to speculation about

the occurrence of life on extrasolar planets.

The challenge of discovering and studying planetary systems around other stars can be approached by a variety of complementary techniques. It is convenient to divide these planet-discovery techniques into two broad classes: direct approaches and indirect approaches. "Direct discovery" involves the detection and identification of photons from the planets themselves, either photons reflected from the central star or energy reprocessed and emitted as thermal planetary photons. "Indirect discovery" involves the detection and measurement of photons only from the central star, and infers the presence and properties of orbiting planets through their effects either on the reflex motion of the star or on the flux of stellar light. Both kinds of approaches have advantages and limitations; a program to discover and study other planetary systems would make appropriate use of indirect and direct techniques. Chapter III of this report describes indirect discovery of other planetary systems, and Chapter IV addresses direct discovery.

REFERENCES

1. Weidenschilling, S. J. 1988, in *Meteorites and the Early Solar System*, ed. J. F. Kerridge, and M. S. Matthews (Tucson: University of Arizona Press), p 348.

2. Safronov, V. S. 1969, *Evolution of the protoplanetary cloud and formation of the Earth and the planets* (Moscow: Nauka).

3. Greenberg, R., Wacker, J., Chapman, C. R. and Hartmann, W. K. 1978, *Icarus*, 35, 1.

4. Wetherill, G. W. 1990, *Ann. Rev. Earth Planet. Sci.*, 18, 205.

5. Wetherill, G. W. 1986, in *Origin of the Moon*, eds. W. K. Hartmann, R. J. Phillips, and G. J. Taylor, Houston: LPI, p 519.

6. Bodenheimer, P. and Pollack, J. B. 1986, *Icarus*, 67, 391.

7. Strom, S. E., Edwards, S. and Strom, K. M. 1989, in *The Formation and Evolution of Planetary Systems*, ed. H. A. Weaver, and L. Danly (Cambridge: Cambridge University Press), p 91.

8. Lissauer, J. J. 1987, *Icarus*, 69, 249.

9. Stevenson, D. J. and Lunine, J. I. 1988, *Icarus*, 75, 146.

10. Beckwith, S., and Sargent, A. I. 1992, in *Protostars & Planets III*, ed. E. H. Levy, J. I. Lunine, and M. S. Matthews (Tucson: Univ. Arizona Press), in press.

11. Strom, S. E., et al. 1992, in *Protostars & Planets III*, ed. E. H. Levy, J. I. Lunine, and M. S. Matthews (Tucson: Univ. Arizona Press), in press.

12. Sargent, A. I. and Beckwith, S. 1987, *Astrophy. J.*, 323, 294.

13. Strom, S. E., Strom, K. M. and Edwards, S. 1988, in *Galactic and Extragalactic Star Formation*, ed. R. E. Pudritz, and M. Fich (Dordrecht: Kluwer), p 53.

14. Lin, D. N. C. and Papaloizou, J. 1986, *Astrophy. J.*, 309, 846.

15. Ward, W. R. and Hourigan, K. 1989, *Astrophy. J.*, 347, 490.

16. Adams, F. C., and Shu, F. H. 1985, *Astrophy. J.*, 296, 655.

17. Boss, A. P. 1989, *Astrophy. J.*, 345, 554.

18. Beckwith, S. V. W., Sargent, A. I., Chini, R. S., and Gusten, R. 1990, *Astron. J.*, 99, 924.

19. Horedt, G. P. 1978, *Astron. Astrophy.*, 64, 173.

20. Lynden-Bell, D. and Pringle, J. E. 1974, *Mon. Not. R. Astr. Soc.*, 168, 603.

21. Lin, D. N. C. and Papaloizou, J. 1980, *Mon. Not. R. Astr. Soc.*, 191, 37.

22. Stepinski, T. F. and Levy, E. H. 1990, *Astrophy. J.*, 350, 819.

23. Herbig, G. H. 1977, *Astrophy. J.*, 217, 693.

24. Graham, J. A. and Phillips, A. C. 1987, *Pub. Astron. Soc. Pac.*, 99, 91.

25. Boss, A. P. 1986, *Astrophy. J. Suppl.*, 62, 519.

26. Prinn, R. G., and Fegley, B., Jr. 1989, in *Origin and Evolution of Planetary and Satellite Atmospheres*, ed. S. K. Atreya, J. B. Pollack, and M. S. Matthews (Tucson: Univ. Arizona Press), p 78.

27. Mumma, M. J., Stern, S. A., and Weissman, P. R. 1992, in *Protostars & Planets III*, ed. E. H. Levy, J. I. Lunine, and M. S. Matthews (Tucson: Univ. Arizona Press), in press.

28. Campbell, B., Walker, G. A. H., and Yang, S. 1988, *Astrophy. J.*, 331, 902.

29. Duquennoy, A. and Mayor, M. 1991, *Astron. Astrophy.*, 248, 485.

30. Becklin, E. E. and Zuckerman, B. 1988, *Nature*, 336, 656.

31. Wolszczan, A., and Frail, D. A. 1992, *Nature*, 355, 145.

32. Smith, B. A. and Terrile, R. J. 1984, *Science*, 226, 1421.

33. Sargent, A. I. and Beckwith, S. 1991, *Astrophy. J.*, 382, L31.

34. Reipurth, B., J. Bally, J. A. Graham, A. P. Lane, and W. J. Zealey 1986, *Astron. Astrophy.*, 164, 51.

35. Buhrke, T., Mundt, R., and Ray, T. P. 1988, *Astron. Astrophy.*, 200, 99.

36. Wetherill, G. W. 1991, *Science*, 253, 535.

III. Indirect Detection

3.1 INTRODUCTION

WE NORMALLY THINK OF A PLANET AS MOVING in an orbit around its parent star. But just as the gravity of the star pulls the planet around in its orbit, so the gravity of the planet pulls the star. The main difference is that the star responds more slowly, and thus has a smaller orbit, because of its larger mass. The challenge is to measure the small orbital motions or displacements of the star that result from the pull of its unseen planetary companion, to deduce something about the planet. How massive is it? Does it have a circular orbit, as expected for objects that formed by accretion in a disk? Is there any sign of a second or even a third planet, to make a planetary system? Do all the orbits lie in nearly the same plane? Is the system at all like our own solar system? Could a small planet like Earth be circling in an orbit where water is liquid, and life is comfortable? We have already started with existing technology to make the measurements needed to address these questions, and we can see how to develop the new technology needed to yield fuller answers.

Two complementary approaches to measuring the reflexive motion of a star in orbit around a planet are spectroscopy and astrometry. The spectroscopic approach utilizes careful measurements of Doppler shifts. The component of the orbital motion along the line of sight causes a periodic variation in the Doppler shift of the spectrum, superimposed on the mean Doppler shift due to the steady motion of the star relative to the solar system. The astrometric approach utilizes careful measurements of stellar positions. The orbital motion of a star across the line of sight causes periodic variation in the position of the star as seen on the plane of the sky, superimposed once again on the mean motion of the star relative to the solar system, and with a component that reflects the annual and daily motion of the telescope as it rides around on the Earth.

The more massive a planet is in comparison with its parent star, the easier it is to detect its effect with these techniques. The amplitude of the Doppler shift and the size of the displacement in position are both proportional to the ratio of the mass of the planet to the mass of the star. The apparent size of the star's orbit grows smaller with increasing distance between the star and Earth, so astrometry works best for the nearest stars. On the other hand, the magnitudes of the Doppler shift varia-

tions do not depend on stellar distance, but brighter stars make the spectroscopic work easier, so it is natural to observe the nearest stars here as well. A serious shortcoming of the spectroscopic approach is that it cannot, by itself, lead to a determination of the inclination angle *i* of the orbit relative to the plane of the sky, and individual masses derived from spectroscopic orbits are ambiguous by the unknown factor sin *i*. This problem can be addressed by statistical treatments of samples of planetary systems, or by estimating *i* with supplementary observations.

Another indirect approach to the detection of an unseen planetary companion of a star would be to look for the effects that the presence of the planet might have on the light from the star. When a planet happens to pass between a star and a distant observer, it blocks a measurable amount of the star's light. Or if the system happens to pass almost exactly between the observer and a more distant star, the effect of the planet can be detected as a distortion in the resulting gravitational lensing.

Any technique capable of timing a star's motion projected along the line of sight can serve as a basis for indirectly discovering orbiting planets. As this report goes to press, observations of the millisecond pulsar PSR1257+12 indicate motional disturbances possibly attributable to the influence of two planet-like objects orbiting the neutron star.[1]

Measuring the small stellar motions induced by planetary gravity presents significant, but manageable, technical challenges. The rewards will be large; the dynamical information obtainable through indirect detection will be highly diagnostic of a planetary system's structural regularities and will make important contributions toward understanding planetary system formation processes. This chapter focuses on the role of spectroscopic and astrometric techniques in a program to discover and study other planetary systems. This chapter also briefly discusses photometry and other innovative approaches that are not highly developed; however, there should always be an awareness that unusual approaches can lead to significant results.

SEARCH SPACE

One way to assess the capabilities of a given planet-detection technique or instrument is to define a desirable "search space" and then determine how much of that search space the technique or instrument can cover. An obvious parameter would be the type of parent star. If we are

interested in learning how planets form, we might argue that we should study all types of stars, covering as wide a range as possible for the masses and temperatures that we find for stars after they settle down to burn hydrogen on the main sequence. If we are more interested in knowing the frequency and characteristics of planetary systems around stars old enough to allow the long period of time necessary for the evolution of life, we might concentrate our studies on stars older than, for example, a billion years—the F, G, K, and M stars. If we are interested in gathering information on a large sample of stars, we might choose the most common type of nearby stars, the M stars. If we are concerned that the presence of a second star might be important, even if it is in a rather distant orbit compared to the planet, we need to know the details of which stars are binaries. We will see below that a dense "forest" of deep lines is necessary in the spectrum of a star for high accuracy in determining changes in radial velocity. For this reason O, B, and A stars are poor candidates for precise radial velocity determinations, since there are relatively few lines, and these lines are shallow with small slopes. However, these early stars remain candidates for astrometric searches.

Table 3-1. Type and Distribution of Nearby Stars

Sp.	M_V	$M_S/M_\odot$	T_{eff} *(K)*	$N_{tot}(10)$	$N_S(10)$
A0	*0.7*	*3.4*	*9900*	*2*	
A5	*2.0*	*2.2*	*8500*		
F0	*2.6*	*1.8*	*7400*	*11*	*5*
F5	*3.4*	*1.4*	*6580*		
G0	*4.4*	*1.1*	*6030*	*26*	*13*
G5	*5.1*	*0.9*	*5520*		
K0	*5.9*	*0.8*	*4900*	*42*	*18*
K5	*7.3*	*0.6*	*4130*		
M0	*9.0*	*0.5*	*3480*	*210*	*63*
M5	*11.8*	*0.3*	*2800*		

Table 3-1 shows the classes of stars within 10 pc of the Sun. Most nearby stars are cool, low-mass M stars. The last column in Table 3-1 gives the number of F, G, K, and M stars that are not known to be

members of a binary or multiple-star system.[2,3] Fewer than half the nearby stars may actually be single. In many binaries, the member stars are so widely separated that planetary systems could have formed and persisted in much the same way as single stars. For many other binaries, it is likely that planets, even had they formed, would have had their orbits rapidly disrupted. For nearby stars, the numbers, both N_{tot} and N_S, should scale as the volume of the sample space, i.e., as the limiting distance cubed. For example, one can estimate from Table 3-1 that a sample out to about 30 pc would be required to get a total of about 1000 "single" F, G, and K stars, whereas 500 single M stars would lie within about 20 pc.

Planetary motions are governed by Kepler's third law as revised by Newton:

$$a^3 = (M_* + m_P)P^2, \tag{3.1}$$

where the masses are expressed in solar masses, the semimajor axis a is in astronomical units (AU), and the period P is in years. Normally, the planetary mass m_P is so small in comparison with the stellar mass M_* that it can be ignored here. The indirect dynamical—spectroscopic radial-velocity and astrometric positional—methods reveal the primary dynamical characteristics of discovered planetary systems: the planetary mass (as a fraction of the star mass) and the orbit characteristics.

What range of periods can be covered by a planetary search program that lasts, for example, 10 years? Toward shorter periods, the astrometric approach loses sensitivity, because the orbital displacement drops off as $P^{2/3}$, whereas toward longer periods the spectroscopic approach loses sensitivity, because the orbital velocity drops off as $P^{-1/3}$. Thus, spectroscopy should be able to detect planets with periods down to the limit where the planet could not survive the gravitational forces of the parent star, or where it could not form so close to the star in the first place. The velocity measurements can be spaced out much more widely than the orbit period itself; with modern period-search techniques, it is straightforward to determine short periods from unevenly spaced data, as long as enough observations are distributed over the orbital phase.

Figure 3-1 illustrates where planets like those of our solar system fall relative to astrometric signatures of 10^{-5} and 10^{-4} arcsec for stars of 0.2 and 1.0 solar masses at a distance of 10 parsecs. The diagonal lines of constant astrometric signature follow from Equation (3.2), and all

points above these lines represent larger signatures. Measurements with accuracies comparable to these signatures are projected for space-based and ground-based instruments respectively. The prospects for indirect detection of planets through radial velocity measurements depend on different circumstances. The less massive the star, the larger the velocity anomaly, and planets orbiting with small semimajor axes are most easily detected. Lines of a constant radial velocity signature of 5 m s^{1} with sin i = 1.0 are also shown in Figure 3-1 for 0.2 and 1.0 solar mass stars with any point above such a line corresponding to a larger signature. The radial velocity signatures would be correspondingly reduced for $i < 90°$. These lines follow from Equation (3.3) below. Signatures for solar-system-like planets, which are independent of distance, can be estimated from Figure 3-1. Measurements comparable to this signature have been demonstrated for bright targets over relatively short time intervals. Radial velocity and astrometry techniques cover complementary parts of the search space.

The relationship between the astrometric measurement uncertainty

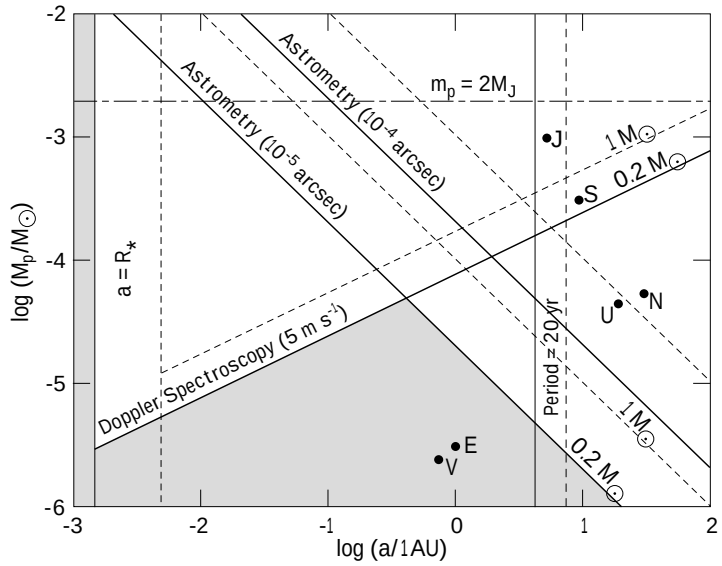


Figure 3-1. The discovery space for solar-system-like planets around stars of 0.2 $M_{\odot}$ and 1.0 $M_{\odot}$ at 10 parsecs. Representative lines are shown for astrometric signatures of 10^{-4} and 10^{-5} arcsec as well as for a radial velocity signature of 5 m s^{1} where sin i = 1.0. Minimum and maximum values of the semimajor axis, set by the stellar radius and an orbit period of 20 years, are also shown.

of any observing instrument and the minimum detectable astrometric signature (amplitude of the periodic angular displacement of a star on the sky) due to a planetary perturbation is a complex function of the period of the orbit and the sampling window (the combined effects of the number, the spacing, and the total time baseline of the observations).[4] The limiting astrometric signatures in Figure 3-1 are expected to be detectable with the projected comparable single measurement accuracies, but only with a sufficient number of measurements over at least an orbital period and with other constraints imposed by the Black and Scargle analysis.[4] Section 3.2 discusses the prospects for various accuracy limits for both ground-based and space-based astrometry programs. The actual detection of a radial velocity signature with a single measurement accuracy comparable to the 5 m s^{-1} line of Figure 3-1 requires a large number of measurements over at least an orbital period and tight control and calibration of systematic errors. Section 3.3 discusses the current state of the art for radial velocity measurements and current and future accuracy limits.

If measurement errors are small compared with the amplitude of orbital motion, the orbital parameters can be obtained from observations lasting less than a full period. However, when the amplitude of orbital motion is not significantly larger than the measurement errors, dozens of observations covering at least one and probably several orbital cycles are necessary to obtain a reliable orbital solution. In our solar system, Jupiter orbits the Sun in about 12 years, and Saturn orbits the Sun in about 30 years. These times define the minimal duration of a planet search to be approximately a decade. Lower-mass stars—such as the M stars, which are most numerous in the solar neighborhood—may have their major planets in faster orbits, allowing definitive studies to be carried out in somewhat less than a decade. Once a planet has been detected around another star, motivation will be strong to mount intensive investigations of that system using both indirect and direct detection methods. A primary early objective will be to ascertain the structure of the system to discover regularities that give clues about formation processes, as discussed in Chapter II.

DISCOVERY SPACE

Astrometric Signatures

The motion of a single planet in a circular orbit around a star, with the parameters given by Equation (3.1), causes the star to undergo a reflexive circular motion around the star-planet barycenter. When projected on the sky, the orbit of the star appears as an ellipse with angular semimajor axis θ given by the expression

$$\theta = \frac{m_P}{M_*} \frac{a}{r} = \frac{m_P}{r} \left(\frac{P}{M_*} \right)^{2/3}, \tag{3.2}$$

where θ is in arcsec when the semimajor axis a is in AU, the mass of the planet (m_P) and the mass of the star (M_*) are in solar masses, the distance r is in pc, and the orbital period P is in years. For example, observing the solar system from a distance of 10 pc, the presence of Jupiter would be revealed as an 11.9-year disturbance in the Sun's motion, with an amplitude of 0.5 milliarcsec. However, if Uranus were the largest planet in the solar system, the amplitude would be only 84 microarcsec with a period of 84 years. If Earth were the only planet circling the Sun, the amplitude of the Sun's motions would be 0.3 microarcsec, with a period of 1 year. For the same planets in the same orbits around a typical M star, the amplitude of stellar motion would be three or four times larger, and the periods would be considerably longer. If the astrometric precision of a single measurement of a star's position is 10 microarcsec, as is projected for TOPS-1, then a full program of observations should detect planets the size of Uranus, in orbits with periods ranging from 3 or 4 years to as long as the duration of the observing program, around stars like the Sun.

Astrometry is most sensitive to nearby planetary systems with massive planets in orbits near the water-ice condensation boundary. This is complementary to radial-velocity techniques, which are optimally sensitive to objects in smaller, shorter-period orbits. In a system with several planets, the actual deviation of the star's motion can be complicated. Figure 3-2 shows the apparent path of the Sun as viewed from a distance of 10 pc from the ecliptic north pole over a period of 65 years from 1960 to 2025 AD. The entire pattern is less than 2 milliarcsec across. If viewed from a vantage point in the plane of the orbit, the apparent path of the Sun would show only one component of this motion. The largest planet in the system is likely to be the first seen. Systematic investigation of the

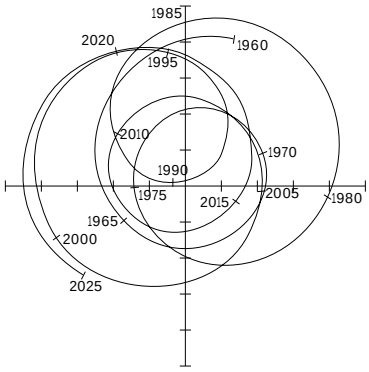


Figure 3-2. The apparent orbit of the Sun around the center of mass of the solar system from 1960 to 2025 AD, as viewed from the north ecliptic pole at a distance of 10 parsecs. Each axis runs from -1.0 milliarcsec to +1.0 milliarcsec, and the tic marks on the axes are in intervals of 0.2 milliarcsec. The positive horizontal axis is in the direction of the Vernal Equinox. These positions are from the JPL DE200 planetary ephemeris.

system over time and with increased sensitivity would reveal the presence and characteristics of the other planets.

The search space illustrated for planets about 0.2 and 1.0 solar mass in Figure 3-1 can be generalized to a variety of stellar masses, spectral types, and planet characteristics. One such representation is shown in Figure 3-3. As an example in this figure, the box indicates a stellar mass range chosen to lie between 0.2 and 1.8 solar masses, corresponding to spectral types M5 to F0, as marked by the vertical lines of the box; the period range has also been chosen for illustration to lie between 0.5 and 15 years, as marked by the dotted horizontal lines. To display the relationship between the period and the size of a planet's orbit in the search space, it is useful to add an axis for the semimajor axis a in Figure 3-3. From Equation (3.1), lines of constant a have a slope of -1/2 in the M_* *vs* P plane. Thus, the axis for a in Figure 3-3 has a slope of 2, with a scale that follows from Equation (3.1).

Figure 3-3 can be used to display the astrometric signature that planets with the mass of Jupiter (10^{-3} solar masses) or Uranus (4.4×10^{-5} solar masses) would have if viewed from a distance of 10 pc. From Equation (3.2), we see that lines of constant θ have a slope of 1 in the M_* *vs* P plane. Thus the axes for θ_J and θ_U in Figure 3-3 have a slope of -1, with a scale for each planet that follows from Equation (3.2) using the masses quoted above. With the aid of the θ_U and θ_J axes, astrometric precision of 40 microarcsec can be seen to be sufficient to allow all Jupiter-sized planets with periods longer than 0.5 year to be detected around all the F, G, K, and M stars within 10 pc, whereas 2 microarcsec accuracy would be required to detect all the Uranus-sized planets. As discussed above,

this assumes that orbits can be detected down to amplitudes as small as the error in a single measurement, which requires a rich time series of observations. The best ground-based astrometric measurements currently range near an accuracy of about 1 milliarcsec.

Radial-Velocity Signatures

When projected along the line of sight, the orbital motion of a star appears as a variable radial velocity with amplitude

$$v = 30 \frac{m_P \sin i}{(a M_*)^{1/2}} = 30 \frac{m_P \sin i}{M_*^{2/3} P^{1/3}}, \qquad (3.3)$$

where v is in km s^{-1}, m_P and M_* are in solar masses, a is in AU, P is in years, and i is the inclination of the orbit to the plane of the sky. For example, when the solar system is viewed from a point in the plane of the

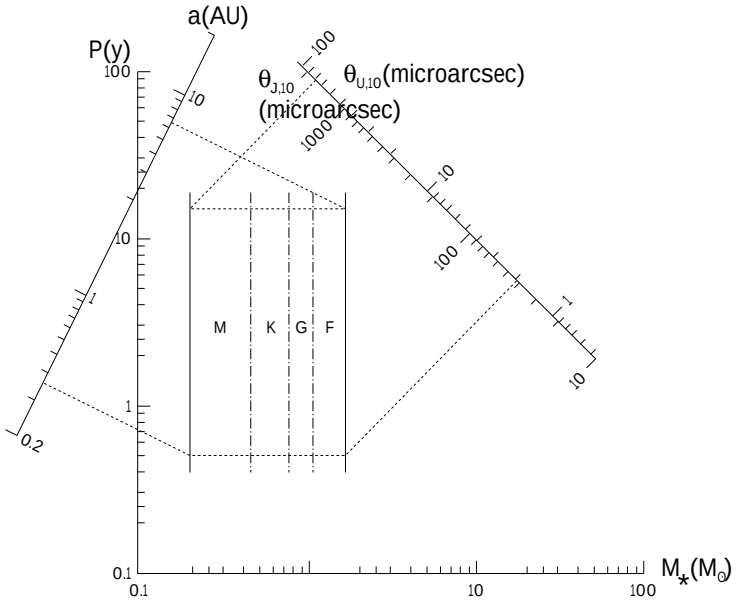


Figure 3-3. Representative search space defined by stellar mass and period of a planetary orbit. The inclined axes are drawn perpendicular to lines of constant semimajor axis and constant astrometric signature respectively. The scale for the semimajor axes is determined from Equation (3.1), and that for the astrometric signature is determined from Equation (3.2) for planetary masses equal to those of Uranus and Jupiter. The astrometric signature axis is for a system at 10 pc.

planetary orbits, the response of the Sun to the pull of Jupiter is a velocity variation of 13 m s^{-1} with a period of 11.9 years, whereas Uranus only contributes a motion of 0.3 m s^{-1} with a period of 84 years. The response of the Sun to the pull of the Earth is a minuscule 0.09 m s^{-1}.

As mentioned earlier, radial-velocity measurements are independent of the distance to the star, but require sufficient photons. Despite their brightness, giant stars are generally poor candidates for this technique because they show intrinsic velocity variations with amplitudes ranging from a few hundred m s^{-1} to a few km s^{-1}, and time scales ranging from 1 week to 2 years.[5,6,7] These variations appear to be due to stellar oscillations and atmospheric motions. Dwarf stars like the Sun do not seem to display such intrinsic variations. Evidence suggests that the intrinsic velocity amplitudes for cool dwarfs are on the order of a few m s^{-1}, although this remains somewhat uncertain. Generally speaking, the interpretation of any low-amplitude velocity variation requires careful

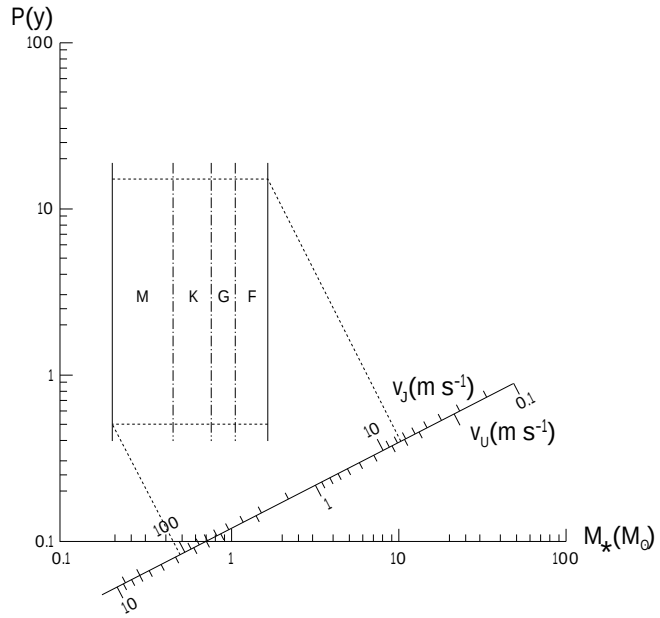


Figure 3-4. Search space as depicted in Figure 3-3, but with the inclined axis drawn perpendicular to lines of constant radial velocity signature. The scale for the radial velocity signatures is determined from Equation (3.3) for planetary masses equal to those of Uranus and Jupiter.

ancillary studies to avoid mistaking possible inherent velocity variations for planetary disturbances.

Figure 3-3 can be further generalized by adding velocity information as shown in Figure 3-4, which also shows radial-velocity signatures produced in stars with Jupiter- or Uranus-mass planets. However, in this case, the amplitude of the disturbance is independent of the star's distance. From Equation (3.3) we see that lines of constant v have a slope of -2 in the M_* *vs* P plane. Thus, the axes for v_J and v_U in Figure 3-3 have a slope of 1/2, with a scale for each planet that follows from Equation (3.3) using the masses quoted above. With the aid of the v_U and v_J axes, it is easy to see that a radial-velocity precision of 8 m s^{-1} is sufficient, through a sequence of many measurements, to allow all Jupiter-sized planets with periods shorter than 15 years, and with reasonably favorable inclinations, to be detected around all the F, G, K, and M stars in the search space bright enough to be observed, whereas 0.4 m s^{-1} accuracy would be required to detect all the Uranus-sized planets. Figure 3-5 shows the radial velocity that would be observed for the Sun if the solar

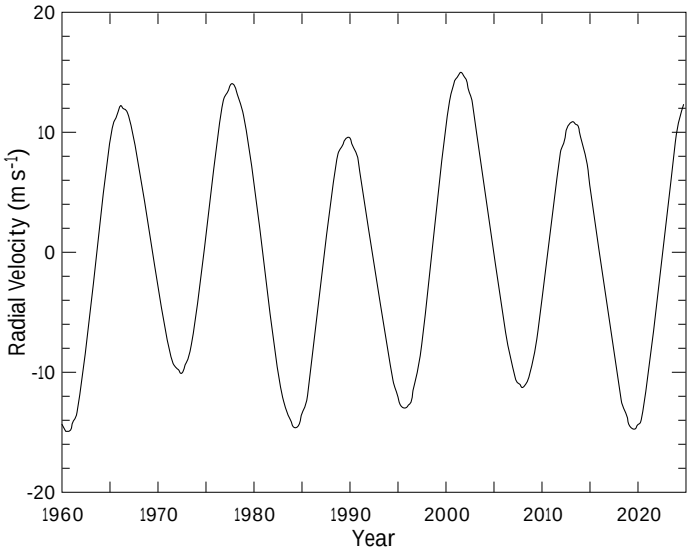


Figure 3-5. The apparent radial velocity of the Sun as it orbits the center of mass of the solar system from 1960 to 2025 AD, as viewed from the direction of the Vernal Equinox (from the positive horizontal axis in Figure 3-2). The amplitude of the observed signal is independent of the distance from which the system is viewed.

system with its nine planets were viewed from the direction of the positive horizontal axis in Figure 3-2 over the 65-year period from 1960 to 2025. Note the dominance of Jupiter's period, but with deviations caused by the other planets.

Doppler measurements alone reveal only the product $m_P \sin i$. However, methods are available for constraining the value of $\sin i$ and thereby getting a better estimate of m_P. Experience with the solar system and current theories of planetary system formation support the assumption that planetary orbits will coincide closely with a star's equatorial plane. Independent astrophysical means exist to estimate the inclination of a star's rotation axis. Stellar absorption lines are broadened by rotation in a way that permits an estimate of $v_{eq} \sin i$, where v_{eq} is the equatorial rotational velocity of the star. If the period of stellar rotation can be determined from starspot modulation of the photometric signal or rotational modulation of the CaII H and K line emission,[8] v_{eq} can be inferred from the estimated radius of that particular spectral type star, and $\sin i$ can be estimated for at least a reasonable fraction of discovered systems, thus allowing an estimate of planet masses themselves.

For many systems, both spectroscopic and astrometric methods will be applicable, especially when astrometry reaches the 5 to 10 microarcsec accuracy objective anticipated for TOPS-1. Where possible, the application of both techniques will produce especially incisive dynamical information about the planet masses and orbit parameters. Spectroscopic Doppler searches can effectively be carried out now from the ground, as part of TOPS-0.

Number of Stars Observable

Both radial-velocity and astrometric techniques lose sensitivity as the target and, for the latter, reference stars become fainter. For targets fainter than some limiting magnitude (depending, of course, on the size of the telescope), the signal-to-noise ratio will become too small to permit meaningful interpretation with a limited number of observations. Table 3-2 gives the absolute visual magnitudes of stars in the search space, the distance at which a star of given spectral type has an apparent magnitude of 10, and the approximate number of stars of the particular spectral type within that distance. The properties of the stars in the middle of the spectral range (e.g., G5 for the G stars) were used to determine the number of observable stars given in the last column of

Table 3-2. The number of target stars that can be reached with the techniques described herein is far larger than can be managed in a ground-based program or in the finite lifetime of an orbiting space-based system. For example, current observing programs are limited primarily by the availability of telescope time and manpower.

Table 3-2. Number of stars observable if $m_V = 10$ is the limiting magnitude.

Sp	M_V	r_{max}*(pc)*	*stars*
M5	*11.8*	*4.4*	*~ 5*
M0	*9.0*	*15.8*	
K5	*7.3*	*34.7*	*~ 750*
K0	*5.9*	*66.0*	
G5	*5.1*	*95*	*~ 11,000*
G0	*4.4*	*131*	
F5	*3.4*	*208*	*~ 45,000*
F0	*2.6*	*302*	

3.2 ASTROMETRY

DESCRIPTIONS OF TECHNIQUES

In astrometric planet searches, a star's position is accurately measured over a period of time and against a suitable reference system to ascertain whether its motion is linear (inertial), or whether it shows the telltale (noninertial) variations produced by orbiting planets. Central to all astrometry is the question of the reference frame; two problems must be addressed. First, the reference frame stars move and thus introduce errors in the apparent motion of the target stars. Second, the reference frame stars are often faint, so that their small collective photon rate may limit the precision of the measurement. This problem is exacerbated in classical astrometry by the need to determine the plate scale and other parameters from the same stars used for the reference frame. Stellar positions can be measured by several techniques, utilizing either single-aperture traditional telescopes or multiple-aperture interferometric instruments. This section discusses first the potential of ground-based astrometric planet detection and then the potential of space-based astrometric investigations.

Single-Aperture Astrometry

Temporarily neglecting the question of the reference stars, the preci-

sion of an "ideal" space-borne astrometric telescope is

$$\sigma \approx \frac{\lambda}{D\sqrt{n}} \tag{3.4}$$

where σ is the photon-statistics-limited standard error of the target star's measured position, D is the telescope aperture, λ is the central wavelength, and n is the number of detected photons. Equation (3.4) assumes that the diffraction-limited star image is cut through its center and allowed to fall on two photon-counting detectors. For off-center division of the image, the precision is degraded. This degradation of precision is important for realizable instruments, as discussed below.

The reference frame is defined by some carefully chosen set of stars, usually in close angular proximity to the target. Either the reference stars must be sufficiently far away to show negligible proper or parallactic motion themselves, or that motion must be accommodated in the data reduction. The Ritchey-Chrétien reflector design is typical of single-field astrometric telescopes. Without a near-focal-plane corrector, which would introduce large astrometric errors, the astrometrically useful field of a 2-meter aperture Ritchey-Chrétien is typically limited to approximately 20 arcmin. The 20 or so brightest stars in that field collectively yield photons at a rate equivalent to a single magnitude-11 star. Thus, for a target brighter than magnitude 11, the error budget from photon statistics is dominated by the reference frame.

In one type of astrometric telescope, a Ronchi ruling is moved in the focal plane over the star-field image to modulate the starlight. The ruling has alternating stripes, clear and opaque, and it is moved perpendicular to those stripes. Detectors are placed to receive the modulated light from the target star and each of several reference-frame stars. The relative phase between pairs of starlight signals yields the corresponding star separation, modulo the angle equivalent of a cycle on the ruling.

In one frequently considered design, the clear and opaque stripes on the Ronchi ruling are equally wide and about the same size as the diffraction-limited star image. Half the starlight is lost by the opaque stripes and, as previously noted, the information rate is lowered because the image-splitting line is not always centered on the image. These factors together increase the integration time for a given accuracy by a factor of about ten.

We next consider a 2-meter Ronchi-ruling astrometric telescope

with a large optical bandpass centered at $\lambda = 0.5$ μm. Since the target stars are of magnitude 9 or brighter, the reference-frame stars' collective magnitude of about 11 dominates the integration time. With a 20% probability of photon detection based on detector efficiency and light loss due to optics other than the Ronchi ruling, about 10^8 photons are detected per minute. By using Equation (3.4) and applying the ten-fold reduction discussed above, we find the photon-statistics limit to measurement precision is 10 microarcsec from a 1-minute observation. Including the effect of reasonable optical bandwidth increases the required integration time for this precision by less than a factor of two.

Atmospheric Disturbances

Space-based instruments, free from the disturbing effects of Earth's atmosphere, will provide the ultimate tools for the discovery and study of other planetary systems. Ground-based astrometric instruments have several significant sources of error, chief among which are differential refraction and atmospheric seeing. Refraction is variable and dependent upon wavelength. Variable atmospheric conditions or variations that cause the system's wavelength-sensitivity profile to change can cause significant changes in the apparent relative positions of stars of different temperatures. Figure 3-6, a plot of the difference in the apparent zenith distances of stars as a function of their zenith distance and the wavelength of the measurement, illustrates the astrometric effect of differential refraction. Under typical observing conditions at sea level, the ultraviolet image of a star may be displaced by several arcsec from the far red image. A number of measures can be taken to limit or calibrate these effects of differential refraction. Even in space-based systems, differential color effects can arise in the instrument itself, and similar measures can

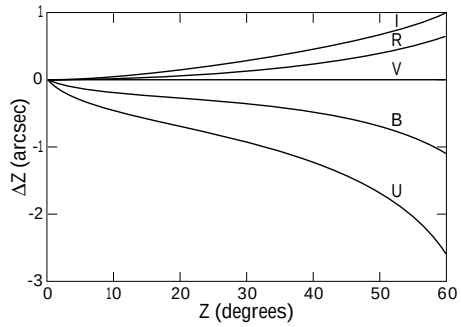


Figure 3-6. Variation of astronomical refraction with wavelength and zenith angle (relative to visual wavelengths).

be utilized to control these effects.

An important limiting error source in ground-based astrometry is the image jitter introduced by atmospheric "seeing." The amount of image motion, often a few tenths of an arcsec, varies with the conditions of the atmosphere and the telescope site. Mountaintop sites, due in part to the fact that they have less atmosphere above them, usually suffer less from this effect than low-altitude sites. Figure 3-7 shows the relative "jitter" of star images as they trail across the fields of telescopes located on Mauna Kea and at the Allegheny Observatory in Pittsburgh. The seeing-induced motion at Mauna Kea is approximately 25% that at Pittsburgh. Seeing studies conducted at the Mauna Kea and Pittsburgh sites reveal the following: the diameters of the images of point sources are 0.93 and 2.2 arcsec respectively; stellar image motions at the two sites yield average values of 0.085 and 0.275 arcsec respectively; and the standard deviations of separations of stars (normalized to 10 arcmin) indicate relative precisions of 0.054 arcsec and 0.143 arcsec respectively in 1 second of observing time.

According to the theory of atmospheric turbulence,[9] the relative image motion σ as a function of the seeing conditions, separation S and integration time T, can be represented by:

$$\sigma \equiv \sigma_0 \left(\frac{S}{S_0}\right)^a \left(\frac{T_0}{T}\right)^{1/2}. \qquad (3.5)$$

σ_0 is a constant, which depends on the seeing (i.e., the strength of the turbulence), and a is approximately 1/3.[10] S_0 is a normalization constant for the separation. In the following, we adopt $S_0 = 600$ arcsec, and $T_0 = 1$ sec. Star-trail data taken with the Canada-France-Hawaii Telescope on

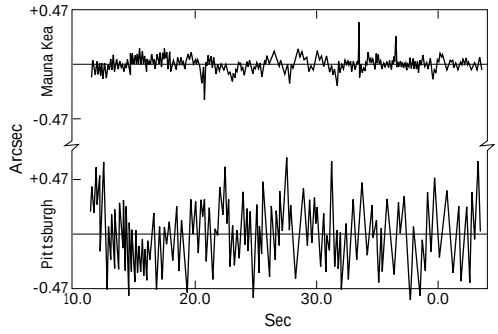
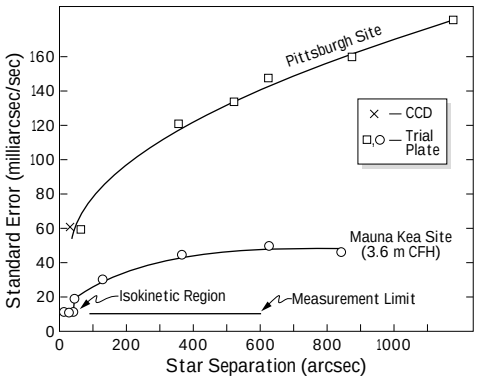


Figure 3-7. Relative image motion for trailed stars at Mauna Kea and Pittsburgh.

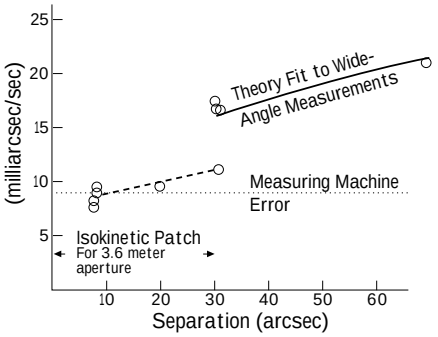
Figure 3-8. Relative image motion of star pairs as a function of their separation.



Mauna Kea are plotted in Figure 3-8, which shows the relative image motion of star pairs as a function of their separations. A least-squares fit to these data in the analytic form of Equation (3.5) yields σ_0 = 0.054 arcsec and a = 0.35.

Over most of its range, Figure 3-8 shows general agreement with Equation (3.5) with the empirically derived values of the constants. Note, however, that the small portion near the origin shows different behavior. The isoplanatic patch (the angular area perpendicular to the direction to the star over which the phase remains within 1 radian of that of the plane wave) grows at the rate of approximately 1 arcsec per inch of aperture.[11] This implies that Equation (3.5) should have an additional term that is dominant at small separations and is directly dependent upon aperture. To illustrate this effect, we have enlarged the central portion of Figure 3-8 in Figure 3-8a. At separations of 30 arcsec or less, the observed relative motions are substantially smaller than predicted by

Figure 3-8a. Magnification of the isokinetic region of relative image motion of star pairs in Figure 3-8.



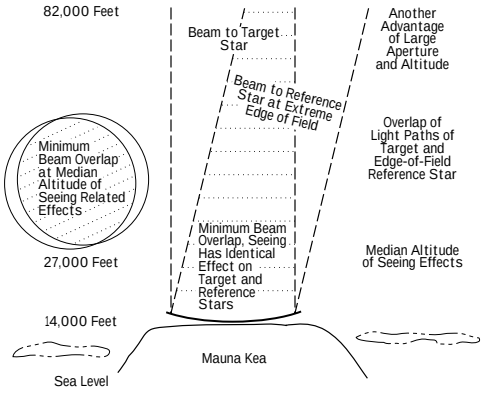


Figure 3-9. Overlap of beams to two stars from a telescope on Mauna Kea.

Equation (3.6). Allowing for the error of the measuring device (approximately 0.5 micron), the position error measured for separations from 0 to 30 arcsec is approximately 8 milliarcsec of error (with a 1-second integration), one-third of the predicted error. This central area of the field is called the isokinetic patch.

Figure 3-9 sketches two starlight beams reaching a telescope after passing through the Earth's atmosphere above Mauna Kea. For the Canada-France-Hawaii 3.6-meter telescope observing stars sufficiently closely spaced, these beams will have some overlap at the top of the atmosphere's disturbing layers. The present data, and Figure 3-9, suggest that the maximum overlap angle for the 3.6-meter telescope on Mauna Kea is 30 arcsec. The same geometry suggests that the Keck 10-meter telescope's isokinetic patch will be approximately 80 arcsec in radius. The potential of such instrumentation remains to be determined from use, but it is not unreasonable to anticipate that astrometric precisions of the order of 100 microarcsec might be achievable in an hour of observation.

Interferometric Astrometry

Combining the starlight received in two subapertures separated by a distance *B*, at least several times larger than their individual diameters *D*, results in angular resolution and astrometric measurement precision better than that achievable with a single aperture of the same area. If we assume an equal number of detected photons and photon statistics of the target star defining the limiting error for both instruments, the perfor-

mance ratio is $\sigma_T/\sigma_I = 3\pi^2 B/8D \approx 3.7B/D$.[12] The accuracy attainable in real instruments may contain significant contributions from other error sources.

Figure 3-10 schematically shows the operation of an astrometric interferometer. Light from the target star is collected by two subapertures and routed via mirrors to a beamsplitter (partially reflective mirror) where the two beams are combined. This combined beam will exhibit constructive and destructive interference; the interference will be at a maximum if there are equal optical path lengths from the source to the beamsplitter via the two arms. If the source direction is shifted relative to the interferometer baseline, an additional path delay results in one beam external to the interferometer. This path delay must in principle be compensated by an equal amount of path delay in the other beam internal to the interferometer to maintain the maximum interference. This relationship can be written

$$x = \vec{B} \bullet \hat{s} + C, \tag{3.6}$$

where $\vec{B}$ is the baseline vector—essentially the vector connecting the two subapertures—$\hat{s}$ is the unit vector to the star, C is a constant (instrumental bias), and the delay x is the amount of internal path length necessary to equalize the path delays. Thus, the delay x is a measure of the angle between the interferometer baseline and the star unit vector. The measured value of x is the sum of the offset of the interference fringe

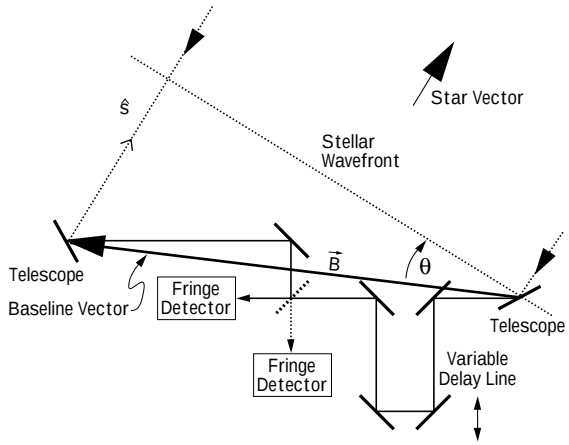


Figure 3-10. Astrometry with an interferometer. The delay at which fringes are observed is $\vec{B} \bullet \hat{s} + C = |B| \sin\theta + C$.

from the point of maximum interference and the additional internal optical path added to approximately equalize the delay.

The photon-noise error in measuring stellar position with an interferometer depends on both the subaperture diameter D and the baseline length B. If n is the total number of detected photons, the photon-noise-induced astrometric error is

$$\sigma = \frac{\lambda}{2\pi B\sqrt{n}}, \tag{3.7}$$

where λ is the wavelength. However, n is proportional to the collecting area d^2, so that $\sigma \propto 1/Bd$, as compared with λ/D^2 for a filled-aperture telescope of diameter D. Thus, an interferometer, even with relatively small subapertures and a modest baseline length, can have a photon-noise-limited accuracy equivalent to a large-diameter telescope.

The interferometer is completely modeled by four parameters: the three-dimensional baseline vector B and the constant C. These parameters can be measured with laser interferometry to monitor systematic errors. Because the interferometer model does not depend on field of view, wide-angle astrometry is possible, in principle, with a single instrument. The observational scenario for ground-based interferometric astrometry differs for wide-angle and narrow-angle astrometry. For wide-field astrometry, instruments such as the Mark III stellar interferometer[13] switch rapidly and repeatedly through a star list of 5 to 15 stars. From the resulting measurements, both the instrument parameters and the relative star positions can be determined. The accuracy achieved with such wide-angle techniques is 5 to 10 milliarcsec over several nights. However, achieving the accuracies necessary for planet detection requires narrow fields—well within the isokinetic patch—as well as simultaneous measurements so that the common-mode atmospheric errors will cancel. Thus, for narrow-angle interferometric astrometry from the ground, the optical feed at each subaperture must be designed to direct the light from both the target and reference star(s) to a central beam combiner, where they are observed simultaneously. Interferometer concepts for space typically employ interferometers on a platform connected by laser metrology; the reference star may be observed simultaneously or sequentially.

GROUND-BASED ASTROMETRY: EXISTING PROGRAMS AND PROJECTED IMPROVEMENTS

Existing Ground-Based Planet Searches

The only current ground-based search for other planetary systems uses the Multichannel Astrometric Photometer (MAP)[14] on the Thaw refractor.[15] In its fifth year, the observing program includes 16 observations per year of each of 16 stars. The Thaw MAP has an average accuracy (in the sense of conformity to the best model) per night of 2.8 milliarcsec and regularly obtains parallaxes and annual normal points with a precision of about 1 milliarcsec. Accuracy has been shown to depend on (1) the photon count rate of the target star (as well as the reference frame stars), (2) the duration of observation, (3) the angular size of the reference frame, and (4) the quality of the astronomical seeing.[14,16]

The MAP determines stellar positions with respect to reference frames defined by a number of nearby background stars in each field.[14] The MAP metric is invested in a Ronchi ruling manufactured on a flat piece of glass ruled with evenly spaced stripes running across its width (spanning the telescope's useful field). In the current ground-based implementation, the clear and opaque stripes are of equal width, each approximately equal to the diameter containing 90% of the light of a stellar image in the telescope's focal plane. The Ronchi ruling is normally considerably longer than it is wide, containing many lines. In use, the ruling is pulled through the focal plane so that the lines alternately block, then pass, the light of each star. The light from selected stars—the target and some number of reference stars—is subsequently measured, each in an assigned detector.

As the ruling passes over a field, the light falling on each detector varies, nearly sinusoidally, with the same period but generally different phases. The relative phase between detected signals is determined several thousand times per hour of observation, giving a well-averaged determination of the relative positions of stars in the field. (The whole number of lines between images is easily obtained.) The perpendicular axis is measured in a similar fashion, after rotating the instrument through 90°.

Current Ronchi-ruling manufacturing techniques yield rulings with precisions exceeding that needed in any ground-based observing program. Moreover, since the stars move slowly, during any one session they may be used as fixed points of comparison. Then, each time a phase is ascertained with respect to the other stars in the field, a measure of the

position of a line on the ruling with respect to all the other lines then in use over the ruling is obtained; this feature of the measurement technique should allow the system metric to be calibrated during measurement itself, eventually establishing a model of the metric to a precision better than required for the measurements.

The only currently operating astrometric planet search has sufficient sensitivity to detect Jupiter-like planets orbiting nearby stars.[14] If several are detected, they will be suggestive of patterns involving the characteristics of both the planets and the stars around which they formed. If no planetary objects are discovered in the current systematic search, that would begin to place a significant limit on the prevalence of planetary systems similar to our own. The program size is severely limited by the observing time available with the present instruments.

Astrometry with the Keck Telescopes

Single-Aperture Astrometry

The unique aperture, site, and suite of telescopes projected for the Keck Observatory on Mauna Kea, Hawaii, combine to create the promise of astrometric measurements previously considered beyond the reach of ground-based systems. Mauna Kea is characterized by unusually high-quality seeing. Moreover, the Keck Telescope is expected to have an unusually large isokinetic patch, because of its 4.2-kilometer altitude and large aperture. In astrometric measurements, the reference-frame stars force the field of view to be larger than would otherwise be desirable. The large aperture of the Keck Telescope allows dim stars near the target to be used for reference. With the Keck Telescope, the light from two stars separated by as much as 80 arcsec shares some parts of their air columns to an altitude 100,000 feet above sea level. As a result, atmospheric refractive effects along both starlight paths are correlated. Atmospheric seeing can be represented by horizontal turbulence layers at a dominant altitude of approximately 8.2 kilometers above sea level, slightly less than twice the altitude of the Keck site.[11] Two beams of starlight converging into the Keck 10-meter aperture at an angular separation of 80 arcsec would be spatially separated a distance of only 1.6 meters at an altitude of 8.2 kilometers. A target and reference star with such an angular separation would have 80% beam area overlap.

Geometrical considerations of this nature suggest that the astrometric isokinetic patch, observed to be 30 arcsec in radius with the 3.6-

meter Canada-France-Hawaii Telescope (CFHT) on Mauna Kea, could have a radius of 80 arcsec with the Keck Telescope, although this remains to be verified observationally. The CFHT data suggest the possibility that relative astrometric precision for target and reference stars falling within 80 arcsec of one another could be of the order of 0.1 milliarcsec per coordinate axis in 1 hour. Confining measurements to the central 67% of this field radius, the average field can be expected to contain 30 reference stars of magnitude 21 and brighter, yielding approximately 5 million reference-frame photons per second. In the presence of 0.5 arcsec seeing, initial modelling of the remaining atmospheric effects suggests that an actual hourly astrometric precision of 0.1 milliarcsec could be achieved for stars as faint as magnitude 16, and with errors less than 0.2 milliarcsec at magnitude 21. These estimated precisions would be sufficient to detect, at 3 standard deviations, Jupiter-mass planets, in decade or longer orbits, around stars within 12 parsecs (40 light-years) of the Sun.

Astrometrically, the most challenging aspect of the Keck optical design appears to be distortion caused by the fast primary mirror, $f/1.75$. A Ritchey-Chrétien of this design has a distortion-induced displacement of only 0.000075 arcsec at the edge of a 1 arcmin field. Although this displacement is too small in itself to cause serious problems, the interaction of distortion with the coma caused by secondary mirror decentration and/or tilt can cause significant linear coma and third-order and higher terms in the model that relate the positions of the star images to the positions of the stars on the celestial sphere.[17] However, because the secondary mirror of the Keck telescope is actively controlled, tight control on its centration and tilt is feasible and seems adequate to resolve major problems in this area. Even with the Keck Telescope on Mauna Kea, differential atmospheric refraction is a source of serious astrometric error. However, multispectral techniques appear to be capable of resolving these effects at a level adequate to support 0.1 milliarcsec astrometry.

Ground-Based, Narrow-Field Interferometry in the Infrared

Another indirect planet-detection method of considerable potential is the application of long-baseline interferometry to narrow-angle astrometry. This approach has recently generated new attention, in large part because of the newly discovered, relatively benign characteristics of at-

mospheric seeing over very narrow fields, as discussed above. This approach works best at infrared wavelengths, where the atmospheric seeing cell is large, the atmospheric time constant is long, atmospheric dispersion is small, fringe tracking is easier, photon noise is less, and suitable reference stars are easier to find. From the Mauna Kea site, at 2.2 microns, a 1.5-meter aperture is usually diffraction-limited with only tip-tilt correction. Two such apertures can be separated by a baseline on the order of 100 meters and the beams from a target star combined. This section discusses image-plane recombination. The alternative method of pupil-plane interferometry is described in Chapter IV, which also contains a general discussion of the differences between the two techniques.

In the image plane, two beams produce a monochromatic image consisting of an envelope given by the close-to-diffraction-limited image from each aperture (having widths on the order of 0.5 arcsec), and containing a modulated signal with a period set by the baseline; in the case that λ=2.2 μm and with a 100-meter baseline, the fringe spacing is 2.2 x 10^{-8} radians, or about 5 milliarcsec. The achievable astrometric precision depends on the level to which this fringe can reliably be divided in the presence of instrumental and atmospheric effects. Division to 1/50 of the fringe would produce an astrometric precision of 0.1 milliarcsec. This level of fringe division for astrometry has been demonstrated on the Mark III interferometer, which used a 30-meter baseline to measure the separation of the Alpha Andromeda system (about 0.1 arcsec). The 100-microarcsec astrometric accuracy was estimated by comparing a single night's observations with a least-squares orbital fit derived from some two dozen observations taken over a 2-year period. The ability to achieve a 100-microarcsec measurement on this very closely spaced system supports the conclusion discussed earlier that very narrow angle astrometry from the ground is likely to offer significant performance improvements over what has been achieved in the past over wider fields.

Well-defined interference fringes are contained in a narrow band within the combined images. The center of that band corresponds to the point at which the optical path lengths through both telescopes to the star are equal. Because of changes in these path lengths caused by atmospheric fluctuations, after one atmospheric time constant, the phase of these fringes will in general have changed by an amount on the order of 1 radian (i.e., they will have moved on the detector by about 1 milliarcsec). However, this movement may be largely avoided by introducing a

compensating change in the optical path of one of the two telescopes by using a delay line controlled by the fringe shift. The delay line is also necessary to keep the fringe pattern centered in the envelope because the star moves relative to the baseline.

The crucial point, as discussed earlier for the case of single-aperture astrometry, is that reference stars with small angular separations from the associated targets share the same optical path through the atmosphere, and telescopes, and therefore their fringes, are also frozen by the fringe tracking. Many reference stars will be needed to calibrate instrumental characteristics. Such stars are in general much fainter, and their fringes are not visible within one atmospheric time constant. However, because their fringes are frozen, they become visible after a longer exposure. At 2.2 microns, the angular radius over which suitable reference stars can be found is thought to be at least 20 arcsec. Beyond this radius, stellar fringes are washed out after a long exposure because of atmospheric fluctuations. This is because the optical path through each telescope from a reference star at large angular separation from the target is uncorrelated with the path from the target star being used for fringe tracking. When the target and reference stars each have visible fringes, the relative phases of these fringe patterns can be measured. This measurement corresponds directly to a measurement of the angular separation along the direction of the baseline. Note that the angular resolution improves with the baseline, but the size of the region available for reference stars depends only on the individual apertures. Therefore, if suitable beam combining could be achieved, the technique would continue to give better angular resolution as the baseline increases, until the target star fringes become invisible in the time the atmosphere fluctuates (they contain less and less of the light), or one of the stars is resolved.

Significant technical developments will be necessary to achieve the conceptual potential of very narrow-angle astrometry. Simultaneous measurements must be made of at least two stars, the target and one reference star. This requires that the light from the two telescopes be split into two interferometers, with two beam combiners, two delay lines, etc. Laser metrology of existing delay lines has shown accuracies better than 2 nm, which over 100-meter baselines would correspond to 0.5 microarcsec, so this is not expected to pose major technical problems. However, the dual interferometer arrangement needed for very narrow-angle astrometry has never been used. Moreover, effective observing method-

ologies may require more than one reference star, perhaps two or three. Such a requirement could substantially diminish the number of observable fields, and could significantly complicate the instrument.

The Potential Role of Ground-Based Astrometry in TOPS

The TOPS Program can benefit significantly from the continuation of existing high-precision, ground-based astrometric planet-search programs. In addition, recent developments promise significant improvements in ground-based astrometry at a level that could make a major impact on the discovery and initial investigation of other planetary systems. Indeed, it seems certain that real advances in astrometric accuracy can be obtained by capitalizing on the best available ground-based astronomical sites and by using large-aperture telescopes for narrow-angle astrometry in both filled-aperture and interferometric modes.

The potential of these new ground-based astrometric techniques remains to be fully understood and defined. The ultimate achievable accuracy will depend on a number of factors yet to be determined, including the actual behavior of the atmosphere over the usable fields of view, the practical observational methodologies that can successfully be developed for defining reference frames, and the actual distributions of a sufficient number of suitable reference-frame stars in the narrow fields of view, as well as other influences of the atmosphere (such as color dispersion) and instrumental effects.

Astrometric accuracies somewhat higher than 0.1 milliarcsec might be achieved. Should this potential be realized, narrow-angle astrometry will be able to provide an important stepping-stone toward the ultimate goals of the TOPS Program. Current ground-based astrometric programs—with best accuracies on the order of 1 milliarcsec—have the sensitivity to detect Jupiter-like planets around a relatively small number of stars in the solar neighborhood. With accuracies on the order of 0.1 milliarcsec for a large number of targets, the search for Jupiter-like planets could be extended to a large number (100 or more) of nearby stars, though it is not clear how many of these possible targets could be studied within the severe limitations of available observing time. An objective of TOPS-1 is to achieve astrometric accuracy of at least 10 microarcsec for at least 100, and preferably 1000 stars in the solar neighborhood; TOPS-1 will provide a definitive search for planets with masses comparable to those of Uranus and Neptune.

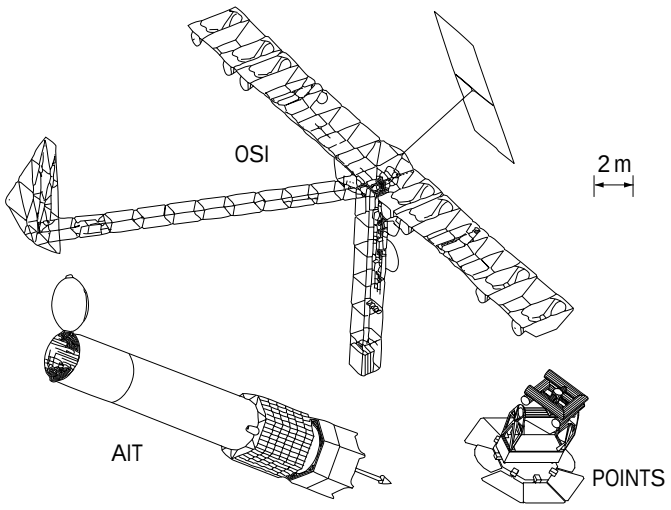


Figure 3-11. The three candidate instruments for implementation during TOPS-1: the Astrometric Imaging Telescope (AIT), the Precision Optical Interferometer in Space (POINTS), and the Orbiting Stellar Interferometer (OSI). The systems are drawn to scale with the reference mark for 2 meters shown in the figure.

The possibility of achieving, from the ground, intermediate accuracies of 0.1 milliarcsec for a statistically significant population of stars would provide a highly desirable transitional step between the presently limited astrometric planet search and the ambitious program designated for TOPS-1. Although significant technical questions remain to be answered, the potential for narrow-angle, ground-based astrometry utilizing the Keck Telescope and associated facilities should be explored and exploited.

PROPOSED TOPS-1 SPACE-BASED PROGRAMS

This report discusses the search for and subsequent intensive study of other planetary systems in terms of several stages of investigation, with each subsequent stage yielding significantly greater conceptual returns and involving increasing technical challenges. Although ground-based investigations—both those currently under way and those contemplated for the future—are expected to make fundamental contributions to the early study of other planetary systems (TOPS-0), the deepest and most definitive investigations will involve space-based instruments of increasing technical sophistication and capability. The TOPS-1 stage of plan-

etary searches will aim to achieve a definitive discovery survey for planets in the solar neighborhood, with masses as low as those of Uranus and Neptune, through an examination of 100 or more stars. Both filled-aperture and interferometric approaches have been proposed to carry out this investigation.

One filled-aperture approach has been proposed for space application to the discovery and study of other planetary systems. The astrometric investigation[18,19] on the Astrometric Imaging Telescope (AIT) is based on principles now being used for ground-based planet detection. Two space-based astrometric interferometers have been proposed: the Orbiting Stellar Interferometer (OSI)[20,21] and the Precision Optical Interferometer in Space (POINTS).[22,23] OSI is based on principles currently being used in ground-based interferometric studies, and POINTS is a space-based astrometric concept. The major conceptual difference between the two interferometer designs is that OSI determines the relative angular separation of two closely spaced (less than 30°) stars, whereas POINTS determines the separation of two stars about 90° apart. OSI uses moving delay lines and siderostats to select a target star from the visible field; POINTS rotates the instrument around the direction to one star and changes the angle between its two rigid interferometers to find the second star. All three candidate mission concepts are sketched in Figure 3-11.

Orbiting Stellar Interferometer (OSI)

The Orbiting Stellar Interferometer is designed to perform both astrometry and limited imaging. It consists of three independent Michelson stellar interferometers arranged collinearly on a single platform. The maximum instrument baseline is 20 meters, with seven 0.4-meter collecting apertures. Observations are to be conducted in the visible and near infrared. Two of the three interferometers would be used to measure the orientation of the spacecraft in order to stabilize the third (science) interferometer. All three interferometers have tilting siderostats and variable delay lines so that the baseline need not be perpendicular to the star direction. For astrometry, the science interferometer switches among target objects within a 30° by 30° field of view; rotating the instrument and changing guide stars brings another patch of the sky into view. Laser metrology is proposed to monitor the relative orientations of the three baselines as well as the internal optical paths to maintain high

astrometric accuracy.

The design goal for OSI is 10 microarcsec for wide-angle astrometry and 1 microarcsec for narrow-angle astrometry (angles <5°). This ambitious goal is limited by photon statistics for sources fainter than magnitude 16 and by systematic errors in the laser metrology for brighter sources. For a magnitude-16 star, the contribution from photon noise in a 10-minute integration would be approximately 2 microarcsec (40 minutes to achieve 1 microarcsec). A 100-picometer error in the three-dimensional metrology of the baseline would contribute 1 microarcsec to the astrometric error. The time-stable part of this error would be partially corrected by calibration.

The most challenging aspect of OSI's metrology design (and of precision metrology systems in general) involves the design and fabrication of the optical fiducials, which define the distances to be monitored or controlled. Laser metrology is used to calibrate the baseline lengths to a nominal precision of 100 picometers (pm), a scale comparable to the size of an atom. Thus, the baseline represents the measured distance between two fiducial points, located essentially at the centers of the siderostats. For OSI, these fiducial points are the apices of corner-cube retroreflectors that are fixed to the surfaces of each siderostat (the first optical surface touched by starlight). It is imperative that the positions of the corner cubes relative to the siderostats be stable, so that measured changes in the distance between the fiducial points accurately reflect changes caused, for example, by thermal-induced distortion of the mirror positions. At present, no corner cubes exist with surfaces that are accurate at the required 10 pm level. Furthermore, no corner cubes exist with three faces perpendicular to each other to an accuracy corresponding to residual errors of 10 pm or smaller. Current technology is about a factor of 1000 worse than is required. However, by mapping the retroreflector error and locking a small laser beam to a spot on its surface, the error can be reduced.

OSI Architecture

The nominal 20-meter size of OSI was chosen to minimize the metrology technology requirements. For a given angular accuracy, the required metrology accuracy scales linearly with baseline length. OSI's 20-meter baseline, which implies a one-dimensional metrology accuracy requirement of 100 pm for 10 microarcsec accuracy, was chosen as a

compromise between something much larger, which would be difficult to put into orbit, and something much smaller, which would impose more demanding metrology requirements.

A feature of OSI is the use of pivoting siderostats to point the instrument at stars. For a large instrument, pivoting the siderostats instead of rotating the entire interferometer allows faster switching among targets by almost an order of magnitude. This faster switching permits faster calibration of the instrument, which in turn permits reduction of errors associated with thermal drifts in the instrument and spacecraft. Faster switching among target stars also improves scientific throughput. With a 20-meter baseline, it should take approximately 1 hour for OSI to pivot by 180°.

Technology Status

Development for OSI is being pursued in three areas. The first is the overall design of the spacecraft and mission, which is coupled to a top-down costing study for the mission. The second is a systems demonstration of a space-based interferometer. As part of the Controls-Structures-Interaction technology program funded by NASA's Office of Aeronautics and Space Technology, a scale model of a space-based interferometer with architecture similar to the OSI concept is being built at the Jet Propulsion Laboratory. The model will have a 15-ft truss structure hung from the ceiling of a high-bay building. The structure will support two independent stellar interferometers (four siderostats) and about 20 individual laser interferometers to monitor vibration. The goal is to demonstrate structural control to the 10 to 30 nm level simultaneously with attitude control at the milliarcsec level.

The third area of development is picometer metrology. A major error source is likely to be the manufacturing errors of the optical fiducials. Metrology errors of 100 picometers in three dimensions are likely to require about 30-picometer accuracy in one-dimensional metrology. Fabrication of corner cubes at this level of accuracy is likely to be impossible. However, the corner cubes need not be perfect if their errors can be calibrated at the 30-picometer level. Preliminary computer simulations of corner cubes made with super-polished optics suggest that the manufacturing errors may be calibrated to a level that permits the needed 30-picometer linear metrology.

Precision Optical Interferometer in Space (POINTS)

POINTS is a wideband optical interferometer dedicated to astrometry. It has been designed to minimize systematic errors and perform as closely as possible to the accuracy limit imposed by photon-counting statistics for stars no brighter than $m_v \leq 10$. POINTS consists of two interferometers oriented at right angles to each other, each with a 2-meter baseline, 25 cm apertures, and a beam-combiner assembly that forms and detects the interference fringes. POINTS is designed for a nominal measurement accuracy of 5 microarcsec, which is to be reached in about 1 minute when observing a pair of magnitude-10 stars. The nominal optical bandpass is from 0.25 to 0.9 microns. After consideration of the times required to slew, settle, and acquire the targets, it is projected that 350 such observations can be made per day. After suspected binaries and giants are excluded, the faintest of 1000 remaining targets of F, G, K, and M stars would be about magnitude 9.5.

Two starlight interferometers observe distinct stars, separated by $90° \pm 3°$ on the sky. The right angle configuration has three attributes: (1) for any given target, there is a large area of the sky (approximately 2000 square degrees) from which to select reference stars; hence, they can be bright; (2) the instrument measures absolute parallax; there is no need for "zero-parallax objects" in the form of faint, distant stars; and (3) instrument biases can be calibrated during the data analysis; wide-angle measurements, rapid all-sky coverage, and 360-degree closure make this possible. Measurement of a star-pair separation is the sum of three parts. Two are the measured separations of the target stars from the corresponding interferometer reference directions; the third is the (approximately 90°) angle between the interferometer reference directions. The former are inferred from the displacements of the beam-combiner assemblies from their nominal positions and the fringe patterns on the detectors. The latter is obtained by laser interferometry.

Systematic error in POINTS is addressed with a combination of three approaches: (1) the use of stable materials and thermal control; (2) real-time metrology using laser gauges; and (3) bias determination based on closure information in the astrometric data. These approaches are addressed below.

Material stability is important on time scales from tens of seconds (single observation) to tens of minutes. The longer times are determined by the ability of the instrument to detect and correct biases via closure.

Even on these longer time scales, creep, radiation damage, and stress relaxation are irrelevant; only vibration and thermal expansion are relevant. The most critical mechanical stability requirement is that for the fiducial blocks, which are solid chunks of glass (e.g., ULE) that are unlikely to show much vibrational distortion. They need to be thermally stable to about 10 millidegrees. Since these components are located deep within the thermally controlled instrument environment, the required short-term stability can be obtained passively by insulation.

POINTS relies on two kinds of laser-driven metrology. Full Aperture Metrology (FAM) surveys the optical components that transfer starlight from the entry apertures to the starlight beamsplitter. FAM starts with laser light injected backwards through the beamsplitter so that all the earlier optical elements that transfer the starlight are illuminated by laser light over their entire starlight-handling surfaces. When the laser light reaches the first starlight-handling element, it encounters a shallow phase-contrast hologram (diffraction grating with curved rulings), which focuses a small fraction of the light into the fiducial block. The laser light is collimated and passed to an auxiliary beamsplitter, where an error signal is generated and used to correct dimensional drift. The laser-light optical path beyond the fiducial block is controlled with an additional laser gauge. The common apices of the clusters of corner-cube reflectors in the fiducial blocks serve as reference points for the corresponding starlight interferometers. The six point-to-point laser gauges connecting these four "fiducial points" measure the angle between the two starlight interferometers.

Star parameters (position, proper motion, and parallax) for the network of observations are modelled by least squares. A key quantity is the redundancy factor M, the ratio of the number of observations in a series to the number of stars observed. When M is about 4.2, all the star-pair separations can be determined (including those that were not, and could not be, measured directly), and the average separation uncertainty is equal to the instrument measurement uncertainty; the grid is said to be "locked up." Covariance studies have shown that an instrument bias model can be included with little increase (less than 10% at M=5) to the star-parameter uncertainties if the number of bias parameters is less than 10% of the number of observations. The information for estimating the parameters of this bias model comes from closure on the sky, but no special observing sequence is required. Thus, the instrument is self-calibrat-

ing on a time scale of an hour: long-term systematic errors will not masquerade as a planetary signature, parallax, or proper motion.

In a POINTS planet-search mission, a sequence of measurements would be made of the separations of pairs of stars, perhaps four times per year for 10 years. The characteristics of such a mission were studied in a series of double-blind mission simulations. In each simulation, one person would enter ranges of parameters into the Monte Carlo program, which would fix the parameters and generate the pseudo data without disclosing the description of the set of planets simulated. Then a second person, without knowledge of the ranges of parameters used by the first, would analyze the data to determine the locations and orbital elements of the simulated planets. (As expected, the search for planetary signatures was found to be interactive, with more interactions required when there were many planets than when there were few.) In the simulations, the number of stars having planetary signatures ranged from 20 to 90; each star had either zero or one planet. There were no false alarms. The probability of detecting a planet was found to range from 1 for an astrometric signature above 25 microarcsec to 0.6 for a signature of 5 microarcsec; below 5 microarcsec, the probability was low.

Technology Status

The key technology area for POINTS is the metrology. Within this area are three critical technologies: fiducial blocks, zone-plate mirrors, and laser gauges per se. Each fiducial block is a collection of retroreflectors that forms the end points of four laser gauges. The block must be mechanically stable, and the apices of the retroreflectors must be coincident to within a few microns, a goal that appears achievable.

The zone-plate mirror (ZPM) is a holographic optical element that makes the full-aperture metrology possible. Test ZPMs have been used to show that the basic properties of the element could be realized, but they lack the uniformity of diffractive efficiency required for some of the desired tests. The technology to make a ZPM that meets flight requirements has been demonstrated by industry for other (Strategic Defense Initiative) applications. A measure of laser-gauge performance has been obtained in tests of a dual gauge in which two laser signals, derived from the same laser, measure the same distance over the same optical path through air, using different and distinguishable modulation schemes. One laser signal was used to control the distance to a constant in units of

the laser wavelength. The second was used to independently measure the controlled distance. This use of a controlled distance is essential to ensure that the measurement reflects performance of the laser gauge, and not temperature-driven fluctuations in the length of the metering rods and air-pressure fluctuations in the test interferometer. With the optical path adequately controlled by the first interferometer, the behavior of the second interferometer mimics its expected performance in the space-based instrument. The two-point (Allan) variance of the measurements of the second interferometer met the preliminary flight requirement of 2 pm on time scales of 1 minute to 1 hour. This requirement for each laser gauge is a conservative goal intended to ensure that the combined laser gauges contribute an appropriately small amount (less than 10 pm) to the total error budget of 50 pm, consistent with a 5-microarcsec accuracy goal for the 2-meter-baseline interferometers.

Astrometric Imaging Telescope (AIT)

The AIT encompasses two instruments, combining the indirect planet-detection method (astrometry) and direct detection (coronagraphic imaging) into a single satellite package that shares the main optical system. (The AIT imaging investigation is discussed in Chapter IV.) The AIT astrometry instrument is the Astrometric Photometer (AP), a space-borne adaptation of the MAP instrument currently used in a ground-based planet search. In the AP, each target field consists of the target star and as many as 25 reference stars. The AP instrument itself consists of a Ronchi ruling and a detector assembly. The ruling is moved across the field of target and reference stars. Behind the ruling, several probes are positioned, one at each star, to act as a photometric light bucket. Lenses and fiber optics on each probe scramble the light to minimize introduction of metric error, and to direct the starlight into photon-counting detectors. The light curve from each star is recorded and transmitted to Earth where phase analysis gives the position of the target star with 10-microarcsec accuracy for each approximately 30 minutes of data collection, depending on the brightness of the target and reference-frame stars. The measurement is repeated in the orthogonal direction. The AP study is designed to investigate at least 100 selected stars in the solar neighborhood with sensitivities sufficient to discover and characterize planetary systems.

The AP metric is invested in a single, passive, multi-lined, glass

Ronchi ruling applied to all the measurements. Stellar positions are determined from numerous individual relative positional measurements, as the ruling—with as many as 10^3 to 10^4 lines—is drawn across the field of view. This results in averaging down the random ruling error by between one and a half and two orders of magnitude, relieving the burden on fabrication tolerance. Because each relative positional measurement is made in a single frame, and because all are made against the same passive ruling metric, the tolerances of the system are highly relaxed in comparison with the precision of the final measurement. And, because all data—including all the target stars and reference frames, over the entire duration of the mission—are measured against the single passive metric, systematic errors can be calibrated internally against the whole data set.

Problems and Limitations

Photon Noise. The AIT astrometric instrument directly measures the centroids of stellar images, and is designed to be limited in astrometric accuracy by photon noise at the 10-microarcsec level. Achieving the measurement accuracy requires collecting sufficiently many photons to divide the image according to $\alpha \sim \theta/\sqrt{N}$, where α is the precision of the astrometric measurement, θ is the size of the stellar image in the AP focal plane, and N is the number of photons in the observation. For the AIT design, θ is about 0.1 arcsec. Thus, about 10^8 photons are required to achieve 10 microarcsec accuracy. Roughly speaking, this number of photons must be collected from each target star alone, as well as from the aggregate of each frame's reference stars. For the baseline AIT, the limiting target-star magnitude is about 11 for an observation of about 30 minutes; however, in the typical case, photon flux from the reference frame limits the speed at which each measurement can be made. A representative number of candidate target stars and their reference fields have been examined to verify that a sufficient number of reference-frame photons are available to conduct the planned program.

Optical Aberrations. The AIT optical system is designed to minimize asymmetric optical aberrations over the system's usable field. This greatly reduces systematic metric errors that otherwise arise in astrometry. Numerical modelling of the optical system has shown that the residual centroid shifts remain below the required accuracy as long as (1) the primary-secondary separation is held to about 100 μm; (2) the primary-secondary decenter is held within a tolerance of 2 mm; (3) the

secondary tilt is held within a tolerance of 2 arcmin; and (4) inhomogeneous variation of mirror reflectance remains within about 90%. (In-orbit optical alignment is facilitated by use of the AIT coronagraphic imaging instrument, which provides a convenient, on-board alignment diagnostic.)

Pointing Jitter. Spacecraft pointing jitter induces centroid positional noise. Pointing jitter should be held within about 0.01 arcsec, in the frequency band 10 to 100 Hz. At other frequencies, it is sufficient to limit the pointing jitter to below 2 arcsec to keep the stars within the fields-of-view of the photometry probes behind the ruling.

Ronchi Ruling Inaccuracies. The Ronchi ruling holds the basic metric for AIT astrometry; errors in the ruling lines translate to errors in stellar centroids. For random line-edge errors, the accuracy of the ruling equals that of one line divided by the square root of the number of lines. For 10 microarcsec accuracy, the ruling-line error must be less than 1 nanometer; thus for a 6400 line ruling, the random error in one line must be less than 80 nm.

Systematic errors in ruling lines do not average down; they need to be controlled by a combination of calibration and manufacturing tolerance. At low spatial frequencies, systematic errors can be calibrated. At high frequencies, systematic errors are averaged by, for example, the finite size of the stellar image. At intermediate frequencies (100-1000 cycles), systematic errors must be held below 1 nm. Inhomogeneous thermal properties of the ruling substrate can introduce errors, requiring careful selection of substrate material and thermal control of the ruling environment.

Technology Status

The AIT has three major technology-development issues. Two of these (supersmooth mirrors and coronagraphic mask fabrication) pertain to the imaging investigation and are discussed in Chapter IV. For the astrometry, the main technology challenge is the Ronchi ruling.

Test rulings have been fabricated using both electron beam lithography and optical photo-reduction. Other techniques, including optical interferometric ruling manufacture, as well as mechanical fabrication, are also still being evaluated. These initial tests indicate that, in principle, the needed rulings could be fabricated to the required specification. But efforts are under way to demonstrate the feasibility of fabricating rulings

of the needed size for which the combination of fabrication tolerance and calibration can meet the overall specifications for both random and systematic errors.

Other technology areas of importance include fiber-optic positioners, photon-counting detectors, and long-lifetime mechanisms. These have ground-based, and in some cases space-based, analogs and are of lesser concern than achieving the basic metric tolerances.

PROPOSED TOPS-2 INDIRECT DETECTION FROM THE MOON

The Moon is an attractive location for a variety of astronomical instruments. Indirect detection of planets could also benefit by operating from the lunar surface. For example, an optical interferometer with a 2-kilometer baseline would have a fringe separation of 50 microarcsec. Centroiding the fringe to one part in 500 would result in 0.1 microarcsec of astrometric accuracy. With moderate to large apertures, e.g., about 1.5 meters, such an instrument would gather sufficient photons to rapidly make astrometric measurements of even distant stars. Astrometric accuracy of 0.1 microarcsec compares favorably with the 0.3 microarcsec amplitude astrometric wobble induced by the motion of an Earth-like planet around a solar-type star at 10 parsecs. A lunar interferometer with these characteristics would resolve the disk of the star, allowing also for the possibility of measurements that could overcome positional ambiguities introduced by intrinsic stellar phenomena.

3.3 RADIAL-VELOCITY MEASUREMENTS

Recall from Figure 3-4 that a Doppler measurement detection threshold of 5 m s^{-1} or better would be able to detect a Jupiter-sized planet around virtually every star in the indicated search space. Existing techniques already have achieved this accuracy for bright stars. Extension to the dimmer stars can be accomplished by larger telescopes, higher photon sensitivity, and/or longer integration times.

Ultimately, the use of radial-velocity measurements for discovering planets will be limited by systematic errors inherent in the target objects. For example, the absorption lines of the Sun are known to be systematically blue-shifted as a result of granular convection. Variations in this effect over the 11-year solar cycle[24] could, in principle, produce an apparent radial-velocity variation that mimics the effect of Jupiter. Although early observations of the Sun[25,26] suggest such relatively large intrinsic ra-

dial-velocity effects, more recent studies indicate that such effects are not large enough to compromise the search for Jupiter-class planets around solar-type stars.[27,28] Moreover, apparent radial-velocity variations caused by intrinsic behavior of a star are expected to be accompanied by related effects observable in spectral line shape.[29] These spectral effects provide a way to correct for extraneous, intrinsic stellar behavior. It is not yet known whether the instrumental systematic accuracy of radial-velocity measurements can be made lower than 1 m s^{-1}. But even if the 10 cm s^{-1} precision level needed to detect an Earth-like planet in a 1-AU orbit about a solar-type star were achievable, intrinsic stellar sources of variation in apparent radial velocity would probably preclude unambiguous detection and interpretation of such a low-amplitude signal.

STATE OF THE ART

Two techniques have demonstrated the capability for measuring stellar radial-velocity variation to precisions on the order of 5 m s^{1}. Both techniques involve accurate, high-dispersion spectral analysis. However, they differ significantly in their actual technical approaches and in the methods used to establish accurate reference spectra against which these very precise measurements can be made.

Fabry-Perot Technique

In the Fabry-Perot etalon approach,[30-36] light from the Newtonian focus image of a star is transmitted to the entrance aperture of a spectrometer via an optical fiber. The fiber isolates the instrumentation from the telescope, and the many internal reflections thoroughly scramble the light so that the output end of the fiber is uniformly illuminated. Light from the fiber end is collimated and passed through a tilt-tunable Fabry-Perot etalon made of ULE glass plates, and operated in a temperature-controlled vacuum chamber. The etalon is a multiband pass filter that transmits only narrow ranges of wavelength corresponding to the many orders of constructive interference, where adjacent regions of transmitted wavelengths are separated by wider spectral-free regions of almost no transmission. The spectrum is thereby sampled at discrete points by the etalon. The spectrum is further dispersed and cross dispersed to separate overlapping orders, and the resulting two-dimensional array is focused onto a charge coupled device (CCD) array. Several hundred points of the spectrum are sampled in a spectral range of about 500 Å near

4300 Å. Near this wavelength, each interference order is 47 mÅ wide, and the sample points are 0.64 Å apart, resulting in distinct, widely separated, monochromatic images of the entrance aperture at the focal plane of the camera. Changes in Doppler shift cause changes in the relative intensities of these images according to the slope of the spectral profile at each point sampled.

If the etalon is tilted about an axis perpendicular to the collimated beam, the transmission windows are shifted in wavelength. With a local reference spectrum, the etalon can be calibrated to two parts in 10^3 corresponding to a Doppler shift of ± 6 m s^{-1}. Thus, it is possible to position the etalon to compensate for the motion of the spectrograph along the line of sight to the target star. Any change in the Doppler shift of the star relative to the solar system's barycenter then shows changes in the relative intensities of the aperture images on the CCD. In the region of the spectrum near 4300 Å, about 10 to 25% (depending on spectral type) of the sample points are on the steep slopes of absorption lines with the remaining sample points being in the continuum with nearly zero slope. The continuum points are used to normalize the intensity from one observation to the next. The difference in the intensities of each point on a steep slope yields a change in the Doppler shift between observations since the sample point has not changed wavelength relative to the solar system barycenter, whereas the stellar spectrum may have shifted. The change in Doppler shift determined at a particular sample point is given by

$$\Delta v = \frac{c\Delta I}{\lambda \frac{dI}{d\lambda}}, \tag{3.8}$$

where I is the intensity, λ is the wavelength at that point, v is the radial velocity of the star, and c is the velocity of light. The determination is weighted by the local slope, and all the determinations are averaged to yield the change in the Doppler shift between the two observations.

Final determination of the residual systematic errors is made by observing sources of known Doppler shift, such as using sunlight reflected from a lunar crater. Random error in the lunar observations (as determined from the standard deviation of a rapid sequence of observations) has a value near ± 5 m s^{-1}. The systematic error is estimated by the quadratic difference between the standard deviation over the entire 4-year

sequence of observations of ±7 m s^{-1} and the random error of ±5 m s^{-1}, yielding ±5 m s^{-1} also for the systematic error. The systematic errors most likely result from uncalibrated changes in system hardware, which generally are probably not Gaussian.

Initial observations using one Fabry-Perot velocity meter began on Kitt Peak in September 1985. The instrumentation was finalized in December 1986, at which point a long-term series of uniformly calibrated observations was begun, targeting 16 stars of approximately solar type. Individual observations are good from ±3 m s^{-1} to ±30 m s^{-1}, depending on the brightness of the star and other observing circumstances. In addition to the lunar studies, 91 observations of the K0 V star σ Draconis, over 3.6 years, show the standard deviation of the four seasonal averages to be ±3 m s^{-1}, indicating both the constancy of the star and stability of the metric. The ultimate error in a determination of a radial-velocity change is always less than the error of a single measurement. A decrease in random error is expected as $1/\sqrt{n}$ where n is the number of observations, although systematic errors frequently prevent this ideal from being realized.

Gas Absorption Cell Technique

A gas absorption cell placed in front of the entrance slit of a coudé spectrograph can provide a reference spectrum against which to measure the shift of a stellar spectrum.[37] Absorption lines of the star will have superimposed on them the absorption lines of the gas, providing an accurate wavelength calibration. Iodine (I_2)—with a strong electronic band ($B\,^3\Pi^+_{Ou}$ - $X\,^1\Sigma^+_g$) in the 5000 to 6000 Å spectral region—produces a rich spectrum of extremely narrow lines, and this gas has been successfully used in an absorption cell.[38] The combined spectrum from the star and I_2 absorption falls onto an 800 x 800 CCD array in the coudé spectrograph. Spectral resolution is 0.030 Å, so that each resolution element is slightly oversampled, and there are many resolution elements over a typical spectral line. Spectra are normalized from observation to observation using the continuum.

To monitor a star for varying Doppler shift, a reference spectrum of the star, without the superposed I_2 absorption lines, and a separate spectrum of the I_2 cell, illuminated through the telescope optical path by a quartz lamp, are recorded. To determine the change in the Doppler shift, the stellar spectrum is compared, in an iterative fitting cycle, with a

model spectrum calculated by combining an appropriately shifted reference stellar spectrum and the fiducial I_2 spectrum. The same procedure is applied to the fiducial iodine spectrum taken at the same time as the reference stellar spectrum to find any instrumental drift. The effect of the solar system motion of the spectrograph, as determined from a planetary and lunar ephemeris, is subtracted to yield the change in the Doppler shift of the star relative to the solar system barycenter.

Systematic errors must be controlled or calibrated to a level significantly below the random errors. Typical sources of systematic errors include such things as temperature variations in the gas absorption cell or spectrograph, variations in illumination of the spectrograph between stellar and flat field observations, and long-term changes in filter characteristics. Systematic errors are identified and isolated by repeated observations of a source with Doppler shifts thought to be well understood. One kind of test is made by observing reflected sunlight from selected areas on the lunar surface. Even the Moon is not an ideal standard calibration object. Because the Moon is an extended object, it illuminates the optics slightly differently than a star. From the lunar observations, random errors (determined from repeated lunar spectra over 3-day observing runs) are found to be ± 4 m s^{-1}. The systematic errors have not yet been quantified because the reduction procedure does not yet include lunar rotation and libration, but the systematic errors are expected to be less than ± 10 m s^{-1}, and probably about half that value.

Random errors depend largely on photon-counting statistics. Thus, the random errors of a single observation should vary inversely as the square root of the number of photons gathered. Repeated observations in a time series of radial-velocity measurements of an object can average the random errors and permit determination of the projected orbital velocity amplitude to better than the error of a single observation. Random error sources can include the limited spectral bandpass, photon statistics, CCD readout noise, determination of mid-exposure time, and short-term small variations in cell or spectrograph temperature.

An I_2 cell was installed on the coudé spectrograph of the McDonald Observatory 2.7-meter telescope in the fall of 1990, replacing an O_2 absorption line instrument. Stability tests of the I_2 cell have shown random errors to be in the 2 to 5 m s^{-1} range.[38] Tests of the I_2 cell on a much less stable coudé spectrograph and a 2.1-meter telescope yielded precisions of 10 to 15 m s^{-1}.

Other high-precision radial-velocity programs include a Canadian program using a hydrogen-fluoride (HF) absorption cell at the coudé focus of the Canada-France-Hawaii Telescope. This program already has acquired data over a period of more than 8 years. The published data set[39] has an error of ±13 m s^{-1}, and includes 23 bright solar-type stars, each observed an average of five times per year. The spectral range covers 8640 to 8780 Å, making it most sensitive to late type stars. Another program at San Francisco State University[40] uses an I_2 absorption cell in front of the coudé spectrograph on the Lick 3-meter telescope. The bandpass is 3800 to 9500 Å with a resolving power of 40,000. The Lick system has been operational since May 1990 with a velocity scatter of ±25 m s^{-1} over a 1.2 year period. Observations encompassing 70 main-sequence F-M stars are planned.

Cross Correlation Techniques

Classical radial-velocity systems making measurements with accuracies in the range 25 to 50 m s^{-1} can play important supporting roles in a planet search program.

The CORAVEL (COrrelation RAdial VELocities) instruments[41] are echelle spectrographs with a two-dimensional metal mask at the focal plane and a refraction block with precision rotation for shifting the spectrum along the dispersion axis (Doppler shift axis) of the mask. The mask has slots corresponding to the absorption lines of the spectrum being observed. The signal passed by the mask is measured photoelectrically, and the shift where the signal reaches a minimum gives the Doppler shift of the star relative to the mask position, since that is the shift where the slots in the mask match up with the absorption lines in the spectrum. The shift in the reference spectrum relative to the mask is also determined to calibrate instrumental drifts. The Doppler shift of the star at a particular observation is the difference between the shift in the reference spectrum and that of the star.

Three instruments that record digital data[42] are in operation. The recorded spectra are mapped to a laboratory-standard wavelength scale using a laboratory-reference spectrum, and then digitally shifted relative to a template spectrum until the cross-correlation between the spectrum and template is maximal. The shift determined in this manner is the Doppler shift relative to the template spectrum. Long-term stability is a critical requirement for any attempt to measure low-amplitude spectro-

scopic orbits. The zero point stability has been achieved at the level of about 100 m s^{-1} over more than 10 years. Stars are observed routinely to magnitude V=15, and some projects have reached as faint as V=18. Improvements in the instrument configuration should provide precisions of 25 m s^{-1}.

Altogether "classical" radial-velocity methods seem capable of achieving precisions of 25 to 50 m s^{-1} on faint stars using moderate-sized telescopes. This can support extensive surveys of moderately low-amplitude velocity variations in a large number of stars, supplementing the higher precision radial-velocity searches for planetary systems. Thus, stars that are members of previously unknown binary pairs or that have brown dwarf companions producing radial-velocity variations in the range 25 to 50 m s^{-1} could be eliminated from search programs using higher-precision instruments. Establishing which stars have no detectable radial velocities down to 25 m s^{-1} would allow the efficient selection of those stars for the more precise surveys. A moderate-sized telescope in the 1.5-meter diameter range should be able to reach more than 1000 stars at a limit of 25 m s^{-1}, whereas the largest telescopes available for ra-

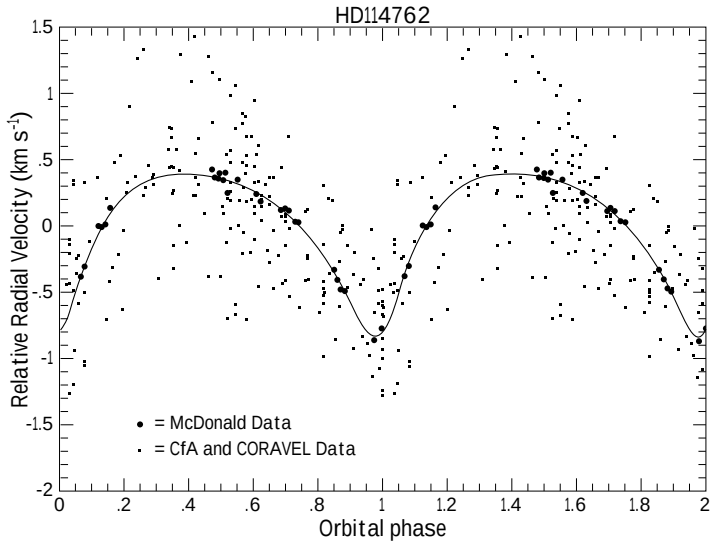


Figure 3-12. Radial velocity curve for HD114762. The McDonald Observatory planetary search data of Cochran et al (1991)[44] are shown as the large filled circles. The original discovery data of Latham et al (1989)[43] are shown as small dots. The solid curve is from the orbital solution derived from the McDonald data.

dial-velocity work are not observing even 100 stars at 5 m s^1. Sufficient observing time is available on moderate-sized telescopes to allow the extensive monitoring of large numbers of stars. The spectroscopic orbit with the smallest confirmed amplitude is the old solar-type dwarf HD114762;[43] see Figure 3-12.[44]

Role of Radial-Velocity Measurements in TOPS

Radial-velocity measurements have reached the sensitivity necessary to detect planets ($\pm$5 m s^{-1}), and current observing programs are well into the discovery phase. Existing high-precision observing programs typically survey two to three dozen stars each. Significant overlap in target choices allows independent confirmation of results. Future radial-velocity measurements in TOPS should embrace three activities: continuation of existing long-term productive programs, development and application of new instruments, and (as data series become longer and more accurate) a gradual evolution of radial-velocity work from detection to characterization of planetary systems. This third activity will continue to complement other measurement disciplines in subsequent phases of TOPS. Radial-velocity planet detection programs are photon-intensive, and different surveys have solved this problem either through a dedicated telescope of moderate aperture or through regular access to a larger aperture telescope. An effective TOPS Program would take advantage of the economy and complementary data provided by intensive radial-velocity investigations by supporting a vigorous program of radial-velocity surveys on a variety of telescopes, including very large aperture systems.

3.4 Photometry

Photometry measures the brightness of stars. Two phenomena exist for which photometry could be used to discover some planetary systems. The simplest occurs when a planet orbiting a star passes between the star and a distant observer; in this case, the apparent brightness of the star is temporarily reduced by the amount of light blocked by the planet.[45,46] A second photometric approach takes advantage of the gravitational focusing that occurs when one star passes between a distant star and an observer.[47] In this second case, the gravity of the intervening star bends and focuses the light, acting like a (gravitational) lens. The primary effect of the lensing is to produce a temporary—and temporally symmetric—brightening of the starlight. However, if the intervening star has a large

planet orbiting it, the resulting gravitational lens is not cylindrically symmetric, and the distortion (caused by the planet's gravity) can induce significant and potentially observable features in the light curve of the event.

Both these effects depend on fortuitous alignments, which are unlikely to occur for any particular objects. Thus, to be efficacious, such programs would have to operate in broad survey modes, monitoring very large numbers of stars, each with a very low probability of showing any effect. Even for a star known to have one or more planets, photometric monitoring will reveal a planet only in the unlikely event that the observer happens to lie almost precisely in the plane defined by the planet's orbit. Similarly, the precise alignment of any two stars, seen by any observer, is a rare event. Until recently, such an undertaking would have been beyond the reasonable reach of existing technology. Now, with large-area CCD detectors, photometric approaches, while remaining challenging and difficult, are no longer out of reach. Photometric accuracies of 0.05 magnitudes would be sufficient for at least some systems (to detect either gravitational lensing or a Jupiter-class planet passing in front of an M-dwarf star). Such photometric accuracies are realizable. However, carrying out long-term, routine, automated measurements of many stars at such accuracies is a technological challenge, as is the required robust and automatic software capable of handling the very large data sets that would be involved. Either technique would require making hundreds or thousands of photometric measurements on each of up to a million stars. Although this is an ambitious requirement in terms of data, it only exceeds by a factor of 10 to 100 the sizes of astronomical measurement programs now carried out or contemplated, something that is not unthinkable in the context of rapidly increasing technical capacity.

Gravitational Microlensing

When a point mass is closely aligned intermediately between an observer and a distant point source of light, the observer (equipped with an infinite-resolution telescope) would see two images of the source. This phenomenon is called gravitational lensing. For separations of several kiloparsecs, and a solar-mass lensing star, the lensing-event cross-section is about one milliarcsec. Because this is beyond the resolution of existing telescopes, the two images would be seen as a single, combined image. If the distant star, the intervening star, and the observer were in a fixed

configuration, the lensing would alter the apparent brightness of the distant star, but would otherwise be undetectable. However, stars separated by several kiloparsecs in galactic radius have relative transverse velocities of about 200 kilometers per second due to differential galactic rotation. For these, any gravitational focusing event is only of limited duration, typically lasting about a month for a solar-mass lens (and proportional to the square root of the lens mass). The amplitude of the light-focused, temporary brightness enhancement depends on the impact parameter of the event: the smaller the impact parameter, the larger the amplitude.

This search method would be optimized by observing dense star fields through dense star fields; observing galactic-bulge stars through the galactic disk provides an optimum approach. The intervening (lensing) objects would include ordinary stars, brown dwarfs, and planets. The probability that any given star in the bulge will be seen to brighten by as much as 0.3 magnitude in any month is less than 10^{-6}. For a single-star lens, the expected light curve is bell-shaped, with only one dimensionless parameter. Most stars in the galactic disk are double or multiple. Because of this, the simple light curve will be very strongly modified in about 10% of lensing events.[47] If every star had a Jupiter-class planet, the lensing light curves would be modified very strongly in about 5% of the events if the planets had random semimajor axes. (However, if Jupiters form preferentially at the water condensation distance from a star, the similarity of this distance to the Einstein ring radius for a star halfway to the galactic center can increase the probability of a light curve modification to nearly 20%).[48] The planetary system would appear as a binary with a high mass ratio. In principle, the same approach could detect isolated brown dwarfs and isolated planets in the galactic disk, with masses as low as Mercury.[49]

A principal shortcoming of this approach is that lensing events, though not themselves difficult to measure, are one-time occurrences. Moreover, because the discovered systems will inevitably be very distant, there is no possibility of verification or follow-up studies using other techniques. However, if feasible, the microlensing approach could be useful for gross statistical studies. An advantage is that this approach might be capable of providing a relatively unbiased survey of systems with component mass ratios ranging from double stars through brown dwarfs to planetary mass companions. It is also the only method that might detect isolated, planetary-mass objects.

Transits

The decrease of a star's apparent brightness caused when a planet transits the stellar disk measures the planet's size. Such an event can be observed only if the planet's orbit plane is nearly perpendicular to the plane of the sky, a low-probability occurrence. The probability that an actual transit will occur is highest for short-period planets in tight, several-year orbits. The typical transit event would be several hours in length, requiring that the frequency of observation of any star be significantly higher than one per several hours, over a period of several years. The required photometric precision exceeds eight parts in 10^5 if the close planets are Earth-like, and the star is similar to the Sun. A planetary transit could in principle be distinguished from other intensity fluctuations by analyzing characteristic variations in color resulting from differential stellar limb darkening as the planet traverses the stellar disk.[45] The color photometry must be done to an order of magnitude better precision than that for the starlight reduction, and would require that observations be made relatively frequently. The same approach might also provide an estimate of the chord traced by the planet across the star's disk, and thereby indicate the time to the next transit. An observing program of this nature would need to be done from space.

The majority of stars in the galactic disks are M dwarfs, much fainter than the Sun. Jupiter- or Uranus-like planets might conceivably form within 0.5 AU of an M dwarf. In such a system, photometric discovery of the planet would be facilitated.[47] With the short orbit period, transits would occur relatively frequently. Also, because M dwarf stars are only slightly larger than Jupiter, the relative brightness decrease would be very large. A study of M dwarfs could be done with a large-area CCD detector, so that a number of M dwarfs, perhaps as many as 100 within 1 square degree and within 100 pc of the Sun, might be observed simultaneously. Searching for eclipses of a few thousand M dwarfs should, within a few years, lead to the discovery of a few eclipsing systems, or at least put meaningful upper limits on their existence. Once a planetary system is found, it can be monitored for a long time to search for other planets. For M dwarfs within 100 pc of the Sun, follow-up astrometric and radial-velocity observations would be feasible, although very large aperture telescopes would be required for studies on the distant and very faint M dwarfs.

REFERENCES

1. Wolszczan, A. and Frail, D. A. 1992 *Nature*, 355, 145.
2. Gliese, W. 1969, *Heidelberg Astronomisches Recheninstitut, Veroeffentlichungen Nr. 22*, Ed. G. Braun, Karlsruhe, Verlag.
3. Gliese, W., and Jahreiss, H. 1979, *Astron. Astrophys.*, 38, 423.
4. Black, D. C. and Scargle, J. D., 1982, *Astrophys J.*, 263, 854.
5. Smith, P. H., McMillan, R. S. and Merline, W. J. 1987 *Astrophys. J. Let.*, 317, L79.
6. Cochran, W. D. 1988, *Astrophys. J.*, 334, 349.
7. Walker, G. A. H., Yang, S., Campbell, B., and Irwin, A. W. 1989, *Astrophys. J. Let.*, 343, L21.
8. Doyle, L. R. 1988 in *Bioastronomy: The Next Steps*, Ed. G. Marx, Kluwer Academic Publishers, p 101.
9. Roddier, F. 1981, *Progress in Optics*, 19, 280.
10. Lindegren, L. 1980, *Astron. Astrophys.*, 89, 41.
11. Woolf, N.J., 1982, *Ann. Rev. Astron. Astrophs.*, 20, 367.
12. Reasenberg, R. D., and Shapiro, I. I. 1986 in *Proceedings of IAU Symposium 114, Relativity in Celestial Mechanics and Astrometry*, Eds. J. Kovalevsyy and V. A. Brumberg, Reidel, Dordrecht, p 383.
13. Shao, M. et al (1988), *Astron. Astrophys.*, 193, 357.
14. Gatewood, G. D. 1987, *Astron. J.*, 94, 213.
15. Gatewood, G., Kiewiet de Jonge, J., Stein, J., Han, and Breakiron, L. A. 1986, *Astrophysics of Brown Dwarfs*, Ed. M. C. Kafatos, R. S. Harrington, and S. P. Martin, Cambridge University Press, London. p 104.
16. Han, I. 1989, *Astron. J.*, 97, 607.
17. Korsch, D. 1989, *SPIE Proc.* 1113, #7.
18. Levy, E. H. et al. 1986, *SPIE*, 628, 181.
19. Lawrence, G. N., Huang, C., Levy, E. H., and McMillann R. S. 1991, *Optical Engineering*, 30, 598.
20. Gershman, R., Rayman, M., Shao, M. 1991 in *Proceedings of 42nd Congress of the International Astronautical Federation, Oct 1991, Montreal*, p 421.
21. Shao, M., Kulkarni, S., Jones, D., eds, *1991 Workshop Proceedings Science Objectives and Architectures for Optical Interferometry in Space*, JPL Document D-8540.
22. Reasenberg, R. D. et al. 1988, *Astron. J.*, 96, 1731.

23. Schumaker, B. L., Eldred, D., Ionasescu, R., Melody, J., Miyake, R., Satter, C., Sonnabend, D., Ulvestad, J., Wang, G., (1991), *The POINTS Instrument for TOPS: 1991 Progress Report*, Jet Propulsion Laboratory Internal Report.

24. Dravins, D. 1985, in *Proc. I.A.U. Colloq. 88: Stellar Radial Velocities*, Ed. A. G. Davis Philip and D. W. Latham, Schenectady, L. Davis Press p 311.

25. Deming, D., Espenak, F., Jennings, D., Brault, J. and Wagner, J. 1987, *Astrophys. J.*, 316, 771.

26. Jimenez, A., Pallé, P. A., Régulo, C., Roca Cortés, T., Isaak, G. R., McLeod, C. P. and van der Raay, H. B. 1986, *Adv. Space Res.*, 6, 89.

27. Wallace, L., Huang, Y. R. and Livingston, W. 1988, *Astrophys. J.*, 327, 399.

28. McMillan, to be published 1992.

29. Gray, D. F. 1988, *Lectures on Spectral-Line Analysis: F, G, and K Stars*, The Publisher, Arva, Ontario.

30. Serkowski, K., Frecker, J. E., Heacox, W. D., KeKnight, C. E. and Roland, E. H. 1979, *Astrophys. J.*, 228, 630.

31. Serkowski, K. Frecker, J. E., Heacox, W. D. and Roland, E. H. 1979, in *Instrumentation in Astronomy III: Proc. SPIE, Vol. 172*, Ed. D. L. Crawford, p 130.

32. Serkowski, K., Frecker, J. E., Heacox, W. D. and Roland, E. H. 1980, *An Assessment of Ground-Based Techniques for Detecting Other Planetary Systems: NASA Conference Publication No. 2124* Ed. D. C. Black and W. E. Brunk, p 175.

33. McMillan, R. S., Smith, P. H., Frecker, J. E., Merline, W. J. and Perry, M. L. 1985, *I. A. U. Colloquium No. 88, Stellar Radial Velocities*, A. G. D. Phillip and D. W. Latham, Eds. p 63.

34. McMillan, R. S., Smith, P. H., Frecker, J. E., Merline, W. J. and Perry, M. L. 1986, in *Instrumentation in Astronomy VI, Proc. SPIE*, 627 p 2.

35. McMillan, R. S., Perry, M. L., Smith, P. H. and Merline, W. J. 1988, In *A. S. P. Conf. Ser. 3, Fiber Optics in Astronomy*, Ed. S. C. Barden, Provo, Brigham Young Univ. Press, p 237.

36. McMillan, R. S., Smith, P. H., Perry, M. L., Moore, T. L., and Merline, W. J. 1990, in *Instrumentation in Astronomy VII, (Proc. SPIE, Vol. 1235)* p 601.

37. Campbell, B. and Walker, G. A. H. 1979, *Pub. A. S. P.*, 91, 540.

38. Cochran, W. D. and Hatzes, A. P. 1990, *Optical Spectroscopic Instrumentation and Techniques for the 1990s: Applications in Astronomy, Chemistry and Physics (Proc. SPIE, Vol. 1318)*, Ed. B. J. McNamara and J. M. Lerner, p 148.

39. Campbell, B., Walker, A. H. and Yang, S. 1988, *Astrophys. J.*, 331, 902.

40. Marcy, G. W. and Butler, R. P. 1992, *P.A.S.P.*, in press.

41. Mayor, M., 1985, in *Stellar Radial Velocities, IAU Coll. 88*, ed. A.G. Davis Phillip and David W. Latham (Schenectady: L. Davis Press), p 35.

42. Latham, D. W., 1985, in *Stellar Radial Velocities, IAU Coll. 88*, Ed. A.G. Davis Phillip and David W. Latham (Schenectady: L. Davis Press), p 21.

43. Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M., and Burki, G. 1989, *Nature*, 339, 38.

44. Cochran, W. D., Hatzes, A. P. and Hancock, T. J. 1991, *Astrophys. J. Let.*, 380, L35.

45. Borucki, W. J., and Summers, A. L. 1984, *Icarus*, 58, 121.

46. Tutukov, A.V. 1987, *Soviet Astron.*, 31, 663.

47. Mao, S., and Paczynski, B. 1991, *Astrophys. J. Let.*, 374, L37.

48. Gould, A., and Loeb, A. 1992, Submitted to *Astrophys. J.*

49. Paczyñski, B. 1991, *Astrophys. J.*, 371, L63.

IV. Direct Detection

4.1 INTRODUCTION

THE ULTIMATE CHALLENGE OF ANY SEARCH for planetary systems is to detect and study Earth-like planets orbiting other stars. Earth supports the remarkable phenomenon of life, and our uniqueness is a profound unsolved question. Chapter III described plans for detecting planets by indirect means to clarify how often planetary systems occur. Giant planets can be detected this way, but the small stellar motion induced by an Earth-like planet is not likely to be measurable. The characterization of Earth-like planets requires direct observations of radiation from the planet itself. This chapter describes the TOPS direct imaging program to determine the prevalence of other planets and to establish their prospects for harboring life.

A basic difficulty in directly detecting the planets of other stars lies in the fact that stars are bright, and planet radiation is faint. Nevertheless, the greatly enhanced capability of developing instrumentation is opening up new horizons. Proposed techniques for directly observing both giant and Earth-like planets should lead to a new era in planetary science.[1,2,3,4,5,6] Once planets are detected, major atmospheric constituents and perhaps surface properties can be determined; abundances of oxygen and water comparable to those on Earth might well signal the presence of similar life. The dramatic implications of finding evidence of life elsewhere would open a new era in the study of the origins of life on Earth.

The direct detection program also addresses a related and no less outstanding question—how do planets form and evolve? Planet formation probably takes place on timescales of less than 10 million years—relatively short in cosmic terms—in the circumstellar disks that are natural by-products of the star formation process. In the nearest star-forming regions, about 150 parsecs from the Sun, hundreds of young stars (less than a million years old) have been observed. Many appear to be surrounded by disks ranging in size from 10 to 100 AU—comparable to the extent of our solar system. Indications are that planetary formation may be proceeding within these disks.

Observational studies of planetary phenomena around other stars are already proceeding. Discoveries across the electromagnetic spectrum from ultraviolet to radio wavelengths encompass systems dispersed all

along the supposed path of planetary evolution. For example, a proto-planetary disk, a dense disk of gas and dust from which planets form, around the young star, HL Tauri, has been resolved by observations at the 3-mm wavelength. A more diffuse, dusty disk within which planets may already have formed has been seen in direct optical observations of the older star, Beta Pictoris. Many such disks will be accessible to study at visible and near-infrared wavelengths from ground-based observatories. Longer wavelength measurements to investigate the cooler, outer regions of these disks require space-borne far-infrared and sub-millimeter telescopes and arrays of millimeter-wavelength radio telescopes.

4.2 PLANET DETECTION

Direct detection and measurement of planetary light offer the exciting opportunity to explore the nature of fully developed planets. Once planetary systems are detected around other stars, there will be enormous pressure to mount intensive investigations of the individual planets. Measurements aimed at obtaining information about the chemical compositions of the atmospheres of discovered planets will be of paramount importance. Very sharp interest is likely to focus on searching for the chemical disequilibria in planetary atmospheres that might indicate the presence of metabolic gases and suggest the existence of life.

PLANETARY RADIATION

Planetary radiation is much weaker than stellar radiation. Figure 4-1 shows the radiation from the Sun, Jupiter, Earth, and Uranus as it would be observed at low spectral resolution from 5 pc. At shorter wavelengths, the planet emission derives from scattered light from the 5800 K Sun. Thermal radiation from the planet itself contributes at longer wavelengths. Neglecting any internally generated heat, planets farther from the star are cooler with weaker emission, since they receive less stellar radiation. The solar and planetary spectra at wavelengths ranging from the optical to the near infrared (around $\lambda = 5$ μm) are very similar in shape. These "Planck curves" are characteristic of a neutral scatterer or emitter (such as a ping-pong ball or a perfect black body). However, the planet flux is clearly nine or ten orders of magnitude weaker than that of the star. In the thermal infrared regime, for example at $\lambda = 10$ μ, the difference is only about six orders of magnitude; beyond this wavelength, the ratio between the stellar flux and that of the planet

becomes progressively smaller.

The TOPS search program addresses familiar stellar families: solar-type stars—F and G dwarfs—and the cooler, fainter K and M dwarfs. G dwarfs have temperatures and masses approximately the same as the Sun; M dwarfs are about a factor of two cooler and less massive. These compose essentially all nearby stars. About 30% are expected to be free of the confusing presence of a stellar companion. Because of their large expected reflex motions, M dwarfs are prime candidates for indirect search techniques. Planets around M dwarfs will be detectable only if they are self-luminous.

The spectral energy distributions shown in Figure 4-1 should be representative of mature extrasolar system planets. Substantially brighter emission is unlikely, since no planets much larger than Jupiter are expected. The maximum expected radius of a planet or any object less massive than 0.1 solar mass is only 13% greater than Jupiter's radius.[7] On the other hand, Figure 4-2 clearly illustrates that gaseous planets like Jupiter are considerably hotter and more luminous when they are younger. They are also much larger. As they cool and contract, gravitational energy is converted into thermal emission. Although undoubtedly less pronounced than it must once have been, this effect currently raises

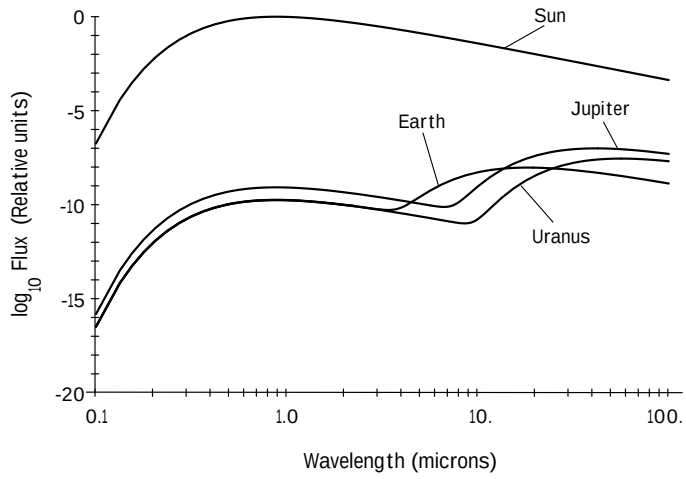


Figure 4-1. The spectral energy distributions of the Sun, Jupiter, Earth, and Uranus as they would appear at 5 pc, averaged over a 10% spectral bandpass. Note the decreased ratio of solar to planetary flux in the thermal infrared, compared to visible wavelengths.

Jupiter's effective temperature to 125 K from the 106 K that would balance absorbed sunlight at 5.2 AU, Jupiter's distance from the Sun. Such excess luminosity makes the direct detection of young "super-planets" a more straightforward proposition than detecting mature planets like those in our solar system.

The scenarios contemplated here are based on rather conventional assumptions about planetary properties. Unexpected factors may enhance radiation from planets around other stars, facilitating their detection. Earth and the giant planets, for example, all exhibit both airglow and non-thermal radio emission. Airglow is a minor effect that is not likely to be helpful in detecting an Earth-like planet around another star. On the other hand, non-thermal radio emissions caused by precipitating radiation-belt electrons dominate Jupiter's radio spectrum. At times, these emissions are extraordinarily intense, exceeding solar radio bursts. Such non-thermal signals from other stars may be detected and will need to be interpreted on a case-by-case basis. Indeed, the aim of

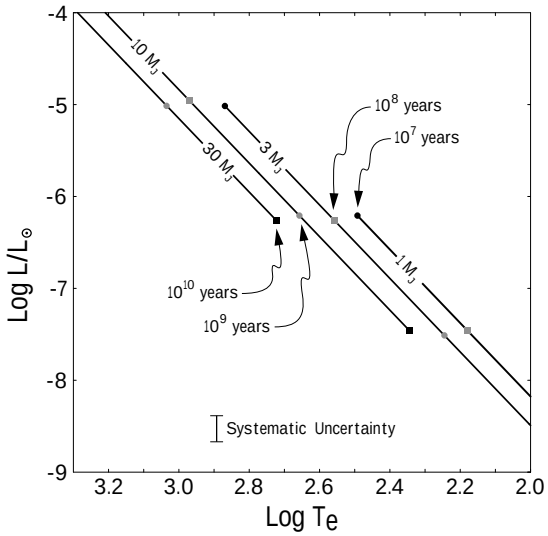


Figure 4-2. The evolution of young massive planets. The curves show how the planetary luminosity and effective temperature are expected to change with time as the planet cools. These theoretical trajectories are uncertain by as much as a factor of two in the luminosity for a given effective temperature because of uncertainties in the opacity. Moreover, the true tracks deviate progressively away from and above those shown at early times because of non-degeneracy. The tracks shown assume negligible change in radius with time.[8]

NASA's Search for Extraterrestrial Intelligence program is to detect radio messages—non-thermal signals of a rather special kind. No TOPS-related project is based on detecting non-thermal radiation, and the remainder of this chapter is restricted to discussions of thermal radiation—reflected starlight, radiation intrinsic to planets, or emission from circumstellar material.

PLANETARY SPECTRA

The gases in planetary atmospheres and the minerals on planetary surfaces will modify spectral energy distributions like those in Figure 4-1. Transitions between energy levels in the atoms or molecules will produce either absorption or emission features in the spectra. The key interactions and modifications to the spectrum occur in the atmospheric zone where the optical depth for a given wavelength, $\tau(\lambda)$, is unity. For reflected radiation, τ=1 occurs at the point on the incident beam of starlight where most of the light has been scattered or absorbed. For thermal emission, τ=1 is the level where most escaping radiation originates. At wavelengths where the atmosphere is transparent, or if there is no atmosphere, the τ=1 level may be at the solid planet surface or at a cloud deck.

In a normal planet atmosphere, the pressure at each height balances the gravitational weight, and the total gas density decreases exponentially with height. The atmospheric temperature and composition and thus the scale height are affected by local processes. Aside from demonstrating the presence of a particular atom or molecule, a spectral feature or line is diagnostic of the conditions where it forms; depending on the particular energy transition, the τ=1 level will occur in regions of different temperature at different heights in the planet atmosphere. If temperature rises with height in the τ=1 region, the line will be an emission feature; if temperature declines, there will be an absorption line; if a temperature inversion is spanned, the core of the line will be inverted. In addition, some atoms and molecules may give rise to multiple spectral lines or bands, the relative proportions of which can be another powerful tool in determining physical conditions in the planetary atmosphere.

Planck curves like those in Figure 4-1 result from low spectral resolution observations; measurements made over broad wavelength bands smear out any spectral features. For planets, such spectra will yield the total emitted power or effective temperature. At higher spectral resolution, absorption features due to molecules, minerals, or ice on the planet

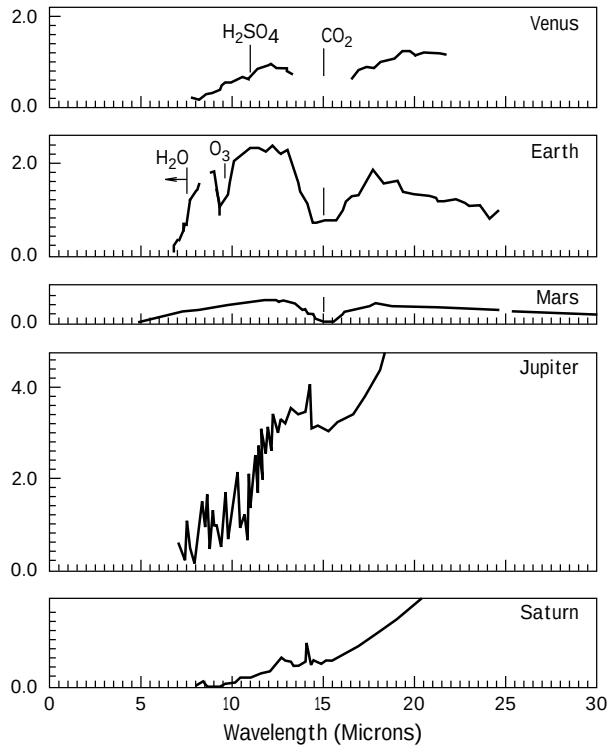


Figure 4-3. Spectra of three terrestrial planets (Venus, Earth, and Mars) and two gas giants (Jupiter and Saturn) showing prominent absorption and emission features due to atmospheric constituents.

may appear in the reflected starlight spectrum, and may produce either absorption or emission lines in the thermal spectrum of the planet itself. Figure 4-3 illustrates the power of spectral line studies for solar system planets. Each planet spectrum is distinct because of the particular mix of molecules in the atmosphere. For example, the 13.7 μm emission feature indicates that acetylene (C_2H_2) is present in the stratospheres of Jupiter and Saturn, and that the temperature increases with height. Water and ozone in Earth's spectrum reflect the presence of life; the ozone derives from oxygen that is many orders of magnitude out of chemical equilibrium. This was not so in the early days of our planet's existence, but the blue-green algae changed Earth's atmospheric chemistry forever.[9] Human life is the fortunate result. Similar inferences might be drawn for planets discovered around other stars if enough photons could be collected to permit spectroscopic studies.

PLANETARY PROPERTIES

Direct discovery of a planet through imaging techniques will indicate the orbit, the characteristic distance from the central star, and by implication, the equilibrium temperature. Longer-term observations will give the degree of circularity of the orbit and its inclination to the star's spin axis—a significant diagnostic of the cosmogonic relationship between the star and the planet. A comparison of the planet's brightness at infrared and optical wavelengths will provide information on its reflectivity (albedo) and any internal heat source and will suggest the planet's size. Dynamical information from indirect detection techniques—astrometric and/or radial-velocity measurements—will estimate the planetary masses as a fraction of the mass of the central star. The latter can be derived with some degree of accuracy from spectroscopic studies. Fundamental planetary properties, such as mass, equilibrium temperature, orbital parameters, and size, can, therefore, be obtained combining direct and indirect techniques.

If the discovered planets' orbits and masses seem consistent with the pattern observed in the solar system, similar properties might be inferred—small, rocky inner planets, with giant, light-element-dominated outer planets. Spectroscopic observations are critical to completing this picture. The infrared portion of an extrasolar planet's spectrum is likely to be an early target for spectroscopic study because of the enhanced fluxes at these wavelengths. Broad-band filter photometry ($\lambda/\Delta\lambda = 2$ to 5) fitted to Planck functions as described above would determine the effective temperature. Narrow band spectrophotometry ($\lambda/\Delta\lambda = 5$ to 20) could address the oxidation state of the element carbon in the atmosphere.

In our solar system, the state of carbon neatly divides the planets with substantial atmospheres into CO_2-dominant (Venus, Earth, Mars) and CH_4-dominant (Jupiter, Saturn, Uranus, Neptune) as illustrated in Figure 4-3. The 15.0 μm CO_2 ν_2 band is very strong and is situated near the peak of a 200 K black body. For atmospheres with reduced carbon, the ν_4 band of CH_4 at 7.7 μm could directly reveal the presence of methane. Additional indicators would be the C_2H_2 ν_5 band at 13.7 μm and the C_2H_6 ν_9 band at 12.2 μm. Both are products of methane photochemistry in a molecular hydrogen atmosphere. A detection of the 6.3 μm ν_2 band of water would be of great interest. The presence of liquid water—and therefore substantial quantities of H_2O gas in the atmosphere—is probably an essential condition for life as we know it. Relying

on a comparison with Earth's present atmosphere, it has been suggested that methane and the 9.6 μm O_3 ν_1 and ν_3 bands would be strong indicators of Earth-like biological activity.[3,10,11]

Many other strategies for characterizing planetary atmospheres will open up if enough radiation can be collected at visible and near-IR wavelengths to compensate for spreading out the spectrum.[10] Increased spectral resolution at all wavelengths will allow searches for a greater variety of less abundant molecules. The resultant cumulative knowledge of the discovered planet would merge with the expanding understanding of planetary atmospheres in our own solar system.

Although biological evolution may follow alternative chemical pathways on other worlds, the presence of molecules far from chemical equilibrium would imply similar disequilibrium processes that might lead to life-forms. Jupiter's spectrum provides a cautionary note: its stratosphere contains acetylene and ethane in much greater abundances than chemical equilibrium predicts, but solar ultraviolet radiation drives the disequilibration. On another planet, we might not be able to rule out exotic biological processes. Although life as we know it might not be responsible for disequilibrium, the effect could be caused by life as we do not know it.

4.3 PROTOPLANETARY AND DEBRIS DISKS

Available information supports our general picture of the way stars and planets form. Nevertheless, direct observations of planetary systems during their first formative 10 million years are vital to addressing the numerous questions that we still confront in understanding the creation and prevalence of planets in our universe. Detailed measurements of a large number of planet-forming and debris disks around stars with ages from 10^6 to 10^9 years are also necessary to comprehend morphological, physical, and chemical evolution. To detect planets directly, the confusing effects of the central star must be reduced—a formidable task. It is easier to observe the circumstellar material in pre-planetary form, as a planet-forming disk, or as the detritus of planetary birth, because of its relatively large angular size and concomitantly greater expected brightness. Many planetary phenomena have already been detected around other stars, and discoveries continue. The TOPS Program will build on and vastly enlarge this knowledge.

PROTOPLANETARY DISKS

Particle growth and the accumulation of planetary bodies most likely take place in disks of gas and dust around very young stars. The detection of far-infrared (25 to 100 μm) and millimeter-wave emission substantially in excess of that expected from the stellar chromospheres alone provides persuasive evidence that about 50% of T Tauri stars—objects similar to the proto-Sun—are surrounded by dense disks of dust.[12,13,14,15] The masses of these disks are significantly greater than a few percent of the mass of the Sun, and at least as large as required for proto-solar nebula.[15,16] In some cases, the circumstellar gas appears to revolve around the stars in Keplerian orbits, as might be expected for a forming planetary system,[17,18] and the dust grains seem to display the fractal characteristics anticipated in pre-planetary disks.[19, 20] Future observations must address the question of whether these disks are capable of forming planets.

Detailed knowledge of the morphology, mass, density and temperature distributions, velocity fields, and chemical gradients of the pre-planetary disks is required. Their evolutionary history must be established and compared with what is known about the solar system. It will be important to ascertain when disk material accumulates into planetesimals and when gaps are swept out by newly formed planets.[21] Very high spatial resolution observations are required for studies of these 10 to 100 AU disks. Likewise, the expected small-scale velocity variations require high spectral resolution. Both demand large collecting areas.

Since they are found in optically obscured star-forming clouds, protoplanetary disks are usually investigated at longer wavelengths. Near-infrared observations, for example, will probe the hot regions inside 5 AU. The cooler outer regions are best studied at wavelengths close to 1 millimeter. Many molecules emit millimeter-wave spectral lines that can be used to determine the chemical composition and velocity structure of the disk (see Figure 4-4). However, to achieve even 1 AU resolution at 1 mm in a disk at 150 pc, the distance of the nearest star-forming clouds, a telescope of 30 km diameter is needed; the ultimate goal of 0.1 AU requires an instrument of untenable aperture size. As we describe in the following sections, a number of techniques have been developed to overcome this problem. At millimeter wavelengths, aperture synthesis interferometry can in principle provide the requisite resolution.

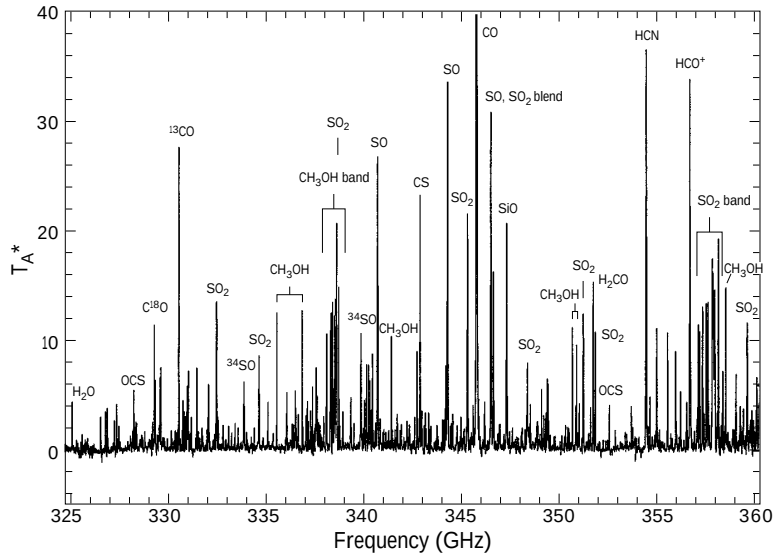


Figure 4-4. The spectrum of the core of the Orion star-forming cloud in the 0.9 mm wavelength band. The vast numbers of spectral lines are diagnostic of different conditions of density and temperature throughout the nebula. High-resolution aperture synthesis maps made in different molecular lines will provide detailed information about the physical and chemical properties of the region.

Several synthesis arrays operating at millimeter wavelengths already exist; Caltech's Owens Valley Radio Observatory and the Berkeley-Maryland-Illinois Array are examples. Observations of protoplanetary disks are under way. One example of such a disk, HL Tauri, was illustrated in Chapter II. At present, the spatial resolution attainable is just under 1 arcsec, or about 100 AU at the distance of the nearest candidates. At this resolution, and at current sensitivity levels, a disk of diameter about 100 AU, comparable to our solar system, has been seen at the core of a large star-forming cloud, L1551,[22] and Jupiter-mass condensations have been detected in the HL Tauri structure.[18]

The planned development in the U. S. of an extensive array of millimeter telescopes will improve spatial resolution at 1 millimeter to tenths of an arcsec. In the meantime, relatively minor increments to arrays that are already operational will significantly increase their power for disk observations. Such studies will contribute a solid foundation for our understanding of how planetary systems form and will complement the optical and near-infrared disk programs envisaged for TOPS.

DEBRIS DISKS

The disks around young stars described in the previous section are the building blocks of planetary systems. As planets accumulate, the material remaining in orbit about the aging central "sun" will become more tenuous and more difficult to detect. By contrast, planetary bodies will become increasingly detectable. In effect, the dust opacity declines as the fraction of material bound up in sizeable bodies increases. These waning disks are associated with stars older than 10 million years, and are often called "debris" disks to emphasize that planets may have already formed.

The Infrared Astronomical Satellite (IRAS) alerted the astronomical community to the possible existence of planetary and debris disks.[23] The clue was again the detection of infrared fluxes in excess of those expected from the stellar photosphere alone. At least 30% of nearby main-sequence stars show evidence for disks at the IRAS sensitivity limit,[24] although the star is 10^5 times more luminous than the disk. Only three have been resolved spatially: those associated with the stars Vega, Beta Pictoris, and Fomalhaut.[25] The presence of a disk around Beta Pictoris has been confirmed by means of ground-based coronagraph imaging.[26] Spectroscopic studies of this disk from the ground[27] and from the Hubble Space Telescope[28] have provided tantalizing hints of gaps and a central hole, as would be anticipated if planets had formed. In addition, there is spectral evidence for the presence of micron-sized silicate particles[29] much larger than typical grains in the interstellar medium.

As with protoplanetary disks, understanding of debris disks will come from complementary high-resolution measurements at optical, infrared, and millimeter wavelengths. The techniques described in the next section to achieve high sensitivity and resolution in the visible regime will be particularly appropriate. Most debris disks are at the limits of detectability for the current millimeter-wave interferometers,[30] although the upgraded arrays will be capable of detecting gaps that could be gravitationally induced in the debris disks by a Jupiter-sized planet.[31] For the nearest objects at about 3 pc, 1 arcsec resolution corresponds to about 3 AU, and a Jupiter-sized planet at about 20 AU from the central star should sweep out a gap about 9 AU wide.

In the future, more sensitive, single-aperture, far-infrared measurements from the European Space Agency's Infrared Space Observatory and NASA's Space Infrared Telescope Facility, together with 3 millimeter observations from NRAO's Greenbank Telescope, will provide sur-

veys of the masses and radial distributions of circumstellar material around significant samples of solar-type stars of various ages, and will identify many more disk candidates. The new infrared instruments will be considerably more sensitive than IRAS and will achieve better resolution—sufficient at mid-infrared wavelengths to resolve disks comparable to Beta Pictoris around Sun-like stars of types F, G, K, and even M dwarfs. Disk evolution will be studied by observing nearby open clusters with well-established ages. Searches for silicate absorption like that seen in Beta Pictoris—and strikingly similar to that detected in Comets Kohoutek and Halley—will also be possible. The overall results will provide a large and varied sample of both planet-forming and debris disks for further analysis in the TOPS Program.

4.4 INSTRUMENT STRATEGIES

Attaining the observational goals of TOPS requires state-of-the-art instrumentation that can only be developed over a period of years. The observing programs envisaged will encompass broad synoptic studies of many objects, as well as highly focused, detailed investigations of individual systems. A long-term objective will be to achieve spatial resolutions significantly finer than 1 AU, although interim resolutions of 1 to 10 AU will contribute essential information, especially about protoplanetary and planetary disks. The most effective approach is to proceed through staged observational programs and evolving instrument development, exploiting every aspect of both filled-aperture telescopes and interferometric arrays.

Three instrument capabilities are fundamental to the investigation of planetary system formation and the detection and characterization of planets. First, the inherent angular resolution must be adequate to spatially distinguish the objects of interest. Second, radiation from the planet must be sufficiently intense to produce a high signal-to-noise ratio (SNR). Third, it must be possible to discern faint planet radiation in the presence of diffracted and scattered light from the bright central star. Even the study of circumstellar, planet-forming disks can be limited by diffraction effects. The problem is severe at both visible and infrared wavelengths. Methods of dealing with these issues are considered below.

ANGULAR RESOLUTION

At all wavelengths, astronomical telescopes are limited by the wave

properties of electromagnetic radiation. Radiation of a given wavelength, λ, passing through a telescope aperture of diameter D, is necessarily diffracted into an angle of approximately $2\lambda/D$—the resolution. Thus, to achieve the same resolution as a 1-meter telescope operating in the visible ($\lambda = 0.5$ μm), an infrared ($\lambda = 10$ μm) telescope 20 meters in diameter is required. Angular resolution requirements for studying planetary formation were presented in Chapter II. The ultimate goal is to achieve about 0.1 AU resolution on planet-forming disks. For the nearest candidates at a distance of 150 parsecs, this corresponds to an angular resolution a little better than a milliarcsec. However, linear resolution of even 10 AU, which is 0.1 arcsec for the nearest clouds, would enable enormous advances in understanding the evolution of disks.

In principle, planets can be detected with only a slight augmentation in telescope size. For relatively nearby stars within 10 parsecs of the Sun, the Earth-Sun distance is equivalent to about 0.1 arcsec; a Jupiter-like planet would be separated from its star by more than 0.5 arcsec. Relatively mundane instruments can provide the necessary angular resolution, but distinguishing the faint image of a planet from that of its far brighter companion star dictates the development of more sophisticated techniques.

SIGNALS AND NOISE

The comparison in Figure 4-1 between the stellar and planetary fluxes expected from a planetary system like our own is sobering. As the following sections will make clear, the dynamic range problem—how to discriminate between a bright star and a close, faint companion—appears to be soluble, but a further limitation is imposed by photon statistics. At optical wavelengths, the total flux expected from an Earth-like planet 5 parsecs distant is about one photon per second per square meter. For a Jupiter-like planet, the expected flux is about four photons per second per square meter. At infrared wavelengths, the number of received photons increases by two to three orders of magnitude, much improving the photon statistics, but other factors can degrade performance. An observing system to detect visible and infrared radiation from other planets must, therefore, combine the ability to average data acquired over extended periods with the large collecting area demanded by the resolution requirements.

At optical wavelengths, the dominant source of background will

consist of diffracted stellar light, a figure error scatter halo, and scatter due to low order aberrations associated with telescope misalignment. The signal-to-noise ratio (SNR) can be expressed as:

$$SNR = \frac{P}{\sqrt{(P+B)}}\sqrt{QAt}, \qquad (4.1)$$

where P is the rate at which planet photons are collected per unit collecting area times the fraction of planet photons in the central image core. B is the rate at which background photons are collected; it is the rate per unit area at which photons arrive from the star times the local value of the stellar point spread function at the planet's location times the area of the image core. Q is the system end-to-end efficiency, A is the telescope collecting area, and t is the integration time. For the case at hand, the stellar background will always be brighter than the planet flux and the dominant source of shot noise. Thermal emission from the telescope and sky will be the limiting source of background noise at infrared wavelengths.

Diffracted and Scattered Light

It might appear that the conditions for planet detection are satisfied if the nominal resolution of the telescope is adequate to discriminate a planet from its parent star, and if sufficient photons can be collected from the planet to achieve acceptable statistics. This is not the case. Even a perfect, filled-aperture telescope fails to contain all the light of the star within its central diffraction spot.

In the focal plane, the radiation intensity of a stellar image follows the familiar Airy pattern $[J_1(\pi\kappa\theta)/(\pi\kappa\theta)]^2$ as a function of angle θ, where κ is the aperture diameter measured in wavelengths. The half-power point is $0.61(\lambda/D)$ from the center of the image, the first null is at $1.22(\lambda/D)$, and from there the diffraction rings ripple outward. The response is plotted in Figure 4-5. The fall-off in diffraction ring (sidelobe) intensity is relatively slow, following an inverse-cube relation far from the central peak. These diffraction rings are so intense that at first sight, a straightforward search for planets by direct imaging seems untenable.

Fourier Transforms

Modern filled-aperture and interferometric techniques provide the means of obtaining very high resolution and suppressing the diffracted stellar radiation that creates excess noise and prevents the detection of

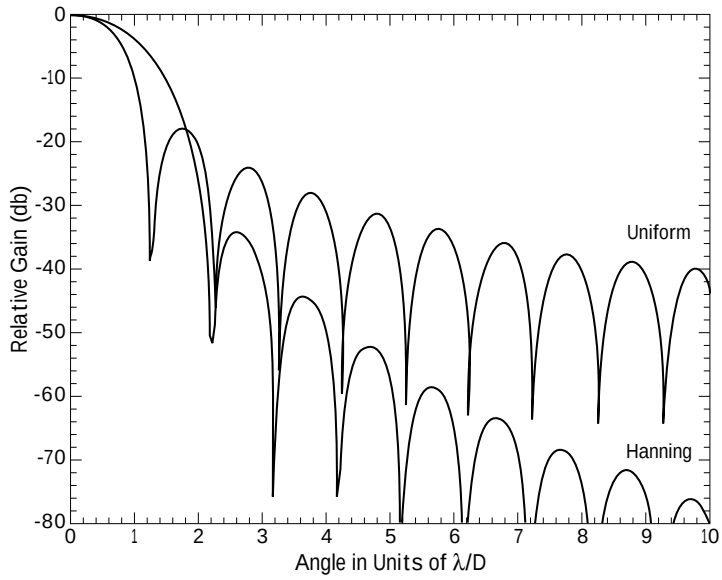


Figure 4-5. The radial dependence of the unblocked aperture (i.e., uniform illumination) compared to the pattern for a Hanning taper (apodization). The response is logarithmic (10 db = one factor of 10); note the dramatic reduction in the far pattern of the tapered response. The increased width of the principal response should also be noted.

planets. The Fourier transform is an operational tool in widespread use in scientific and engineering fields. As described below, Fourier methods can be used to achieve very high resolution and overcome diffraction effects.

Most reasonable functions can be expressed as superpositions of sine waves of all "frequencies," with amplitude and phase adjusted properly for each "frequency." The amplitudes and phases are called the complete spectrum, and the Fourier transform is simply the mathematical operation that carries the given function into its spectrum representation, and vice versa. (Frequency is in quotes above because, when the independent variable of a given function is time, the Fourier dual is frequency. Any dual variables will behave similarly.) It is possible to switch between conjugate functions (the function and its spectrum) and their dual spaces freely, depending upon the physical situation.[32] For a uniformly illuminated circular aperture, the particular distribution of Fourier components can be manipulated and given different weights before the final image is formed. This process is called apodization by opticians and illu-

mination tapering by radio engineers; when performed in the Fourier representation, it is often referred to as Fourier weighting.

One example is the case where the aperture receives an illumination taper—an apodization proportional to radial distance, falling to zero at the edge. The diffraction pattern drops off as the sixth—rather than the third—power of the angular distance from the central peak.[33] This highly desirable suppression of the intensity is, unfortunately, offset by loss of area and a corresponding degradation of the equivalent angular resolution of the central diffraction peak. Three-quarters of the collecting area is lost, and the width of the central diffraction peak is twice that of the Airy distribution. A compromise, the Hanning taper, combines relatively high area efficiency and good sidelobe suppression. The Hanning function is simply a raised cosine, maximum at the origin with its first minimum at the edge. Half the area is lost, but as illustrated in Figure 4-5, the sidelobe level is dramatically lower than for the case of uniform illumination.

Telescope optical manufacturing and alignment errors cause the redirection of energy from the central core of the point spread function into the sidelobes. For the reasonably smooth optical system under discussion here, redirection is well described by the grating equation; a sinusoidal figure error of spatial period P diffracts light into an angle inversely proportional to P and proportional to the wavelength of observation. For example, 10 cm period errors at 0.5 μm diffract energy into a field angle of one arcsec. Although this is entirely a diffractive redirection of light, it is commonly called mirror scatter.

The efficiency with which light is scattered is proportional to the square of the rms depth of the error in waves. An optical system with fixed figure errors will look 20 times smoother in waves at 10 μm than it will in the visible, for example. Consequently, the scattered radiation level will be 400 times lower at 10 μm. Low-order system aberrations, such as focus or coma and astigmatism arising from secondary mirror misalignment, will also redirect energy into the stellar halo. Although usually thought of as low spatial frequency phenomena, these actually produce a spatially infinite scattered light distribution that is asymptotically parallel to the diffracted light halo.

The effects of scattered light can be significantly reduced if the mirror can be polished to an unusually smooth surface. Techniques for super-polishing developed for micro-electronics have emphasized modest

aperture spherical mirrors. Recent work has demonstrated a figure about ten times smaller than that attained with conventional polishing, on a spherical mirror of 0.3 m diameter. By contrast, for planet detection, apertures on the order of 2 meters with aspheric figure will be needed. In addition, the level of scattered light must be reduced by several orders of magnitude. A major technical development program must be implemented to meet these goals.

CORONAGRAPHS

Suppressing sidelobe light in the point spread function clearly requires innovative technology. Each optical surface in a telescope system has imperfections so that further surface roughness sidelobe degradations are contributed at each stage. Errors are usually uncorrelated from stage to stage, and add randomly. Prime-focus telescopes are not likely to be applicable for planet searches, and even a simple Cassegrain configuration introduces difficulties. A perfect secondary mirror is itself an obscuration in the entrance aperture, with an accompanying set of diffraction perturbations that must be minimized.

In the 1930s, Bernard Lyot invented his coronagraph to observe the solar corona.[34] This device has found new applicability in the diffraction-limited regime against a point source. The Lyot coronagraph has three concatenated imaging elements that may be thought of as Fourier transforms. The telescope, which forms a first field-image plane, has a single, small circular mask on the optic axis. The second element is a small lens or mirror that forms an image of the entrance aperture; on this an oversized mask intercepts the edges of the pupil image. The last optical element is another lens or mirror that reforms the field image on the detector. The mask in the image plane operates on the elements of the image; the mask in the pupil plane weights the Fourier components.

For an off-axis source, the small circular occulting mask is bypassed at the first image plane, and has no significant effect. A very small amount of light is lost to the undersized pupil-plane mask. As a result, the detector receives essentially the same off-axis image it did in the absence of the coronagraph. For an on-axis source, however, the core of the image is occulted by the mask. The resulting pupil plane image is the Fourier transform of the point spread function with its central core removed. In other words, the occulting mask is a high pass filter. Since the pupil has high spatial frequency content only at its edges, the resulting

pupil image shows a concentration of light at the edges; this is removed by the slightly undersized pupil plane mask. The light is then reimaged at the second focal plane. The final image that results from the high-pass pupil filter applied by the occulting mask and the low-pass image filter applied by the Lyot stop shows reduced diffraction sidelobes.

Given the field-of-view requirements of the planet detection problem, the Lyot coronagraph can suppress the Airy wing by two orders of magnitude with acceptable penalties in collecting area. Recent coronagraphs have gained an additional factor of ten or more in rejection of diffracted light by hybridizing the basic design with graded-transmission masks on both the first focal plane and the pupil plane (tapering or apodizing). Theoretical models show that the graded image mask renders

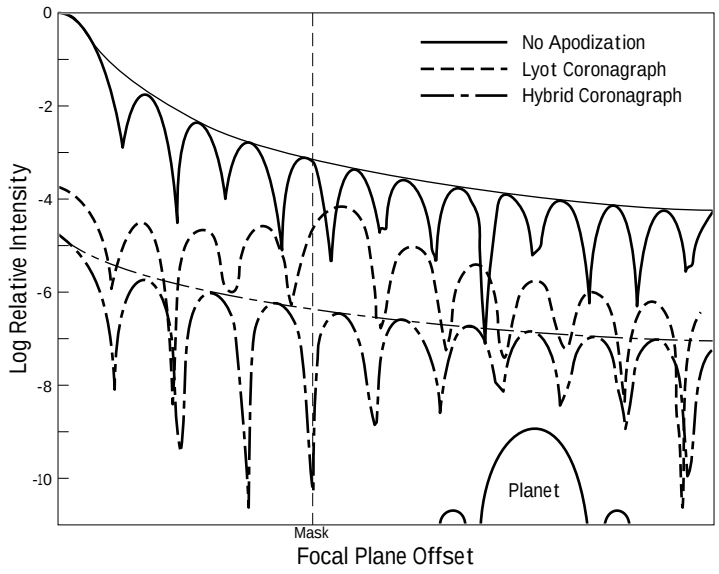


Figure 4-6. Hybrid coronagraph performance in reducing diffracted starlight. The intensity of diffracted starlight in the focal plane is a function of distance from the center of the Airy disk. The solid lines indicate the falloff in intensity of a solar-type star normalized and compared to the intensity of a Jupiter-size planet located at 5 AU and at 10 pc distance. No apodization was used in this case to show that the relative brightnesses of the star and planet are about 9 orders of magnitude apart. The dashed lines show the reduction in the diffracted light for a classical Lyot type and a hybrid type coronagraph. The coronagraphic efficiency of the hybrid design can be tuned over a large range, but is shown here reducing diffracted starlight by about 3 orders of magnitude.

the Lyot stop more efficient, while allowing some transmission very close to the central star—a valuable portion of the field of view. For targets near the center, the coronagraphic efficiency is virtually independent of angle, after mask transmission losses are taken into account. The grading reduces the diffraction wings of the pupil plane image. With a more compact pupil plane image, the pupil mask can be smaller, and less light is lost for off-axis sources. The performance of a hybrid coronagraph in reducing diffracted starlight is illustrated in Figure 4-6.

INTERFEROMETERS

More than a decade ago, infrared Michelson interferometers were proposed as a way of cancelling out starlight, allowing nearby planets to be detected directly.[35] The Michelson interferometer provides a concrete example of Fourier transform measurement. First realized as an optical device, then used far more extensively by radio astronomers, the Michelson interferometer is illustrated in Figure 4-7. The instrument allows separate beams of radiation from a source to interfere: if the path lengths are equal, the amplitudes add; if they differ by half a wavelength, they cancel. In this example, each entrance aperture receives radiation from a point source. The path lengths are equalized by inserting a variable delay in one arm of the interferometer. An optical system (schematized by a lens) focuses and merges the beams. In this image plane representation, the diffraction pattern of each entrance aperture

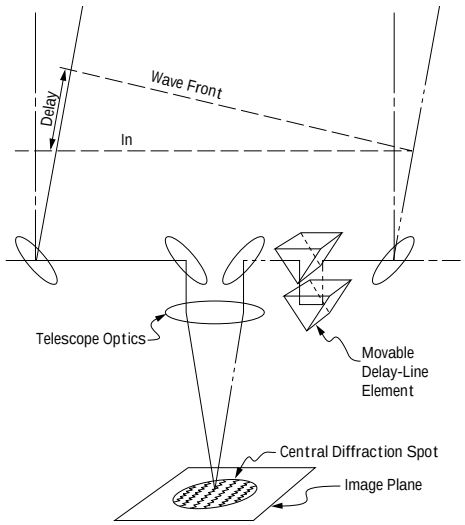


Figure 4-7. A Michelson interferometer in the image plane version. The incoming rays are reflected off plane mirrors, and the diffraction spots are merged by the optics of a telescope (a lens is shown, but a set of reflecting optics works as well). The two beams alternately interfere destructively or constructively, giving fringes in the image plane. The path lengths are equalized by delay line optics.

(the Airy pattern described above) is formed on an area detector such as a charge-coupled device (CCD). The image, composed of the superimposed diffraction patterns, will have a central diffraction spot-size subtending an angle approximately λ/D; D is the aperture diameter, and λ is the observation wavelength. This image is crossed by interference fringes or alternating bands of maxima and minima with the maxima being separated by λ/B. B is the baseline length or the distance between the independent elements.

The intensity of the radiation from every source in the sky can be described in terms of position, usually right ascension (α) and declination (δ), wavelength (λ) or frequency (ν), and time (τ). The basic principles of various telescope systems can be discussed with reference to this intensity, $I\ (\alpha, \delta, \nu, \tau)$. Each measurement is an average over a band of frequencies or wavelengths, and over some span of time; here, the single-frequency, instantaneous approximation is made to simplify discussion. The resulting specific intensity $I\ (\alpha, \delta)$ that describes the distribution of source brightness on the sky has a Fourier conjugate I (u,v); the dual coordinates (u,v) are usually called the spatial frequency. Since the spatial frequency is two-dimensional, given values of u and v are said to be located in the u-v plane. When the Fourier conjugate I (u,v) is the quantity measured, summing over the u-v plane is equivalent to making a measurement in the image plane of a telescope.

The fundamental theorem of Michelson interferometry is that the fringe amplitude and phase are proportional to I (u,v), which is usually treated as a complex amplitude. The u-v plane is normal to the direction in which the interferometer elements are pointed; u and v are the coordinates, measured in wavelengths, of the baseline projected onto this plane. This geometry is illustrated in Figure 4-8. If separate fringe measurements were made for all possible baseline lengths and orientations

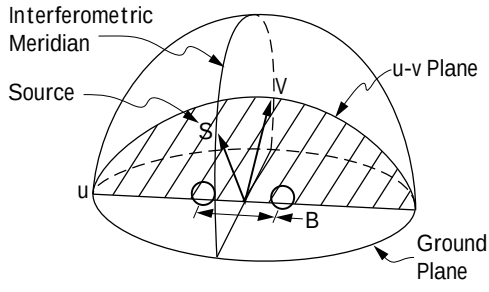


Figure 4-8. Representation of the u-v plane for a Michelson interferometer. In this case, the source is in a direction perpendicular to the interferometer baseline B. The u-v plane is constructed normal to the observing direction.

within a circle of diameter *D*, the information in the derived image would be equivalent to that in an image obtained by a conventional telescope of aperture *D*.

Many individual elements, combined by pairs as Michelson interferometers, make an interferometric array, often called a synthesis array.[36] Since *N* elements can be combined pairwise in $N(N-1)/2$ ways, the number of points measured instantaneously in the u-v plane grows approximately as the square of *N*. As Earth and the array rotate, a separate set of u-v plane positions can be sampled. An interferometric array on a spacecraft can be physically rotated to accomplish the same end—rotation synthesis.[37] In the data reduction process for such array observations, the u-v measurements are gridded and Fourier transformed to give a raw map of the observed source. The image of a point source obtained in this way is termed the dirty beam in radio interferometry, or the point spread function in the optical domain. In effect, the raw map is just the desired image convolved with the "dirty beam" or the point spread function. A variety of techniques (e.g., the CLEAN algorithm, maximum entropy, hybrid mapping, self-calibration) can be applied to reconstruct this image.

The close relationship between the Michelson interferometer and a conventional filled-aperture telescope is shown in Figure 4-9. After reflecting off the parabolic mirrors, all the rays from a distant point source arrive at the focus with the same phase. Two portions of the primary mirror can be regarded as elements of an interferometer: the rays interfere constructively at the focus, to which the light has been conducted by the optically determined ray path. In the example of Figure 4-7, the optical paths were different, but constructive interference also resulted. Radio astronomers use electrical transmission lines instead of geometrical optics to achieve equivalent results. A conventional telescope can be regarded as a filled array of Michelson interferometers, with the results summed at the focus.

The generalization of interferometry to the synthesis array concept raises possibilities that hold major promise for both planetary detection and for studying circumstellar material. Interferometry at centimeter radio wavelengths is already highly developed;[36] the Very Large Array (VLA) is the largest example. More recently, arrays that operate at millimeter radio wavelengths have also been constructed: examples in the U.S. are the Owens Valley Radio Observatory Array and the Berkeley-Illinois-Maryland Array. Sub-millimeter arrays that may allow measure-

ment of the critical H_2O line near 600 μm are planned. Two-element Michelson interferometers currently in operation at infrared and optical wavelengths are illustrating both the difficulties inherent in the technique and practical solutions to the problems.

Aside from signal-to-noise considerations, the physical principles of optical and radio interferometry are not essentially different, and the hardware components are almost equivalent. Figure 4-9 compares the modern optical and radio versions of a Michelson interferometer in the pupil plane realization in which no image is formed. The radio signals enter a hybrid junction, a four-terminal pair network that takes two inputs and gives two outputs—the in-phase and out-of-phase combinations of the input signals. These outputs then go to separate detection systems. In effect, the product of the field of amplitudes is taken by operating on the two outputs. The optical realization in the lower diagram is equivalent, although the elements have a physically different form. A beam-splitter corresponds exactly to a hybrid junction. No amplifiers intervene because they would add too much noise to any optical system. The in-phase and out-of-phase combinations are detected separately;

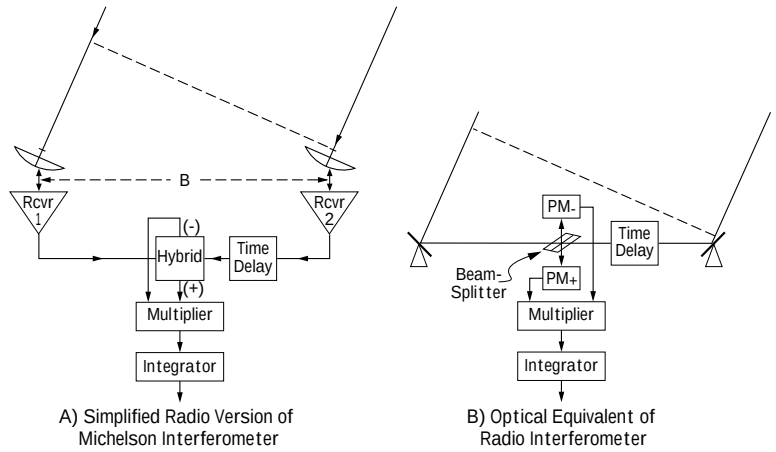


Figure 4-9. Comparison of modern versions of a Michelson interferometer. In A, the signal processing is often performed digitally. In B, there are no amplifiers, since they would add an unacceptable amount of noise at optical wavelengths. A beam-splitter functions exactly as its radio analog, the hybrid junction. One output beam gives the sum of the amplitudes with 180° phase shift. The two outputs, after detection in the photomultipliers PM– and PM+, can be processed to give the product—just as in case A.

these outputs are then processed by an electronic system that produces the complex fringe amplitudes by operations equivalent to the radio case.

The lack of amplification in the optical interferometer example illustrated in Figure 4-9 is driven by fundamental physical reasons. For even the brightest stars (excluding the Sun), we receive far fewer than one photon per Hz of bandwidth per second per atmospheric seeing disk. For a full 10-meter aperture, we receive only about 0.005 photon/sec/Hz. Even with the active optics described below, the situation is not significantly improved. It is not possible to amplify this signal without losing phase information. By contrast, in the radio regime, phase-preserving amplification is the norm; it is possible to combine N telescopes into $N(N\text{-}1)/2$ baselines simultaneously without loss of signal-to-noise ratio, by amplifying and recording the signals from each telescope, and subsequently recombining them. In the optical regime, the signal from each telescope would need to be divided in real time among the other N-1 telescopes with a corresponding reduction of signal-to-noise ratio unless the number of photons is sufficiently large. Interferometry at infrared wavelengths presents an intermediate case.

Two factors limit the effectiveness of optical and infrared interferometry. First, mirrors have wavelength-and-polarization-dependent phase shifts, which vary from sample to sample. Mismatches in the mirrors limit the depth of the null. Second, an uncompensated pointing error shifts the nulling pattern, whereas pointing jitter decreases the depth of the null. Space-based interferometric starlight cancellation techniques are expected to be more effective than ground-based techniques, in which varying atmospheric turbulence corrupts the phase fronts and therefore the fringes. To prevent over-resolution of the star and consequent lessening of the degree of cancellation, a modest-baseline space interferometer is preferred for planet detection at visible and infrared wavelengths.

The interferometric approach to planet detection will, at least in the short term, concentrate on very simple interferometric systems. Moreover, although we talk of direct detection of a planet by interferometric imaging, no image of the planet per se will be formed with the first generations of interferometers. Direct detection will consist of finding a statistically significant increase in brightness at a reasonable distance from a star. Confirmation at this stage will depend on repeated detections in a series of positions that could reasonably represent the planetary orbit.

Intensive study of interferometric arrays will, of course, continue not only from the standpoint of planet detection, but also with a view to investigating the circumstellar material. These studies can span a wide range of the electromagnetic spectrum. A correspondingly broad selection of techniques will be brought to bear, including infrared arrays that use optical or quasi-optical methods as well as millimeter- and sub-millimeter-wave arrays using radio technology.

ADAPTIVE OPTICS

At visible and near-infrared wavelengths, the angular resolution of telescopes with apertures greater than 10 to 20 cm is limited by turbulence in Earth's atmosphere rather than by the inherent, diffraction-limited image size of the system. Nevertheless, telescopes larger than the critical site- and wavelength-dependent size r_0 can be equipped with adaptive optics subsystems to compensate for atmospheric turbulence effects and achieve imaging on scales that approach the diffraction limit. The efficiency of an optical or infrared interferometer with individual input apertures larger than r_0 will also be seriously reduced by wavefront errors resulting from atmospheric turbulence. Unless the light beams from each aperture are adaptively corrected before being combined, the fringe pattern could be completely destroyed by turbulence effects. Adaptive optics systems are crucial to the success of many TOPS goals. Adaptive optics will allow better resolution of circumstellar disks and sharpen the images of more evolved stars, thereby enhancing the sensitivity of coronagraphic searches for planetary companions.

Adaptive optics systems continuously measure the wavefront errors resulting from turbulence. Using a point-like reference source situated above the distorting layers of Earth's atmosphere, compensation is effected by rapidly adjusting a deformable optical element located in or near a pupil plane of the optical system. For existing large (3- to 5-meter) telescopes, the technique has the potential to provide nearly diffraction-limited performance at visible and infrared wavelengths and is expected to yield similar performance for the new generation of 8- to 10-meter telescopes.

Turbulent degradation of astronomical images has traditionally been ascribed to three sources:[38, 39] thermally induced turbulence in the telescope and dome, the boundary layer between the observatory and the free atmosphere, and the free atmosphere itself. In recent years, astronomers have learned to control the thermal environment of the telescope and

dome, reducing this contribution to image degradation to negligible levels. The Mauna Kea site in particular does not disturb the free atmosphere flow sufficiently to induce any appreciable boundary layer. For the Keck telescopes, adaptive optics need compensate only for wavefront errors induced by the turbulence in the free atmosphere, which is largely concentrated in narrow shear layers several kilometers above the telescope site.

The optical effects of atmospheric turbulence are due to variations in refractive index. These, in turn, correspond to temperature-induced density variations in the adiabatic turbulence. Studies[40] have demonstrated that the structure functions of all these parameters follow a Kolmogorov power spectrum; the wavefront degradation is normally characterized by the Fried parameter, r_0. In terms of the structure function C_N^2:

$$r_0 = 0.432 \left[(2\pi/\lambda)^2 \int C_N^2(h)\, dh \right]^{-3/5}, \qquad (4.2)$$

it increases significantly at longer wavelengths.[41] Other relevant parameters are the isoplanatic angle (determined by the distribution in height of the turbulent layers and the diameter of the telescope), and the coherence time (determined by r_0 and the wind speed at the turbulent layers).

In a conceptual adaptive optics system, errors in the wavefront from a known high-contrast source are sensed and compensated at a small, deformable mirror, upon which the pupil is reimaged. The image quality is improved over the isoplanatic patch (i.e., within a fraction of the isoplanatic angle of the reference source). If the wavefront error is sensed and corrected within each r_0 diameter of the telescope aperture at a rate faster than the coherence time, image quality approaching the diffraction limit should be achieved. Intermediate approaches such as sensing and correcting each $4r_0$ diameter or some limited number of independent subapertures (Zernike terms) in the wavefront can also lead to major improvements in image quality.[42] Although the specific requirements for TOPS-0 imaging and interferometry have not yet been determined, Figure 4-10 shows the substantial gains in Infrared Telescope Facility (IRTF) image quality from applying tip-tilt corrections only.

The Fried parameter r_0 and the coherence time both increase as $\lambda^{6/5}$. Thus the number of turbulence cells to be sensed and corrected falls off with wavelength as $\lambda^{12/5}$; the number of photons sensed to provide an error signal (cell area times measurement time) rises as $\lambda^{18/5}$. Since the tar-

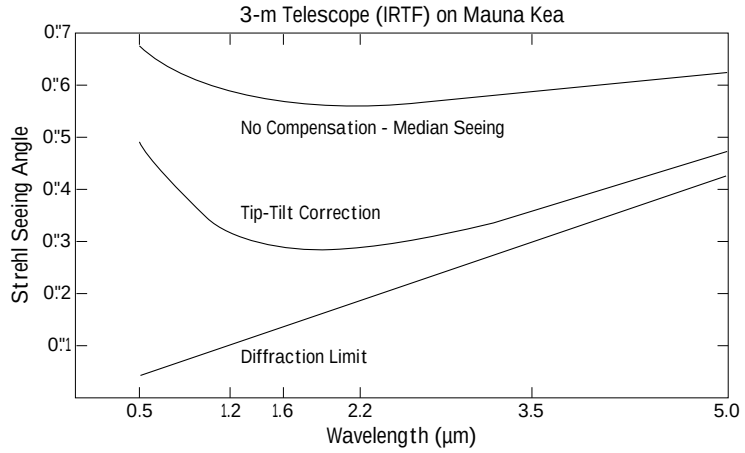


Figure 4-10. Improvements in IRTF image quality after tip-tilt corrections.

get object must be within the isoplanatic patch surrounding the reference star, it is critical to the success of adaptive optics techniques that a sufficient number of well-dispersed bright reference stars be available. At longer wavelengths, star counts rapidly increase at fainter magnitudes and, therefore, sky coverage markedly increases for adaptive techniques utilizing natural reference stars. The isoplanatic patch is also larger. Thus, the TOPS program will exploit this enhanced potential for adaptive techniques in the infrared.[43] A way to eliminate the need for a natural reference star within a few arcsec of the target object has also been envisaged. Ground-based lasers, beamed along the line of sight of the telescope, can generate "artificial guide stars" in the sky, close to sources of interest.

For most areas of optical and infrared astronomy, enormous gains will be achieved through adaptive optics. Indeed, the recent report by the National Academy of Sciences, *Decade of Discovery in Astronomy and Astrophysics*, gave highest priority to aggressive development of this technique. Simple tip-tilt adaptive correction techniques are already under study, and deformable mirrors, wavefront sensors, wavefront reconstruction processors, and other adaptive optics components have been developed. Components of the Starlab Wavefront Control Experiment, a space demonstration recently cancelled by the Strategic Defense Initiative Organization, will now be available for adaptive optics technology for ground-based astrophysics and planet detection studies. NASA's Office of Aeronautics and Space Technology has recently initiated technol-

ogy development for a 12-meter-diameter adaptive beam director for laser power beaming. This segmented optical system will be designed to achieve diffraction-limited performance with r_0 as small as 2 cm. TOPS stands to benefit significantly from this broad community interest in adaptive optics technology.

4.5 TOPS-0: THE KECK II TELESCOPE

The W. M. Keck Observatory on Mauna Kea, Hawaii, will be the cornerstone of the ground-based observing program for TOPS-0. The Keck Observatory's two 10-meter telescopes may be operated independently, employing various focal-plane instruments, or combined to form a single interferometric system. The Keck design, incorporating 36 segmented mirrors, has produced the world's largest photon-collecting surface for visual and infrared astronomy—an extremely important asset for TOPS research.

Investigations of the morphology and composition of both protoplanetary and post-planetary (debris) disks require observations that push the very limits of technological capability of ground-based facilities. The Keck telescopes, augmented by adaptive optics and exploiting the superior atmospheric conditions on Mauna Kea, are expected to operate very close to these limits. Figure 4-11 shows the potential effect of tip-tilt corrections and independently controlled subaperture compensations (Zernike terms) on the Keck telescope. Direct infrared imaging and spectroscopy with Keck II should lead to major advances in our understanding of the formation and evolution of stellar and planetary systems. Indirect searches for planets will also benefit from Keck II observations, since the greatly enhanced light-gathering power of the 10-meter mirror will vastly increase the number of stars that can be examined. One of the most ambitious goals for the ground-based phase of TOPS is the direct detection of massive substellar objects around other stars. The high-resolution imaging capability requirements for this project might be realized by an interferometer system composed of an array of 1.5-meter outrigger telescopes and at least one of the Kecks.

Other telescopes will be used to augment or complement the ground-based TOPS-0 investigations, especially for studies of a synoptic nature, which tend to require large amounts of observing time. The 3-meter Infrared Telescope Facility (IRTF) on Mauna Kea, which is wholly supported by NASA, has recently undergone substantial instru-

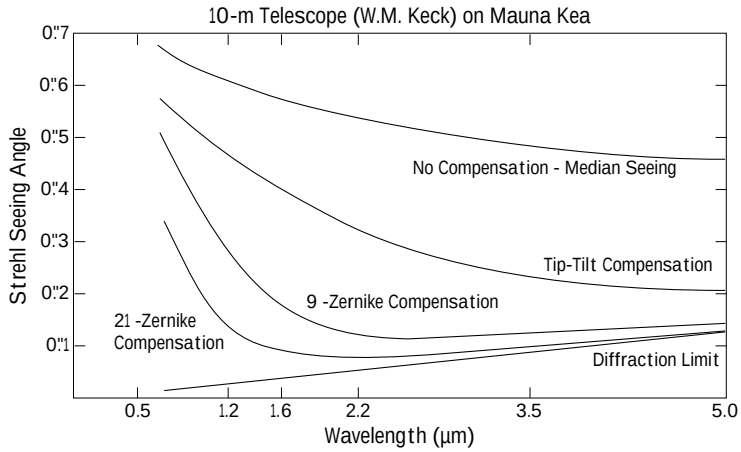


Figure 4-11. Expected improved performance of Keck I with tip-tilt and Zernike compensation.

mental and environmental improvements and meets many of the image-quality criteria for the TOPS-0 program. The median seeing in the near infrared (around 2 μm wavelength) is now 0.5 arcsec. The introduction of active optics is expected to provide sustained image quality of 0.4 arcsec or better throughout much of the near infrared wavelength region. Using modern cameras and moderate ($\Delta\lambda/\lambda \approx 1/100$ to $1/1000$) and high ($\Delta\lambda/\lambda \approx 1/10{,}000$ to $1/40{,}000$) resolution spectrographs with state-of-the-art CCDs and infrared arrays, the IRTF should greatly assist in collecting information necessary for detailed planning of later phases of the TOPS program and assure the most efficient use of Keck II when it comes on line.

TOPS-0 Instrumentation

The California Association for Research Astronomy (CARA) already has a well-chosen set of basic observational instruments under construction for Keck I; the suite of instruments is shown in Table 4-1. CARA will select a similar or complementary set for Keck II. Although a number of elements of the TOPS-0 science can be accomplished with similar instrumentation, the specific observational requirements of the TOPS-0 program are likely to mandate additional specialized instruments. Possible focal plane instruments for Keck II that are particularly well-matched to the TOPS-0 goals are described in the following paragraphs. However, at some appropriate time, an Announcement of Opportunity

Table 4-1. Complement of instruments for Keck-I.

1. NEAR-INFRARED CAMERA (NIRC) 1-5 μm 256 x 256 InSb Array *PI: Soifer / Mathews (Caltech)*
2. LONG WAVELENGTH INFRARED CAMERA (LWIC) 8-14 μm 10 x 64 PCA or 20 x 64 BIB *PI: Arens / Jernigan (UCB - SSL)*
3. HIGH RESOLUTION SPECTROGRAPH (HIRES) 0.3-1.0 μm 2048 x 2048 CCD *PI: Vogt (UCSC - LO)*
4. LOW RESOLUTION IMAGING SPECTROGRAPH (LRIS) 0.4-1.0 μm 2048 x 2048 CCD *PI: Oke (Caltech)*
5. LONG WAVELENGTH SPECTROGRAPH (LWS) 8-20 μm 96 x 96 BIB *PI: Jones / Puetter (UCSD)*

(AO) will be issued, soliciting proposals for instruments specific to the TOPS-0 science goals. The actual TOPS-0 instruments and their PIs will be selected at that time.

As illustrated in Figure 4-12, the Keck telescopes have a multiplicity of focal-plane positions for mounting various scientific instruments. The two $f/15$ Nasmyth platforms are expected to be the principal locations for most instrumentation. Five $f/15$ Cassegrain foci, one straight-through and four bent, are also available. The four bent Cassegrain positions are arranged in two groups around the elevation axis ring. A short $f/25$ Cassegrain focus is provided; located on the optical axis of the telescope, it is primarily for IR instrumentation and operates with a chop-

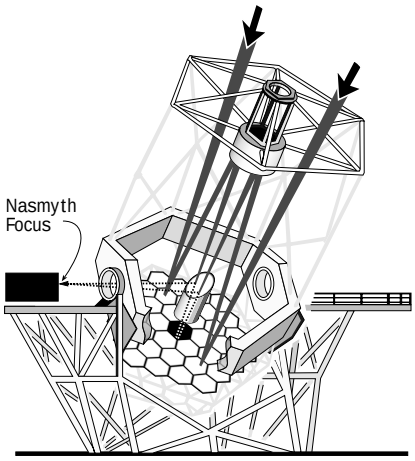


Figure 4-12. Keck telescope optical configuration and focal positions.

ping secondary. Since space is limited at this focus, CARA plans to add a new IR-optimized Nasmyth and Cassegrain focal ratio of $f/40$, using the same chopping secondary as the $f/25$ beam. The $f/40$ Nasmyth focus is expected to be the primary location for TOPS-specific instrumentation on Keck II. Provision has been made, but not yet implemented, for a coude beam on both Keck telescopes. This optical path will be required if the two telescopes are used in an interferometric mode.

Several basic observing modes—imaging, spectroscopy, astrometry, photometry, and radial velocity—have been identified for the TOPS-0 ground-based program. Instruments specific to these modes are described below.

Coronagraphic Imaging Camera (CIC)

The CIC, operating in either the CCD visible or 2.2-μm passband, is planned primarily to study disks surrounding relatively bright stars. It is best suited for disks that extend from a few to several tens of arcsec from the star. The coronagraphic technique employed by the CIC uses a graded-apodizing, focal-plane mask, and an appropriate exit-pupil mask to minimize the diffracted light from various components of the telescope. The Keck instruments must effectively eliminate diffraction from the edges of the 36 mirror segments. The performance of this complex mask has already been modelled during studies for the Astrometric Imaging Telescope (AIT), one of three candidate instruments for the TOPS-1 program. (A more extensive discussion of AIT follows in Section 4.6.)

Rotational Shearing Interferometer (RSI)

The RSI is a direct detection imaging instrument for use over broad spectral bandwidths in the infrared, where coronagraphic apodizing techniques are not as efficient as in the visible domain. The RSI suppresses the effect of the central star by destructive interference of on-axis starlight, but will not null out an off-axis object, such as a planet or brown dwarf. It is especially useful for imaging objects very close to the star. Whereas a coronagraph uses a mask to block the star out to four to eight diffraction rings, the RSI has a special advantage within the region from four diffraction rings inward to the first diffraction minimum. In practice, the degree of interference cancellation depends on the rms accuracy of the wavefront entering the RSI; high-speed adaptive optics

must be implemented. However, at λ = 10 μm, radiation from the central star can probably be suppressed by a factor of 1000.

High-Resolution IR Imaging Spectrometer (HRIIS)

The HRIIS is a cryogenic infrared (2 to 5 μm) echelle grating spectrometer that features two-dimensional detector arrays (256 x 256 InSb and HgCdTe) for spectral-spatial mapping of molecular material in circumstellar disks, and for spectral studies of candidate brown dwarfs or young "super-planets." The instrumental resolution, in the range $\Delta\lambda/\lambda$ = 1/50,000 to 1/300,000, is sufficient to resolve the Keplerian velocities of the circumstellar material or of any objects in orbit about the central star. These velocities will reflect the distance from the central stars through absorption spectroscopy. HRIIS also permits study of the gas phase abundances of molecular material in the inner regions of protoplanetary disks. Emission spectra of the inner region in the disk-clearing phase will also be enabled. Spectral line emission from a given species is diagnostic of the temperature and density in the formation region. Thus, this instrument will provide information about physical and compositional gradients as well as velocity structure in inner protoplanetary and planetary disks.

Astrometric CCD (ACCD)

The ACCD is envisioned as employing an advanced state-of-the-art CCD, probably 2048 x 2048 pixels, fabricated for metrological precision. Studies indicate that centroiding to 0.002 pixels is possible, although atmospheric effects are expected to limit the positional accuracy to 100 to 200 microarcsec. This yields a capability equivalent to a barycenter detection of Jovian planets at distances of up to 15 pc. Early testing of such a CCD on Keck I would establish the feasibility of developing an instrument of this type. The typical results expected were outlined in Chapter III.

Multichannel Astrometric Photometer (MAP)

The MAP obtains astrometric precision by moving a Ronchi ruling across a field containing a target star and 15 to 20 reference stars. The photometric signal obtained as the ruling occults these stars will provide an unusually accurate measure of their relative positions. The version envisioned for Keck II would employ appropriate detectors for both the

visible and 2.2 μm spectral regions. Precision of about 100 to 200 microarcsec is expected with an integration time of 104 seconds, sufficient to detect Jovian planets around a solar-type star at distances of up to 15 pc. Specific goals for the MAP may be found in Chapter III. The concept may also be used in the Astrometric Imaging Telescope, described in Section 4.6.

Radial Velocity Spectrometer (RVS)

The requirements on the Keck RVS for detecting a Jupiter-like planet around another star were discussed in Chapter III. Radial velocity detection of planets is not directly related to the distance of the star. Rather, the limits are imposed by the apparent brightness of the star and by factors associated with the physical characteristics of the stellar photosphere. Current RVS-type instruments, employing either a tunable Fabry-Perot etalon or absorbing molecular gas, produce comparison lines of exceptionally high fineness. It is already possible to detect shifts as small as 5 m s^{-1} in the stellar absorption lines.

THE KECK I AND II TELESCOPES

With their enormous light-collecting power, the 10-meter Keck telescopes are vital to the success of the TOPS-0 program. With the implementation of tip-tilt corrections and adaptive optics systems (see Figure 4-11), they should be diffraction limited at 10 μm—equivalent to angular resolution $\lambda/D = 0.2$ arcsec, where D is the telescope diameter. For a star at 6 pc from the Sun, this corresponds to a linear scale of about 1 AU, the distance at which planets might be expected. Direct detection of warm planets and brown dwarfs should be possible using just one Keck as a filled-aperture telescope or both Kecks as an interferometer.

Brown dwarfs are self-luminous objects, about the same size as Jupiter but at somewhat higher temperatures. Conventional imaging, using a single telescope, is likely to be the preferred detection technique, provided the star-object separation is greater than 1.5 arcsec. Take, for example, the case of a 250 K, self-luminous, Jupiter-size brown dwarf at a distance of 6 pc from a star. Assume, somewhat optimistically, that the telescope optics are at 275 K, with 35% emissivity and 70% system/detector efficiency. The optimal field of view for detection is about $3(\lambda/\Delta)$.[43] In the thermal infrared regime, the signal-to-noise ratio, $P/\sqrt{B}$, is limited by the background due to the atmosphere and the warm op-

tics (P is the total number of detected photons from the object, and B is the total number of background photons in the detector field of view). At 10 μ, the atmosphere is quite transparent, and optics emission is dominant; with 20% fractional bandwidth, the source rate is 7000 photons/sec, whereas the background rate is 6.5 x 10^{10} photons/sec. This implies that a signal-to-noise ratio of 5 could be achieved after about 10 hours of integration. Of course, calibration/referencing procedures to reduce systematic errors to below the fundamental statistical uncertainties must also be included.

Jupiter-like planets, by contrast, are not self-luminous but are in thermal equilibrium with their parent star. To be heated to detectable levels, they must be located within a few AU of this star. As described earlier, this small angular separation makes it difficult to distinguish the planet in the glare of the adjacent star. In this case, the planet and stellar images might be separated by employing both Kecks as an interferometer, using the phase reference imaging techniques first developed at radio wavelengths.

At a distance of 6 pc from the Sun, star-planet separations of a few AU are equivalent to angular separations of about 0.25 to 1.5 arcsec. Resolution much larger than the diffraction limit of a single Keck is necessary. Keck I and II are about 85 meters apart; in the interferometer mode, the achievable resolution is increased by a factor of 8.5 to 24 milliarcsec at 10 μm. For phase-referenced interferometry, observations of the star-planet system are made simultaneously at a short wavelength, where the star contribution dominates, and at a long wavelength, where the planet is relatively more luminous. In both cases, radiation from the system follows the same path through the turbulent atmosphere. Thus, the fringe phase measured at the short wavelength can be used as a reference for that measured at the longer wavelength. At the longer wavelength, the relative phase, $\Delta\phi$, will vary with projected baseline vector b as $\Delta\phi = R\sin(kb\cdot\theta)$, where R is the ratio of intensities for the planet and

Figure 4-13. Sinusoidal modulation of the phase difference in u-v space (quadrants 1 and 2) for a target star-planet system observed from -4 to +4 hours for the noise-free case.

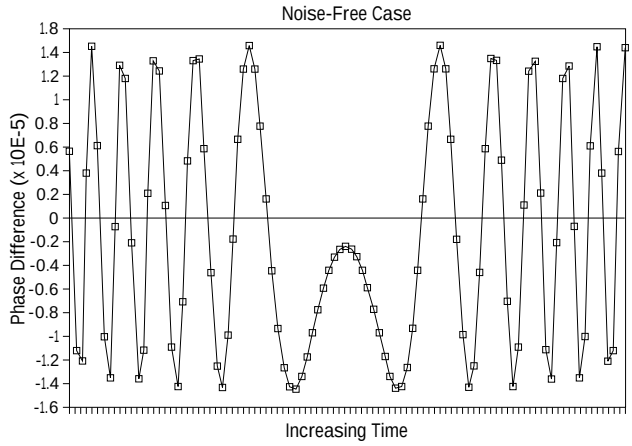


Figure 4-14. Phase variation in a single spectral channel as a function of time as the projected baseline varies due to Earth rotation for the noise-free case.

star, k is wave number, and θ is the angular planet-star separation vector. Such phase measurements are made at a number of wavelengths within the spectral band and at a number of projected baselines, using Earth rotation to foreshorten and rotate the baseline.

The signal-to-noise ratio for phase-referenced detection is a factor of $\sqrt{2}$ worse than that for a filled-aperture photometric detection. However, an interferometer produces synthetic images through observations of an object's Fourier transform and can be effective in distinguishing the presence of a weak companion very close to a bright star. This is illustrated in Figures 4-13 through 4-17 for the simulated observation of a star-planet system at 0° declination, with the planet at 0.25 arcsec from the star, and at 45° azimuth angle. The interferometer elements are separated by 120 meters, and the source is tracked from 4 hours before transit to 4 hours after transit. Figure 4-13 shows the sinusoidal phase modulation expected from a planet-star system in Fourier (u-v)

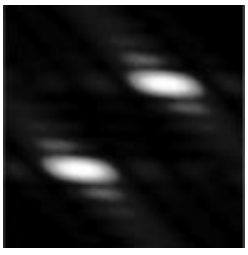


Figure 4-15. Reconstructed planet image for the noise-free case. The simplified processing used here produces a symmetric planet image.

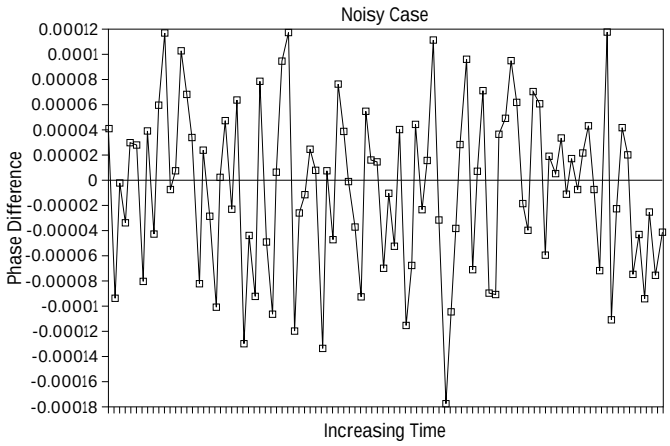


Figure 4-16. Phase variation in a single spectral channel as a function of time when noise is present.

space. It is clear that, although Earth-rotation and wavelength synthesis improve the u-v coverage, only a portion of the u-v plane is observed by a two-element interferometer. The variation of phase in a single spectral channel as the projected baseline changes with Earth rotation is given in Figure 4-14. Figure 4-15 is the Fourier transform of the noise-free data in Figure 4-13; the presence of the planet is evidenced by two enhancements in the pattern. One is merely a symmetric image, resulting from the fact that the phase-referencing method measures only the imaginary part of the Fourier transform. Figures 4-16 and 4-17 are analogous to Figures 4-14 and 4-15, but include noise. The total signal-to-noise ratio, $P/\sqrt{2B}$, is 10, but is built up from 4300 u-v points, each with a signal-to-noise ratio of only 0.15. It is not surprising that no planet pattern can be discerned by eye in the u-v data of Figure 4-16. By contrast, two images are clearly visible in the Fourier transform of Figure 4-17. The planet is unambiguously detected.

Figure 4-17. Reconstructed planet image when noise is present; the total interferometric signal-to-noise ratio is 10.

Nevertheless, the limited u-v coverage of two fixed telescopes constrains their imaging capability to fairly simple systems such as two distinct objects. To study multiple planets and the more complicated protoplanetary disks, many more measurements at different values of u and v are needed. Such observations demand a variety of baselines and hence an array of telescopes. The construction of smaller outrigger telescopes around Keck I and Keck II would provide a straightforward and cost-effective solution.

OUTRIGGER-ENHANCED INTERFEROMETRY

To improve its imaging capability, the present Keck interferometer concept includes four movable outrigger telescopes, each of diameter 1.5 to 2.0 meters. Since for N telescopes, the number of instantaneous baselines is $N(N\text{-}1)/2$, the outriggers form a stand-alone interferometer

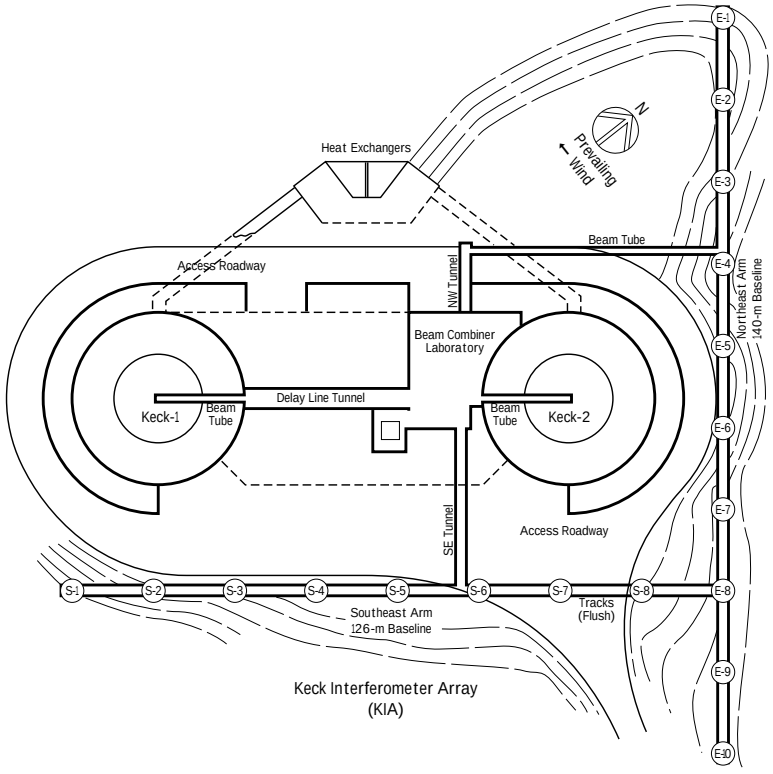


Figure 4-18. Site view of the Keck Interferometer concept, incorporating the two Kecks plus four movable outrigger telescopes.

with six simultaneous baselines. They can also be combined with the two Kecks to improve both light-gathering capability and u-v-plane coverage. The sensitivity will be enhanced on nine of the 15 simultaneous baselines of the six-element Keck-plus-outrigger array, since at least one of the Keck telescopes will contribute to each of these nine baselines.

Figure 4-18 is a site view of the Keck interferometer concept. The four outrigger telescopes can be moved among about 18 fixed stations, providing a variety of baselines with lengths between 12 and 165 meters. Since the incompleteness of the u-v coverage limits the complexity of the images that can be obtained, mobility improves coverage when only a few telescopes are available.

Signals from the outriggers and Keck I and II will be routed to a beam-combining room in the basement of Keck II. Adaptive optics are planned for both Kecks to phase their apertures for use in interferometric combination at wavelengths longer than 2.2 μm. Fast guiding mirrors for the outrigger telescopes should be adequate to phase their smaller apertures. Coarse delay control will be accomplished with long delay lines occupying the length of the coude tunnel. Fine delay control will use active cat's eye delay lines of the type used on other ground-based interferometers. Provision will be made for multiple beam combiners, optimized for different wavelengths or observational goals.

The u-v coverage for several possible modes of the Keck interferometer is displayed in Figure 4-19: (a) reflects the case of the two Kecks used alone; (b) represents the coverage achievable for the Kecks plus the outriggers over one night, and (c) shows the coverage possible if the outriggers are moved to different positions, and the observations are continued on a second night. Greatly increased u-v coverage clearly results from inclusion of the outriggers.

The six-element Keck Array will be capable of imaging a wide variety of extended objects with angular resolution at infrared wavelengths slightly more than an order of magnitude higher than a single Keck. Observations will probably be made at 10 μm initially. The implementation of adaptive optics will extend the array capability to 2.2 μm. Among the most exciting objects that can be investigated are protoplanetary disks. Figure 4-20 shows an imaging simulation for such a disk 150 pc from Earth and at declination +25°. It is assumed that a region of 2 AU radius around the central, solar-type star is clear of material and

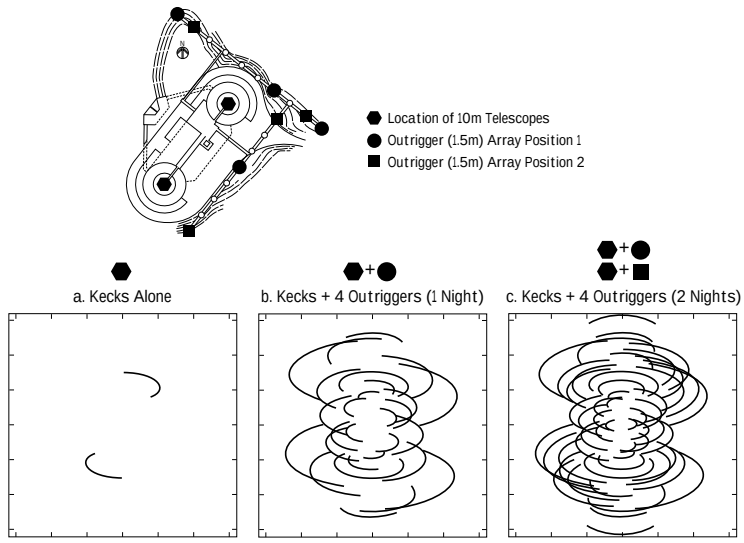


Figure 4-19. U-v coverage available with the Keck Interferometer.

that the temperature, T, of the optically thick disk falls with distance from the star, r, according to $T = 400\ r^{-1/2}$. When the Kecks and outriggers are combined, the radically improved imaging performance and the further reduction in sidelobe energy gained from a second night of observations are dramatic. For debris disks, which are usually 10 times closer than the protoplanetary candidates, the near infrared capabilities of the array will produce even more detailed images.

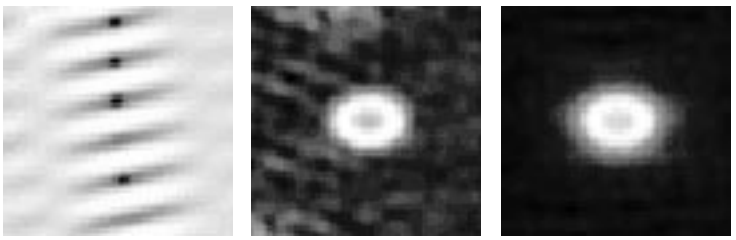


Figure 4-20. Imaging simulations of a protoplanetary disk for the different u-v coverages illustrated in Figure 4-19.

4.6 TOPS-1: ASTROMETRIC IMAGING TELESCOPE

TOPS-1 is the second phase of the program to search for planetary systems around other stars. This phase will exploit space-based observations to broaden and advance the knowledge gained from TOPS-0. Three instruments—the Orbiting Stellar Interferometer (OSI), the Precision Optical Interferometer in Space (POINTS), and the Astrometric Imaging Telescope (AIT)—will be considered as examples. The OSI and POINTS concepts, specifically designed for indirect detection programs, were discussed in Chapter III. The indirect aspects of the AIT concept were also discussed in Chapter III; direct detection applications of AIT are described below.

The Astrometric Imaging Telescope is a proposed 1.5- to 2-meter-diameter, Earth-orbiting observatory designed to incorporate two separate instruments to search for and characterize extrasolar planets and protoplanetary systems. It combines the Coronagraphic Imager and the Astrometric Photometer discussed in Chapter III. The Coronagraphic Imager is specifically designed to image faint objects close to bright sources. The main scientific goals of this instrument are direct imaging of extrasolar planets, understanding of the processes that lead to stellar and planetary formation, and exploration of astrophysical phenomena hidden by bright sources in the field of view. The Astrometric Photometer is an astrometric instrument designed to carry out an exhaustive search for Uranus-mass planets. Together, these instruments will exploit the complementarity between direct and indirect approaches.

The heart of the AIT is a super-smooth optical system; the mid-frequency figure errors are 15 times smoother than the Hubble Space Telescope. At 1 arcsec from a particular star, this figure will give the Coronagraphic Imager a sensitivity hundreds of times greater than that of any existing instrument. The Coronagraphic Imager is equipped with several instruments designed to take advantage of this optical system. Chief among these is a new-design, high-efficiency coronagraph optimized to reduce diffraction sidelobes by several orders of magnitude within a few Airy rings of a bright point source without significant sacrifice of the field of view.

The improvements to be realized from a coronagraph were discussed earlier in this chapter. Essentially, an apodized or graded transmission occulting mask obstructs the central source. Apodization of the focal plane mask results in a significantly more compact pupil plane image,

while allowing some transmission very close to the parent star in the focal plane. The compact pupil plane image renders the application of a Lyot stop more effective at reflecting diffracted light from the system. For targets near the parent star, the coronagraphic efficiency remains virtually independent of radius, even when transmission losses through the mask are taken into account. Consequently, imaging capability is maintained into the central maximum of the Airy pattern.

Theoretical performance predictions for both classical Lyot and high-efficiency coronagraph designs are already confirmed. Figure 4-21 illustrates a laboratory demonstration of diffraction reduction by a factor of 1000. In principle, the coronagraph can work at arbitrarily high background rejection; the limiting background source can always be arranged to be small angle scatter from the optical system. Diffraction reductions below about 10% of the scatter background produce no net gains in system performance.

The overall planetary detection strategy was outlined in Section 4.4 and illustrated in Figure 4-6. The coronagraphic efficiency of the apodized coronagraph can be tuned over a large range, but was shown in Figure 4-6 reducing diffracted starlight by about three orders of magnitude.

The Coronagraphic Imager will implement several additional instrument strategies for capitalizing on the super-smooth optical systems. Various combinations of focal- and pupil-plane masks will allow imaging of faint regions near the limb of a planet or the suppression of far

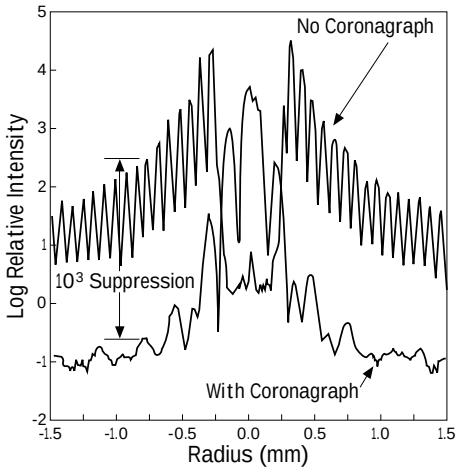


Figure 4-21. Laboratory experiments in diffraction control with the apodized coronagraph have demonstrated significant background suppression. The results shown illustrate one case in which the apodized mask intensity profile has a half power point at three diffraction rings. Thus a source at four or five rings is hardly affected by the coronagraph. The factor of 1000 suppression shown is wholly in agreement with analytical models.

field point spread functions in the crowded fields characteristic of star-forming regions. Tests of a primary mirror 30 cm in diameter indicated that existing metrology techniques are close to being able to characterize mirrors at the required level of smoothness. The integrated figure of this mirror over a 24 cm clear aperture from spatial frequency F (in cycles per aperture) to infinity was 20 Å/$\sqrt{F}$. Subsequent analysis has shown that the intrinsic reduction in mid-frequency figure was actually larger, since a significant portion of the measured mid-frequency power was ringing from low frequency power. It has been demonstrated that the new coronagraph can suppress this ringing, and that the residual figure error meets the flight quality requirements of scatter 1000 times below diffraction.

Systems engineering analysis suggests that low-order aberrations like focus, coma, and astigmatism, characteristic of telescope secondary mirror misalignment, are similarly reduced by the coronagraph. Pointing studies have resulted in a 0.01 arcsec requirement to maintain coronagraphic efficiency. The Coronagraphic Imager constraints on AIT alignment and pointing are modest, consistent with a diffraction-limited telescope of this size, and no more stringent than those achieved by the Hubble Space Telescope. In terms of traditional full aperture figure requirements, the AIT optical system will be about one three hundredth of a wave and will require thousandth wave metrology. In terms of optical fabrication, these are not unrealistic expectations. The technically challenging aspect is to achieve mid-spatial frequency figure requirements.

4.7 TOPS-2: Characterization of Extrasolar Planets

TOPS-2 reflects the very long-range goals of the TOPS program. Observatories in this phase of the project will be flagship missions, intended to revolutionize exoplanetary studies. Whereas TOPS-0 and TOPS-1 aim to identify significant numbers of Jupiter-like planets around other stars, the ambitious objective of TOPS-2 is to discover Earth-type planets that may, like Earth, sustain life. In TOPS-0 and TOPS-1, it is hoped to characterize the orbits of the newly discovered planets. The TOPS-2 goal is to determine more basic planetary properties: the nature of the surface, its temperature, whether there is an atmosphere, and possibly the constituents of that atmosphere.

The concept of a space- or lunar-based instrument designed to detect

radiation from Jupiter-sized planets was introduced more than 10 years ago.[43,44,45,46] Now, avenues of research that may lead to instruments capable of detecting Earth-like planets are opening up. Several workshops—the LOUISA workshop, the Lunar Astrophysics workshop, the New Generation Space Telescope (NGST) workshop—have been held to address the subject. Several approaches are discussed below.

Infrared Detection of Earth-like Planets

Direct detection of Jupiter-size planets may be possible from the ground with 8- to 10-meter-class telescopes. Because of the expected low signal, this is not the case for Earth-sized planets, even if they are at temperatures of 250 K. In Earth orbit or on the Moon, conditions for detecting Earth-like planets are more auspicious, particularly at infrared wavelengths where confusing thermal emission from the atmosphere and the warm telescope optics is a major problem. This effect may be reduced in space by maintaining the telescope optics at very low temperatures. Moreover, the ratio of planet to stellar flux is significantly higher at infrared wavelengths. Nevertheless, at both optical and infrared wavelengths, detection of planets is complicated by the scattered and diffracted radiation from the bright central star, and some form of cancellation is necessary to prevent swamping of the planetary signal. Technological development to achieve the required nulling is progressing. Space- or lunar-based interferometer arrays and filled-aperture telescopes are currently being considered.

Cold Optics

The optics of a ground-based telescope emit about a million times more infrared radiation than the zodiacal dust. If the optics can be maintained at very low temperatures, the system is then background-limited by thermal emission from the zodiacal dust. Several infrared space-based telescopes (e.g., the Infrared Astronomy Satellite (IRAS) and the Cosmic Background Explorer (COBE)) have employed superfluid helium-cooled optics at less than 2 K to lower the infrared background. In both cases, the mission life was limited by the time liquid helium takes to boil away—typically 1 or 2 years.

For a long-duration mission, the optics cannot be cooled by liquid helium. Passive cooling of the optics in high Earth orbit or on the nighttime lunar surface is likely to be more practicable. Both IRAS and

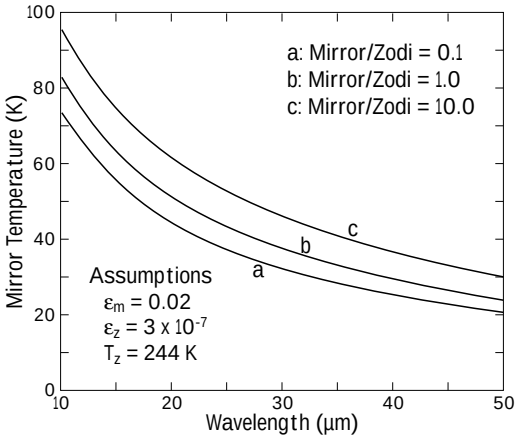


Figure 4-22. Mirror temperature required at various wavelengths to maintain indicated ratios of the thermal radiation from the mirror to the zodiacal intensity.

COBE were designed to keep the optics relatively cool (70 to 100 K) by passive means after the helium reserves had been depleted. Recent studies by ESA and NASA suggest that it may be possible to passively cool large optical systems to temperatures well below 70 K and perhaps as low as 20 K. Figure 4-22 shows calculations of the mirror temperature required at various wavelengths to maintain particular ratios of the thermal radiation from the optics to the zodiacal flux, assuming the zodiacal light is heated to 244 K by solar radiation (i.e., not far from Earth). For optics at 70 K, zodiacal-limited sensitivity will be attained at observing wavelengths shorter than about 10 μm. At longer wavelengths, emission of the optics becomes the limiting factor unless the system is further cooled.

In low Earth orbit, the infrared heat load from Earth itself, as well as the Sun, prevents the implementation of passively cooled 20 K systems. A very high Earth orbit or a site on the Moon is required. For even lower temperatures, an orbit around Jupiter might be envisaged. In this case, solar illumination and the zodiacal component would be considerably lower, reducing the background by several orders of magnitude.

PLANET DETECTION

In the 10 μm region, Jupiter, at 125 K, is approximately the same brightness as Earth, which, although much smaller, is hotter: 275 K. The detectabilities are, therefore, similar. At a distance of 10 parsec, the Earth-Sun separation is only 0.1 arcsec, and the Jupiter-Sun distance is 0.5 arcsec. To resolve the pair, a telescope or interferometer must have

sufficient resolution, 1.22 λ/D for a telescope of diameter D at wavelength λ, or λ/2L for an interferometer where the elements are separated by distance L. In practice, the significant difference in brightness between the star and planet demands over-resolution by factors of four to six. This suggests telescopes with diameters greater than 100 meters, or interferometers with baselines of 40 meters.

High angular resolution is also needed to resolve the planet from the zodiacal dust around the star. The total infrared emission of the zodiacal dust from the solar system is 100 times brighter than emission of Earth. If our solar system were imaged with 0.1 AU resolution, however, Earth's infrared radiation would be comparable to that of the zodiacal light in its vicinity. Over-resolving the star-planet system by a factor of about 10 is therefore necessary to avoid dust-planet confusion. Again, unusually large apertures are necessary.

The collecting area needed to detect an Earth is considerably less than that of a 100-meter telescope. A cryogenic infrared telescope of diameter 25 to 50 meters with a 20% bandpass could detect the emission from an Earth-like planet at 10 pc with signal-to-noise ratio of 5 in about 15 minutes. Several ways of directly detecting planets around other stars are being considered. A space-based, 16-meter telescope operating at infrared wavelengths, for example, would need only a surface accuracy of around 200 Å to reduce the scattered light below the zodiacal level. Various adaptations of the filled aperture concept have been investigated. A large-aperture, infrared telescope using coronagraph techniques combined with rotational shearing interferometry is one possibility.[47] These large telescopes would not have the angular resolution to separate an Earth from the Sun at 10 pc, and are therefore limited to systems much closer or much brighter than 300 K.

Interferometric arrays with large separations could achieve the desired resolution. Baselines of about 250 meters would be required. Such an array could still be quite sparse, with less than 1% of the synthetic aperture filled with glass. The Moon has been suggested as a possible site for this type of instrument. Tables 4-2 and 4-3 contrast the several options, including filled apertures and interferometer concepts.

Much more collecting area is also needed if the goal is spectroscopy of the planetary atmosphere to search for molecules that may signify the presence of life. For a signal-to-noise ratio of 5 in each of 100 spectral channels, the collecting area would have to be increased by a factor of

Table 4-2. Predicted performance for some of the proposed lunar observatories. The interferometers use rotational shear as a nulling mechanism. The indicated performance is for imaging. The interferometers can also be used for astronometric detection. SALSA is an acronym for the submillimeter interferometer.

	Lunar O/IR Interleaf	16 m telescope	SALSA	16m as part of Interferometer
Type	interferometer	telescope	interferometer	interferometer
Baseline	2 km	16m	1 km	2 km
Total collecting area	25 sq meters	200 sq meters	115 sq meters	225 sq meters
Configuration	(12) 1.5m telescopes	(1) 16m telescope	(12) 3m telescopes	(12) 1.5m's, (1) 16m
Optics temp	70 K	70 K	70 K	70 K
Starlight suppress	Nulling interf.	Nulling interf.	not needed	Nulling interf.
Integration time	100 hrs	100 hrs	300 hrs (100μm)	100 hrs
Jupiters	357	172	N/A	1200 (mode 1)
Earths	734	17	N/A	1200 (mode 1)
Protoplanetary disks	yes (1 mas/0.15 au)	yes (125 mas/18 au)	yes (10 mas/1.5 au)	same as w/o 16m

10, to 250-500 square meters (corresponding to a 15- to 20-meter-filled-aperture telescope). The high image modulation of a filled aperture can offer many advantages. The technologies associated with highly dilute systems, such as structural sensing and control systems, can also be applied to large filled apertures. Using thin meniscus substrates and advanced active optical control systems, it is possible to obtain very large controlled optical surfaces with areal densities of only a few kilograms per square meter. These large telescopes will share little technology with the structured or monolithic mirror blanks of today and can be erected in space or on a lunar platform. Even with passive cooling, such telescopes could operate over a broad wavelength range.

Table 4-3. Examples of predicted performance for an IR interferometer and an IR telescope in space. Two-thirds of the candidate stars are double. Early stars are types OBAF; late stars are GKM.

	IR OSI	IBIS
Type	interferometer	telescope
Baseline	100m	16m
Collecting area	(2) 3.7m telescopes	(1) 16m telescope
Optics temp	70 K	70 K
Starlight suppress	Nulling interf.	Nulling interf.
Wavefront correction	single mode filter	Adaptive optics
Integ time/detection	100 hrs	100 hrs
Candidate Jupiters	292	172
Nearby early/late stars	0/292	136/36
Candidate Earths	174	17
Nearby early/late stars	31/143	2/15

THE FUTURE

TOPS-2 can be thought of as being composed of three phases: first, the direct detection of Earth-like planets; second, spectroscopy of the planetary atmospheres; and third, a multi-pixel image of the planet. We can envisage large aperture interferometers cooled passively to 70 K in space or on the lunar surface. One interesting possibility is a hybrid system featuring moderately dilute apertures. Fill factors of 10 to 50% can retain the u-v plane advantages of a filled aperture, while offering resolution gains of factors of 2 to 3 over a monolith. The Moon is a viable base for massive long-lived observing facilities of immense proportions.

Attaining our goals will require consistent and continuing advances and technology development over the next decades. As long as we can build one large telescope, there will always be an incentive to build a second some distance away and phase them. Once two telescopes are phased there will be some impetus to fill in the effective aperture and improve the sensitivity for high-resolution studies. Thus the evolution of filled- and dilute-aperture systems will leapfrog into the future.

Building on such innovative technologies, a far more ambitious stage of exploration of other solar systems should follow. Its justification will necessarily depend upon the results of the TOPS-0, TOPS-1, and TOPS-2 programs that have been the subject of this report. The scale of the undertaking would be enormous, but its goal would be the detailed study of Earth-like planets. If an Earth-like planet is to be imaged as a 10 x 10 pixel image, each pixel will compose only 1% of the infrared emission of the whole planet. The collecting area and the baselines of an interferometric array would have to be expanded one hundred-fold from those discussed here. The TOPS formula aims to ensure that this gigantic step is taken from a firm foundation. The endeavor is not beyond the reach of technology, and even now the general outlines of the project can be drawn.

The universality of this undertaking makes it a compelling project for the international scientific community, and it would be especially appropriate for peoples of our world to work together in the quest for knowledge of other, more distant worlds. The goal of reaching out to other Earths and perhaps other life is within our reach; the first steps can be taken now.

REFERENCES

1. KenKnight, R. G. 1978, *Icarus*, 36, 422.
2. Bracewell, R. N. & MacPhie, R. H. 1979, *Icarus*, 38, 136.
3. Angel, R., Cheng, A., & Woolf, N. 1986, *Nature*, 322, 341.
4. Ibid. *Nature*, 324, 518.
5. Burke, B. F. 1986, *Nature*, 322, 340.
6. Ibid. *Nature*, 324, 518.
7. Hubbard, W. B. 1984, in *Planetary Interiors*, (Van Nostrand Reinhold: New York).
8. Stevenson, D. J. 1991, "The Search for Brown Dwarfs," *Ann. Rev. Astron. Astrophys.*, 61, 531-568.
9. Cloud, P. 1974, *Am. Scient.*, 62, 54-66.
10. Angel, R. 1989, in *The Next Generation Space Telescope*, eds. P. Bely, C. Burrows, and G. Illingworth (STScI: Baltimore), p 81.
11. Owen, T. 1980, in *Strategies for Search for Life in the Universe*, ed. M. D. Papagiannis (Reidel: Dordrecht), 177.
12. Shu, F. H., Adams, F. C., and Lizano, S. 1987, *ARA&A*, 25, 23.
13. Strom, K. M., Strom, S. E., Edwards, S., Cabrit, S., & Skrutskie, M. F. 1989, *AJ*, 97, 1451.
14. Cohen, M., Emerson, J. P., and Beichman, C. A. 1989, *ApJ*, 339, 455.
15. Beckwith, S. V. W., Sargent, A. I., Chini, R., & Gusten, R. 1990, *AJ*, 99, 924.
16. Safronov, V. S., and Ruzmaikina, T. V. 1985, in *Protostars & Planets II*, eds. D. C. Black and M. S. Matthews (Tucson: University of Arizona Press), 95.
17. Sargent, A. I., and Beckwith, S. 1987, *ApJ*, 323, 294.
18. Sargent, A. I., and Beckwith, S. 1991, *ApJ*, 382, L31.
19. Beckwith, S. and Sargent, A. I., 1991, *ApJ*, 381, 250.
20. Meakin, P. & Donn, B. 1988, *ApJ*, 329, L39.
21. Strom, S. E., Edwards, S., & Skrutskie, M. F. 1992, in *Protostars & Planets III*, ed. E. H. Levy & J. Lunine (Tucson: University of Arizona Press), in press.
22. Koerner, D., Sargent, A. I., Beckwith, S. & Padin, 1992, in preparation.
23. Aumann, H. H., Gillett, F. C., Beichman, C. A., de Jong, T., Houck, J. R., Low, F. J., Neugebauer, G., Walker, R. G., & Wesselius, P. R. 1984, *ApJ*, 278, L23.
24. Backman, D. E., & Gillett, F. C. 1987, in *Cool Stars, Stellar Systems & the Sun*, ed. J. Linsky & R. E. Stencel (Berlin: Springer-Verlag), p 340.
25. Backman, D. E., & Paresce, F. 1992, in *Protostars and Planets III*, ed. E. Levy & J. Lunine (Tucson: Univ. of Arizona Press), in press.

26. Smith, B. A., & Terrile, R. 1984, *Science*, 226, 1421.

27. Boggess, A., Bruhweiler, F.C., Grady, C.A., Ebbets, D.C., Kondo, Y., Trafton, L.M., Brandt, J.C., & Heap, S. 1991, *Ap J*, 377, L49.

28. Lagrange-Henri, A. M., Beust, H., Vidal-Madjar, A., & Ferlet, A. 1989, *A&A*, 215, L5.

29. Telesco, C. M., & Knacke, R.F. 1991, *ApJ*, 372, L29.

30. Becklin, E. E., & Zuckerman, B. 1990, in "Submillimetre Astronomy" ed. G. D. Watt & A. D. Webster (Dorecht: Kluwer), p 147.

31. Norman, C., & Paresce, F. 1987, in *The Formation and Evolution of Planetary Systems*, ed. L. Danly and H. Weaver (Cambridge: University of Cambridge Press), p 151.

32. Bracewell, R. N., 1978 in *The Fourier Transform and its Applications*, (McGraw-Hill: New York).

33. Born, M. and Wolf, E. 1980, *Principles of Optics*, 6th edition, (Pergamon: Oxford), p 461.

34. Kuiper, G. P. 1953, in *The Sun*, (University of Chicago Press: Chicago).

35. Bracewell, R. N. 1978, *Nature*, 274, 780.

36. Napier, P. J., Thompson, A. R. & Ekers, R. D. 1983, *Proc IEEE*, 71, 1295.

37. Shao, M., Colavita, M. M., Staelin, D. H., Simm, R., & Johnston, K. J. 1984, in *IAU Symposium No. 109 on Astrometric Techniques*.

38. Goodman, J. W. 1985, in *Statistical Optics*, (New York: John Wiley and Sons), Chapter 8.

39. Roddier, F. 1981, in *Progress in Optics*, ed. E. Wolf, (New York: North-Holland), Chapter V.

40. Tatarski, V. I., 1961, *Wave Propagation in a Turbulent Medium*, translated from the Russian by R. A. Silverman, (New York: McGraw Hill).

41. Fried, D. L. 1964, *J. Opt. Soc. Amer.*, 56, 1372.

42. Noll, R. J. 1976, *J. Opt. Soc. Amer.*, 66, 207.

43. Roddier, F., Northcott, M., and Graves, J. E. 1991, *Publ. Astr. Soc. Pacific*, 103, 131.

44. "Lunar Bases and Space Activities of the 21st Century," 1985, ed. W. W. Mendell (LPI: Houston).

45. "Future Astronomical Observatories on the Moon," 1988, ed. J. O. Burns & W. W. Mendell, NASA Conf. Publ. 2489.

46. Burke, B. F. 1987, in I.A.U. Coloquium 99 "Bioastronomy—the Next Steps," ed. George Marx (Reidel: Dordrecht), p 139.

47. Diner, D. J., Tubbs, E. F., Gaiser, S. L., & Korechoff, R. P. 1991, in *IBIS*, 44, 505.

48. Shao, M. 1991, *Proc. SPIE*, 1494, 347.

Editor-in-Chief:	Bernard F. Burke, *Massachusetts Institute of Technology*
Managing Editor:	Robert A. Brown, *Space Telescope Science Institute*
Senior Technical Editor:	Terri Ramlose, *Science Applications International Corporation*
Designer:	Carl A. Schuetz, *Foxglove Communications*
Technical Editors:	Alan P. Boss, *Carnegie Institution of Washington*
	Robert A. Brown, *Space Telescope Science Institute*
	Eugene H. Levy, *University of Arizona*
	Stanton J. Peale, *University of California, Santa Barbara*
	Anneila I. Sargent, *California Institute of Technology*
Picture credits:	*Courtesy of California Association for Research in Astronomy*
Frontispiece:	*Courtesy of* Charles Lada, *Harvard College Observatory*
Cover Art:	Robert Priest, *Massachusetts Institute of Technology*