Deep Space Optical Communications

Hamid Hemmati, Editor

Jet Propulsion Laboratory
California Institute of Technology

DEEP SPACE COMMUNICATIONS AND NAVIGATION SERIES

DEEP SPACE COMMUNICATIONS AND NAVIGATION SERIES

Issued by the Deep Space Communications and Navigation Systems
Center of Excellence
Jet Propulsion Laboratory
California Institute of Technology

Joseph H. Yuen, Editor-in-Chief

Published Titles in this Series

Radiometric Tracking Techniques for Deep-Space Navigation
C. L. Thornton and J. S. Border

Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation
Theodore D. Moyer

Bandwidth-Efficient Digital Modulation with Application to Deep-Space Communications
Marvin K. Simon

Large Antennas of the Deep Space Network
William A. Imbriale

Antenna Arraying techniques in the Deep Space Network
David H. Rogstad, Alexander Mileant, and Timothy T. Pham

Radio Occultations Using Earth Satellites: A Wave Theory Treatment
William G. Melbourne

Deep Space Optical Communications
Hamid Hemmati

Deep Space Optical Communications

Hamid Hemmati, Editor

Jet Propulsion Laboratory
California Institute of Technology

DEEP SPACE COMMUNICATIONS AND NAVIGATION SERIES

Deep Space Optical Communications

October 2005

The research described in this publication was carried out at the
Jet Propulsion Laboratory, California Institute of Technology, under
a contract with the National Aeronautics and Space Administration.

Table of Contents

Foreword

The Deep Space Communications and Navigation Systems Center of Excellence (DESCANSO) was established in 1998 by the National Aeronautics and Space Administration (NASA) at the California Institute of Technology's Jet Propulsion Laboratory (JPL). DESCANSO is chartered to harness and promote excellence and innovation to meet the communications and navigation needs of future deep-space exploration.

DESCANSO's vision is to achieve continuous communications and precise navigation—any time, anywhere. In support of that vision, DESCANSO aims to seek out and advocate new concepts, systems, and technologies; foster key technical talents; and sponsor seminars, workshops, and symposia to facilitate interaction and idea exchange.

The Deep Space Communications and Navigation Series, authored by scientists and engineers with many years of experience in their respective fields, lays a foundation for innovation by communicating state-of-the-art knowledge in key technologies. The series also captures fundamental principles and practices developed during decades of deep-space exploration at JPL. In addition, it celebrates successes and imparts lessons learned. Finally, the series will serve to guide a new generation of scientists and engineers.

Joseph H. Yuen
DESCANSO Leader

Preface

The ever-increasing demand for data from planetary probe spacecraft is pushing the frequency of telecommunications from radio frequency (RF) bands to the optical and near-infrared regime. Such a transition offers the potential to increase data rates by one to two orders of magnitude over conventional RF links. Early NASA spacecraft telecom systems relied on the S-band frequency. Nearly twenty years later, X-band frequencies were implemented. Over twenty years later, the Ka-band systems are beginning to be implemented in deep space. For the optical band, we are now in the technology maturation and demonstration phase. It is expected that after a number of successful and convincing technology validation demonstrations, the optical band will also move into the implementation phase.

This reference text is intended to summarize and document the optical work performed at the Jet Propulsion Laboratory (JPL) since inception of the Free-Space Optical Communication Group in late 1970s. This text provides an overview of nearly a quarter of century of research and development, performed by JPL's Optical Communication Group, its associated researchers, and other optical-communications researchers throughout the world. The focus of the research effort has been deep space telecommunications. In recent years, the near-Earth communication technologies have been addressed also. The flight transceiver, the ground receiver, and uplink transmitter technologies were addressed.

During the past 25 years, the focus of the component and subsystem technology efforts had to be adjusted frequently to keep pace with the rapid developments in laser, detector, detector array, and fiber-optic technologies. Therefore, a significant portion of the group's effort was concentrated on addressing this challenge. This book is intended to bring a novice in the field up to date, and be informative to those interested in learning about the status of

optical communications technology. As a reference book it should help the people in the field to build upon the prior knowledge and become aware of the important design variations and critical differences between them. Also, this book is intended to provide information on the state-of-the-art in component and subsystem technologies, fundamental limitations, and approaches to reach and fully exploit new technologies.

The text is organized into seven chapters in which Chapter 1 provides an overview of deep-space optical communications technology and a historical perspective of deep-space optical communications technology developments by JPL. Chapter 2 discusses the link and the system design drivers. Parameters that influence the design of an optical communications systems and the link control table that takes into all relevant link parameters are discussed here. The atmospheric channel is discussed in Chapter 3. Cloud statistics, atmospheric transmission, background light and sky radiance, laser beam propagation through the turbulent atmosphere and atmospheric issues driving the selection of a ground receiver site are discussed in this chapter. Chapter 4 deals with modulation and coding, including the statistical models for the detected optical fields, modulation formats, rate limits imposed by constraints of modulation, performance of uncoded optical modulation schemes, optical channel capacity, channel codes for optical modulations, and performance of optical modulations. Chapter 5 deals with the subsystems that constitute the flight terminal. Subchapter 5.1 is on acquisition, tracking and pointing. The most challenging aspect of deep-space Optical Communication technology has been and remains as the tracking and pointing function. This subchapter deals with precise beam pointing throughout the Solar System, options, design drivers and requirements, and examples of system implementation. Subchapter 5.2 deals with the laser transmitter. Flight laser transmitters continue to be a major risk item due to current less-than desired lifetime. Requirements, wavelength effects, candidate sources, modulators, laser efficiency, timing jitter, and thermal management are discussed in this subchapter. The opto-mechanical subassembly including a description of general requirements, the optical channels, design approaches, transmit/receive isolation, stray light control, structure materials, and optical design examples are described in Subchapter 5.3. Flight qualification of lasers and detectors, including environmental requirements, flight qualification approaches and procedures are described in Subchapter 5.4. Chapter 6 discusses the Earth-based terminal architecture. Single-station downlink reception and uplink transmission are discussed in Section 6.1.1. Options and approaches, site diversity, receiver stations located above clouds (e.g., balloons, airplanes, or spacecraft) uplink beacon, safe laser beam propagation, and atmospheric effect mitigation are among the topics discussed in this section. Section 6.1.2 discusses arraying of telescope receivers, including trades, implementation schemes, and performance analysis. Subchapter 6.2 discusses photodetectors, including both single element (6.2.1) and array of photodetectors (6.2.2).

Requirements and challenges, a description of photon-counting detectors, implementation options and performance are discussed here. Subchapter 6.3 discusses receiver electronics, including demodulator architectures, synchronization and post-detection filtering, demodulator variations, and system models and architectures. Chapter 7 discusses future prospects and applications, including certain technology developments to date, navigational tracking, and light science.

Hamid Hemmati,
Pasadena, California
October 2005

Acknowledgments

Numerous individuals have contributed to the work summarized in this book. The Optical Communications Group at JPL was the brainchild of James Lesh who diligently and tirelessly worked on formulating meaningful technology development activities and promotion of this technology within NASA, industry, and other government organizations. Joseph Yuen had the vision for producing this book and relentlessly reminded the authors of its importance and the need for quick publication.

Many individuals have contributed significantly to the technical accomplishments reported here. Among them are Joseph Katz, Don Sipes, William Marshall, Marc Rayman, Deborah Robinson, David Erickson, Kelley Cowles, Edward Kerr, Bonny Schumaker, Joseph Kovalik, Janet Wu, John Sundusky, Norman Page, Stephen Monacos, Homayoon Ansari, Juan Ceniceros, Jeffrey Charles, Vachik Garkanian, Carlos Esproles, and Jerry Neal. Very effective technical leadership of Richard Matheson, William Weber, William Rafferty, Alaudin Bhanji, Kent Kellogg, Leslie Deutsch, Joseph Yuen, Stephen Townes, Fabrizio Pollara, Stephen Lichten, Ben Parvin, MiMi Aung, Samuel Zingales, and many others who have been crucial to the successful accomplishment of the tasks described here

The work reported here could not have come about without the significant level of NASA funding over the years through such visionary scientists at NASA Headquarters as Ramon DePaula and Barry Geldzahler.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA).

October 2005

Contributors

All contributors are with the Jet Propulsion Laboratory, California Institute of Technology (Caltech), Pasadena, California, USA.

Abhijit Biswas received his PhD from Southern Illinois University (Carbondale, Illinois) in molecular science in 1986. He has participated in optical communications technology development at JPL since 1992. Most recently Dr. Biswas served as the ground network systems engineer for the Mars Laser Communication Demonstration Project.

Chien-Chung Chen received his PhD from University of Illinois at Urbana-Champaign in 1987, where his dissertation was on free space optical communications. He joined JPL in 1987, and he has worked on both optical communications and deep space radio frequency (RF) systems development. He has extensive experience in both the development and operations of the deep space communications links. Dr. Chen is the principal investigator for the JPL Mars Laser Communications Demonstration Project.

Samuel J. Dolinar received his PhD in electrical engineering from the Massachusetts Institute of Technology (Cambridge, Massachusetts) in 1976, where his master's and doctoral theses were on optical communications. He worked at MIT Lincoln Laboratory, before joining JPL in 1980. Dr. Dolinar has focused his research on channel coding and source coding for the deep-space channel, especially turbo codes and low-density parity-check (LDPC) codes during the past decade. He teaches data compression at Caltech.

William H. Farr attended Caltech from 1976 through 1980, and has traversed a variety of fields including chemistry, neurobiology, electrical

engineering, and computer science, resulting in publications in seven fields. Prior to joining JPL in 2001, he worked for several engineering and high tech firms, including the Nucleonics Development Corporation, where he was head of the research and development department developing radioisotope and optical instrumentation for industrial monitoring applications, and had two patents. Mr. Farr is now the manager of JPL's Optical Communications Technology Program of the Interplanetary Network Directorate.

Andrew A. Gray received his PhD in electrical engineering in 2000, and his master of business administration (MBA) in 2004, both from the University of Southern California (Los Angeles). Prior to joining JPL in 1998, he worked at the NASA Goddard Space Fight Center for three years. He is a group supervisor in the Communications Architecture and Research Section. The primary focus of the group is development of first-to-the world prototypes for communications and radar systems. He holds three patents. Dr. Gray is also an affiliate faculty member at the University of Washington (Seattle, Washington).

Jon Hamkins received his BS from Caltech in 1990, and PhD from the University of Illinois at Urbana-Champaign in 1996, both in electrical engineering. Dr. Hamkins has been at JPL since 1996, where he is the supervisor of the Information Processing Group, which performs research in optical communications, information theory, channel coding, data compression, and synchronization.

Hamid Hemmati received his MS in physics from the University of Southern California, and his PhD in physics from Colorado State University (Fort Collins, Colorado) in 1981. Prior to joining JPL in 1986, he worked at the NASA Goddard Space Flight Center and at the National Institute of Science and Technology (NIST, Boulder, Colorado) as a researcher. He is now the supervisor of the JPL Optical Communications Group, which is developing laser-communications technologies and systems for deep space and satellite communications. Dr. Hemmati holds seven patents. He has taught optical communications courses at the University of California at Los Angeles (UCLA) Extension.

Chi-Wung Lau received his BS in physics from the University of California at Berkeley in 1996 and his MS in electrical engineering from the University of Southern California in 2001. He has been with JPL since 1996 involved with such projects as Deep Impact, the optical array receiver, and the telecom forecaster predictor tool. Mr. Lau is currently working on applying quantum theory to communications.

Shinhak Lee received his PhD from the University of Washington in electrical engineering in 1997, and he has been with JPL since then. He has made contributions to the acquisition, tracking, and pointing technology. Dr. Lee is a member of technical staff of the Optical Communications Group in the Communications Research Section.

James R. Lesh received his PhD from the University of California at Los Angeles in electrical engineering in 1976, and he has been with JPL since 1971. He has held numerous technical and managerial positions, including head of the Optical Communications Program, and currently he is the chief technologist and manager of the Technology Program of the Interplanetary Network Directorate. He holds three patents, and he has taught classes in communications theory, information theory, channel coding, and signal processing at Caltech. Dr. Lesh is a fellow of the Institute of Electrical and Electronics Engineers (IEEE) and of the International Society for Optical Engineering (SPIE).

Bruce E. Moision received his PhD from the University of California at San Diego in electrical engineering in 1999, and has been with JPL since 2000. He has worked primarily on the design and implementation of error correction codes and modulation schemes for optical communications links. Dr. Moision is a member of the Information Processing Group of the Communications Research Section.

Gerardo G. Ortiz received his PhD from the University of New Mexico in opto-electronic engineering in 1997, and he has been with JPL since 1987. He has made contributions to the development of high electron mobility transistor (HEMT) ultra low noise amplifiers for deep space radio frequency communications, multiple wavelength vertical cavity surface-emitting laser arrays for backbone networks, and acquisition, tracking, and pointing (ATP) technologies for free-space optical communications. Dr. Ortiz is a senior member of the staff in the Optical Communications Group.

Sabino Piazzolla received his PhD in electrical engineering from the University of Southern California in 1997. He has been at JPL since 2004, focusing on optical communications. Dr. Piazzolla is also a part-time faculty member at the University of California at Los Angeles, and at the University of Southern California.

William T. Roberts received his PhD in optical sciences from the University of Arizona in 2001. He has been with JPL since 2001. Dr. Roberts has focused his effort on development of deep-space communication lasers, flight qualification of communication terminal parts, and the conversion of

large astronomical telescopes to perform as deep-space optical communication receivers.

Meera Srinivasan received her BS from Caltech in 1990, and her PhD from the University of Illinois at Urbana-Champaign in 1996, in electrical engineering. Dr. Srinivasan has been with JPL since 1996. Her research interests lie in the areas of optical communications, wireless and spread-spectrum communications systems, array signal processing, and detection and estimation theory.

Victor A. Vilnrotter received his PhD from the University of Southern California in electrical engineering in 1978, specializing in optical communications. Dr. Vilnrotter has been at JPL since 1979, and has conducted research in the application of optical and quantum communications to the deep space optical channel, electronic compensation for deep-space RF antennas via focal-plane signal processing, optical focal-plane detector arrays to mitigate atmospheric turbulence effects, and development and demonstration of fundamental optical array receiver concepts.

Keith E. Wilson received his PhD from the University of Southern California in 1980. Before joining JPL in 1988, he worked in laser research at Hughes Research Laboratories, Allied Corporation, and Litton Guidance and Control. He was a faculty member in physics at California State Polytechnic University at Pomona. Dr. Wilson has managed JPL's successful optical communications demonstrations with spacecraft. He is a principal engineer in the Communications Architectures and Research Section at JPL.

Malcolm W. Wright received his PhD in physics from the University of New Mexico (Albuquerque, New Mexico) in 1992, and was with the Air Force Research Laboratory researching high power lasers before coming to JPL in 1998. Dr. Wright's current work has focused on developing high power fiber lasers for downlink and uplink, space qualification of semiconductor and fiber based lasers and communication performance of various free space optical link demonstrations in the lab and the field.

Chapter 1
Introduction

James R. Lesh

1.1 Motivation for Increased Communications

This monograph represents the collective knowledge and experience of more than 25 years of concentrated research and development effort by a dedicated team of talented technologists at the Jet Propulsion Laboratory (JPL). The work began in the late 1970s and continues today. The vision was an optical communications capability that provides orders-of-magnitude more data return from deep-space missions than is possible with conventional radio frequency (RF) techniques, and the dream has been to see that capability being matured, demonstrated, and used operationally within the professional careers of those who contributed to its earliest analytical and experimental developments.

Communication over deep-space distances is extremely difficult. Communications beams spread as the square of the distance between the transmitter and the receiver. As the distance increases, the difficulty becomes quadratically more difficult. For example, conventional satellite communication from Earth orbit often uses satellites in geosynchronous Earth orbit (GEO) to communicate with the ground. The GEO altitude is approximately 40,000 kilometers (km). From such a distance, quite high data rates in the gigabits per second (Gbps) can be established and maintained. However, the distance from Earth to Neptune or Pluto can be on the order of four billion (4,000,000,000) km. After propagating over such a distance, the communications beam from a spacecraft will spread to an area 10 billion times larger in area than if the beam from the same system traveled from just the GEO distance. The weakened

beam would make communications with the Earth 10 billion times more difficult. Stated differently, a system capable of transmitting 10 Gbps from GEO to the ground would only achieve 1 bit per second (bps) from nominal Pluto/Neptune distances.

One could, of course increase the capabilities of the distant spacecraft's communications system, as well as improve the sensitivities of the Earth reception systems. Indeed, both of these approaches are used for present-day deep-space missions. The net effect has been to raise the nominal data rates from Mars distances to the range of tens to hundreds of kilobits per second (kbps), with correspondingly lower data rates for the outer planets. But further increases are hard to accommodate. Current missions are already flying antennas that are difficult to squeeze into protective launch shrouds, and increases in transmitter power are discouraged due to the difficulties of both generating electrical power at far solar distances as well as removing the waste heat resulting from the corresponding inefficiencies of the various transmitter energy conversion components. On the Earth end, increasing sensitivity is likewise difficult. Current National Aeronautics and Space Administration (NASA) Deep Space Network (DSN) antennas are already enormous (34-m and 70-m diameters), and the receiving system low-noise amplifiers are already operating at but a few degrees above absolute zero. More advances in conventional communications capabilities are planned, and even larger improvements are being researched for future consideration, but practical realities will eventually limit the degree to which such improvements can be made.

As an example, consider the Mars Global Surveyor (MGS) mission that was, and continues to be, an outstanding success mapping features of the Martian terrain. During the entire prime mission phase, the project was only able to map 0.3 percent of the Martian surface at high resolution. More has been mapped during the extended mission phase, but even with this extension, the mission will produce high-resolution maps of only a few percent of the surface. This coverage has been limited by the capabilities of the communications system that was affordable at the time the mission was defined and developed.

Although conventional capabilities will likely rise in the future, so will the needs for even higher instrument data volumes. Most of the planets have had initial flyby pathfinder missions, and a few have had initial-characterization orbiters. However, the spatial and spectral sensitivities of those instruments have been very limited by the data-return capabilities and are orders-of-magnitude below what scientists are doing for Earth observations today. Figure 1-1 shows these future needs. The horizontal axis is the data rate, and the vertical line near the left side is the MGS capability when scaled to Saturn distance. The vertical dimension has no meaning other than to show that things

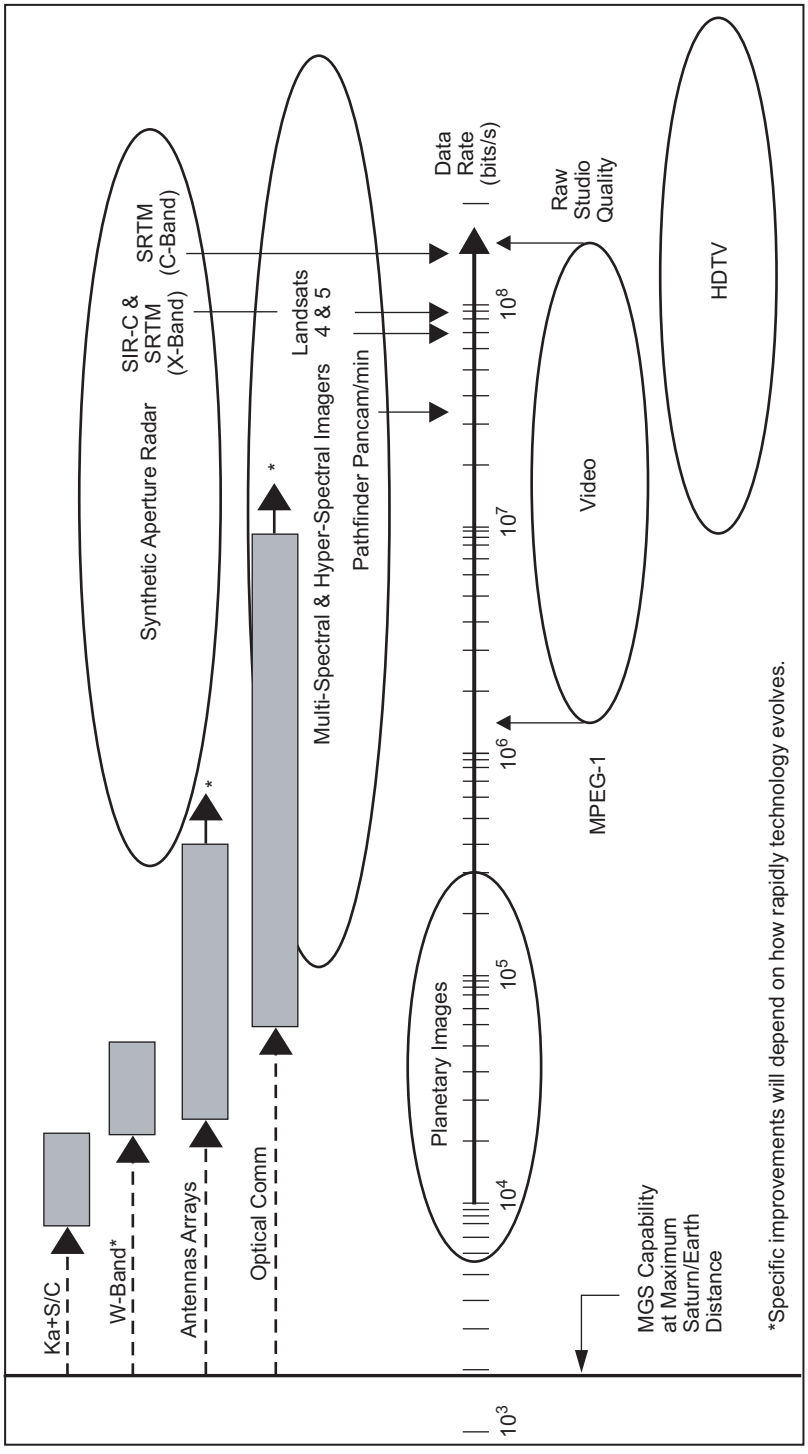


Fig. 1-1. Future data return needs relative to current capabilities based on Mars Global Surveyor at maximum Saturn range.

above the central data-rate-axis arrow are representative of scientific investigation needs, whereas those below just provide a rough measure of telecommunications needs for enhanced public engagement. The ovals represent horizontal data rate regions where corresponding instruments are expected to operate. Regions of anticipated capability improvements are shown for several candidate communications technologies. Technologies ultimately chosen and how far to the right those improvement bars can be extended depend on current and planned technical research and system designs, as well as thorough life-cycle-cost analyses. However, the anticipated performance capability improvements of optical communications are clearly evident.

The promise of improvement comes, to first order, from the much higher frequencies of the optical signals. Over the history of the DSN, conventional RF performance has improved about 12 orders-of-magnitude due to significant and sustained research and development (R&D) efforts at JPL. Improvements have come from many technological advances. However, the biggest improvements were achieved when the operating carrier frequency of the communications signal was increased. Currently, the primary frequency used for deep-space communications is X-band (approximately 8 GHz), although new missions will soon be transitioning to Ka-band (32 GHz). The change from X-band to Ka-band has a theoretical improvement (due to frequency-squared) of 11.6 dB, although practical factors (e.g., atmospheric losses) have limited that improvement to about 6 dB. The promise of optical communications is much more since the frequency is very much higher (approximately 300,000 GHz). Although practical factors (e.g., atmospheric losses, receiver sensitivities) will also be present, they are more than offset by the frequency-squared benefit of the higher carrier frequency.

Figure 1-2 diagrams the much lesser beam spread offered by optical transmission. The left side of the figure shows the transmitted beam sent back toward the Earth from the Voyager spacecraft. The transmitting antenna is 3.7 m in diameter (a dominant architectural feature of the spacecraft), and the transmitted frequency is X-band. By the time the beam reaches Earth from Saturn, diffraction (a fundamental property of all transmitted electromagnetic beams) has caused the signal to spread out over an area 1000 Earth-diameters wide. Contrast this with the right-hand side of the figure where the beam from a small (10-cm) optical telescope is transmitted back to the Earth. Assuming an optical wavelength of 1 μm (frequency of 3×10^{14} Hz), the resulting spot size at the Earth is only one Earth diameter wide. That represents a factor of 1000 concentration of the received energy in both the horizontal and vertical directions (factor of 10^6 in power density), and that is achieved with a very much smaller transmitting antenna (0.1 m versus 3.7 m) on the spacecraft. The wavelength-squared advantage over X-band is approximately 90 dB, although quantum effects and practical implementation considerations limit current realistic gains to about 60 dB.

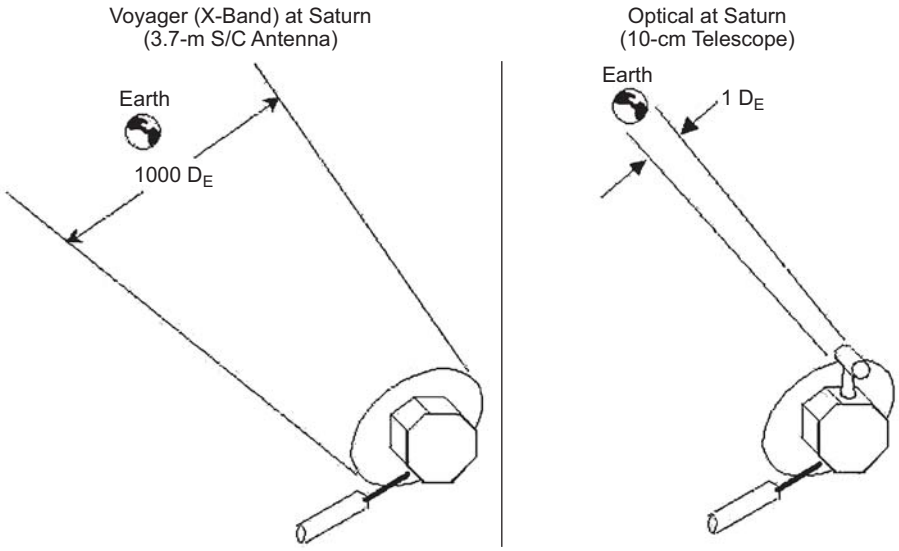


Fig. 1-2. Comparison of RF and optical beam spreads from Saturn.

1.2 History of JPL Optical Communications Activities

JPL began considering optical communications in the late 1970s. Dr. John R. Pierce was the JPL Chief Technologist at the time and had been looking into some interesting attributes of quantum detection theory. In 1978, he wrote a paper predicting that multiple-bit-per-photon optical communications appeared to be possible [2]. Unlike conventional RF communications that used phase modulation of the carrier, Pierce suggested using direct photon detection with a high-alphabet pulse-position modulation (PPM). PPM modulation uses a time interval that is divided into a number of possible pulse locations, but only a single pulse is placed in one of the possible positions. The position of that pulse is determined by the information (word) that is to be transmitted. Given the experience JPL had in deep-space communications, the prospect of extremely power-efficient communications looked very attractive. However, to realize the potential of multiple-bit-per-photon optical communication, it would be necessary to use codes that were efficient at filling in channel erasures. The model for the optical channel under these circumstances was a pure erasure channel where the dominant error source was the quantum uncertainty of the signal itself, and this resulted in pulses for which the weak received photon field contained inadequate probability to reliably constitute detection of a received photon. The laws of quantum mechanics dictated how often these pulses would be "erased." Thus, contextual information (i.e., codes) that bridged these erasure events would be required.

At the time, JPL was routinely using Reed–Solomon (RS) codes in concatenation with convolutional codes. The RS codes were very efficient at correcting the bursts of errors that would result when the Viterbi decoder that was decoding the convolutional code made an erroneous branch path decision. Since the loss of a single PPM optical pulse detection caused a similar burst of errors (loss of a pulse meant that the data bits associated with that pulse were chosen at random), it appeared that such codes were well matched to the PPM channel.

To prove this out, a technology task was started to demonstrate multiple-bit/detected-photon communications in the laboratory. At the time, the heterodyne detection "quantum limit" was understood to be 1 nat/photon (1 nat = 1.44 bits), and this limit was believed to apply to all optical communications systems. However, Pierce in [2] was predicting that much higher photon efficiencies could be achieved. Accordingly, the objective selected for the technology task was 2.5 bits/detected photon, a comfortable margin beyond the classical quantum limit.

Figure 1-3 shows the experimental setup that resulted. Inside a light-controlled enclosure was a semiconductor laser diode, some calibration detectors, calibrated optical attenuators, and a cooled photo-multiplier tube (PMT) detector. The laser diode was driven by the signal from a PPM modulator.

The output of the PMT detector was integrated over the time intervals of each of the possible pulse locations (slots) and compared with a decision threshold. Integrals exceeding the threshold were declared as detected pulses, whereas those that did not exceed the threshold were declared to be non-pulse locations. Timing synchronization was hard wired from the PPM modulator to

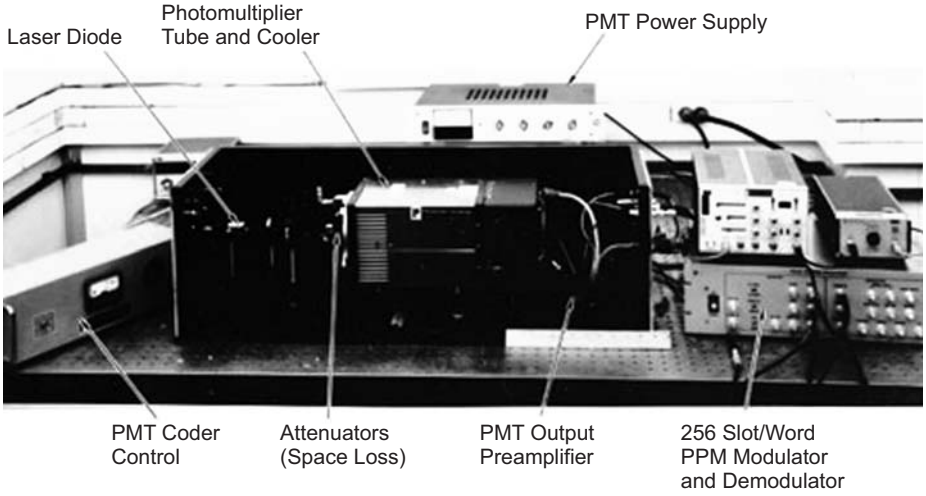


Fig. 1-3. Photograph of the 2.5 bit/detected photon demonstration.

the detector circuit so that timing errors were eliminated from the demonstration. The RS code and associated decoder were not implemented in hardware but were simulated on a computer. Data from the PPM detection process were fed into the computer. Since these data represented hard slot decisions within each PPM word, the performance that an actual RS decoder would have provided could be easily calculated. The resulting measurements demonstrated that reasonable error rates (below 10^{-6}) could be achieved at signaling efficiencies up to 3 bits/detected photon [3–9]. This demonstrated clearly that the classical "quantum limit" did not apply for this channel. There was some concern that it might be difficult to establish timing synchronization for such a link. A subsequent demonstration effort showed that the hard-wired synchronization assumption could be eliminated, at least for signals represented in this demonstration [10,11].

The interest generated by [2], and the resulting multi-bit/photon demonstration, sparked a flurry of theoretical studies to better understand the optical channel and the techniques that could be used to exploit it [12–15]. These studies considered channel capacities and computational cut-off rates for the optical channels [16–23], optimal modulations for achieving higher energy efficiencies [24–27], and codes that were well matched to those modulations [28–30]. But two things were becoming very clear. The first was that the complexity of the systems required to exceed approximately 3 bits/detected photon would increase rapidly. Indeed, Butman, Katz, and Lesh [31,32] showed that the uncertainty principle alone would limit systems to below 20 bits/photon, and that practical timing limits would hold that number even lower. Furthermore, McEliece, Rodemich, and Rubin [33] showed that, based on computational cut-off rate arguments, it would be very difficult to exceed 10 bits/photon. The second realization was that the real challenges to the utilization of optical communications were not in squeezing more out of the modulation and coding efficiencies, but in the sizes, weights, powers, and reliabilities of many of the required optical-communication component and subsystem technologies. As a result, JPL research efforts were diverted away from the information-theoretical aspects of the field and were concentrated on those key component and subsystem technologies.

1.3 Component/Subsystem Technologies

An optical-communication system requires many component technologies. Virtually any one of them can be critical depending on the specific system requirements. It would be impractical to describe them all here, but there are a few component technologies that frequently make the critical list, and these are described below.

1.3.1 Laser Transmitters

One of the most important component technologies involves laser transmitters. When JPL began work in optical communications, laser transmitters had limited powers (less than 100 mW), their efficiencies were very low (less than 1 percent), and they were very unreliable. Some efforts, sponsored by the United States Air Force, had developed a cavity-dumped neodymium-doped, yttrium-aluminum-garnet (Nd:YAG) laser for possible space applications (later downgraded for an airborne laser communication demonstration), but its wall-plug power efficiency was about 0.5 percent [34]). Such power conversion efficiencies were too low to be viable for deep-space missions where power generation is extremely difficult and costly. Building on research already underway at the California Institute of Technology (Caltech), JPL began doing research on monolithically integrated semiconductor laser arrays [35–39]. Semiconductor lasers were much more power efficient than conventional solid-state (e.g., Nd:YAG) lasers, but their output powers were much lower. It was thought that by combining many laser diode elements together in a phase-locked array transmitter, the power output could be increased to the requisite (1–3 W average) levels, and the resulting transmitters would be extremely efficient (perhaps 40 percent). Additionally, one could also consider electronic beam steering of the beam from a laser diode array.

Initial progress, both at JPL/Caltech and elsewhere, was very promising, and significant increases in power levels were achieved [40]. Additionally, phase steering was demonstrated in many devices [41,42], but two problems remained. First, despite the increases in average power levels, the PPM modulation required that the laser energy be concentrated in high peak pulses. When the full average power levels of semiconductor laser arrays were concentrated into short-duration pulses, the instantaneous power densities at the laser facets far exceeded the device damage thresholds. Additionally, high-power laser arrays required efficient thermal conduction from the lasing epitaxial layer and hence required wafer-side mounting to the copper heat sinks. But, access to that same wafer side was required to control injection currents to accomplish electronic beam steering. Hence, the mounting required for high-power generation would short out all the control signal lines for the electronic beam steering.

The semiconductor laser arrays really functioned like efficient optical batteries (i.e., efficient converters of electrical energy to continuous wave (CW) optical energy). What was needed was the equivalent of an optical capacitor that could store and accumulate that optical energy until it was needed for a short optical (PPM) pulse. Nd:YAG laser rods could act as an optical capacitor, storing the optical energy in the fluorescent lifetime of the Nd ions, but they were just too inefficient in converting electrical energy into excited Nd ions.

In 1984, Don Sipes had an idea to improve this energy efficiency [43]. He noted that the contemporary designs of Nd:YAG lasers surrounded the Nd-doped YAG rod with laser diode pumps, but stimulated Nd:YAG laser emission along the central axis of the laser rod. Furthermore, the rod material was highly absorbent at the pump laser wavelength (by design), giving rise to very high excitation levels near the circumference of the rod, but the pump power density in the region of the rod where lasing occurred was much lower. He then reasoned that if he could inject the diode laser pump energy along the same axial space where the laser cavity mirrors were stimulating the Nd:YAG laser emission, then the conversion factor would be much better. This could be done with proper anti-reflection coatings on the cavity mirrors. Additionally, if both the pumping mode and the lasing modes were made very small in diameter inside the rod, then the conversion factor would be even larger. With a research investment of only a few thousand dollars, he assembled such a laser that produced greater than 5 percent electrical conversion efficiency on the first try [43–45].

Further improvements on this approach over the years have increased the power levels to more than 10 W [46–48]. Figure 1-4 shows a later version of this design that produced 2 W of pulsed and frequency-doubled (green, 532-nm) output laser power, and up to 11 W of pulsed laser power at the Nd:YAG fundamental wavelength of 1064 nm. This design demonstrated that laser powers and efficiencies realistic for deep-space optical communication were possible. This laser structure was the only viable deep-space laser approach for almost 15 years until fiber-amplified lasers began to emerge.

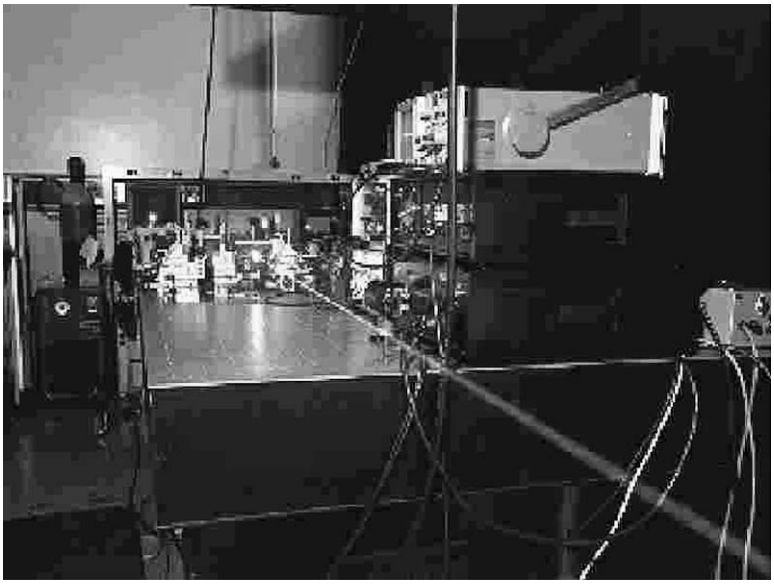


Fig. 1-4. 2-W frequency-doubled laser (11 W if not doubled).

1.3.2 Spacecraft Telescopes

Another key technology component is a thermally stable and lightweight optical spacecraft telescope. Serving as the optical version of an antenna, this telescope was required to keep surface deformations under a small fraction of an optical wavelength (a small fraction of a micrometer [μm]) and to do so over a large temperature range. Thermally stable glasses had been used in many applications, but they required too much mass. Through a Small Business Innovative Research (SBIR) contract with SSG Inc., a 30-cm-diameter telescope that was very precise and thermally stable was developed. Made entirely of silicon carbide, the telescope had a mass of only 6 kg. Figure 1-5 shows the delivered telescope.

1.3.3 Acquisition, Tracking, and Pointing

As mentioned above, one of the most important reasons for considering optical communications is the narrow beam divergence that allows the transmitted power to be concentrated on the receiving target location. However, that narrow divergence benefit comes with the penalty that the beam must be precisely pointed, or the entire benefit is lost. This pointing must be accomplished in the presence of attitude changes of the host spacecraft that are perhaps a thousand times larger than the laser beam divergence. Additionally, platform jitter disturbances can be many beam-widths in magnitude and can have characteristic frequencies of a hundred or more hertz. Finally, the transmitted beam from a spacecraft must be offset (pointed ahead) from the apparent location of the receiving target to compensate for cross velocities between the host spacecraft and the reception location. The normal way of accomplishing all these functions is for the spacecraft terminal to acquire and track an uplink beacon signal from the intended receiving target. That beacon is

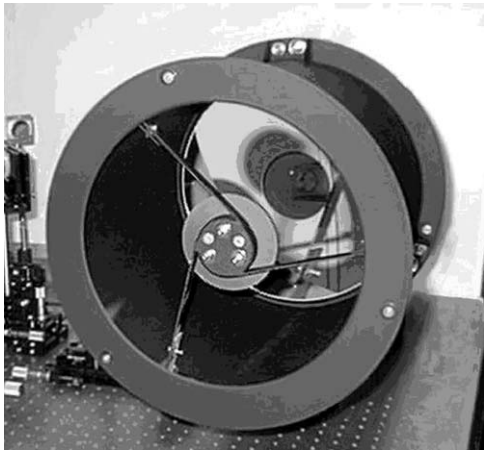


Fig. 1-5. 30-cm silicon carbide telescope.

used to precisely calibrate the attitude orientation of the spacecraft's transmitting aperture. The beacon signal is also used to measure the vibrational components of the host spacecraft. Correction of both the telescope line-of-sight error, as well as compensation for vibrational disturbances, is then accomplished using one or more fine-steering mirrors in the optical path to the telescope. This compensation scheme can also be used to implement the needed point-ahead angle calculated from mission trajectory and planetary orbital predictions.

Initial work at JPL in the late 1980s resulted in the development of an Integrated Optical Communications Test Bed (IOCTB) shown in Fig. 1-6. The IOCTB contained the necessary components to simulate a beacon signal and accomplish the required beam-pointing functions. It served as a familiarization test bed until newer acquisition, tracking, and pointing techniques were developed.

One of the early concerns was the difficulty of getting a sufficiently strong laser beacon signal out to a spacecraft when it is at one of the outer planets. As an alternate approach, techniques were investigated that relied on the solar-illuminated Earth itself as a beacon. Several strategies have been investigated over the years that use different tracking reference sources, either from visible sunlight reflected off the Earth or from the infrared emissions of the Earth as seen against the cold sky background [49,50]. To date, while promising, these techniques have yet to prove that they can provide adequate reference signals under all the various conditions and still be competitive with direct beacon tracking.

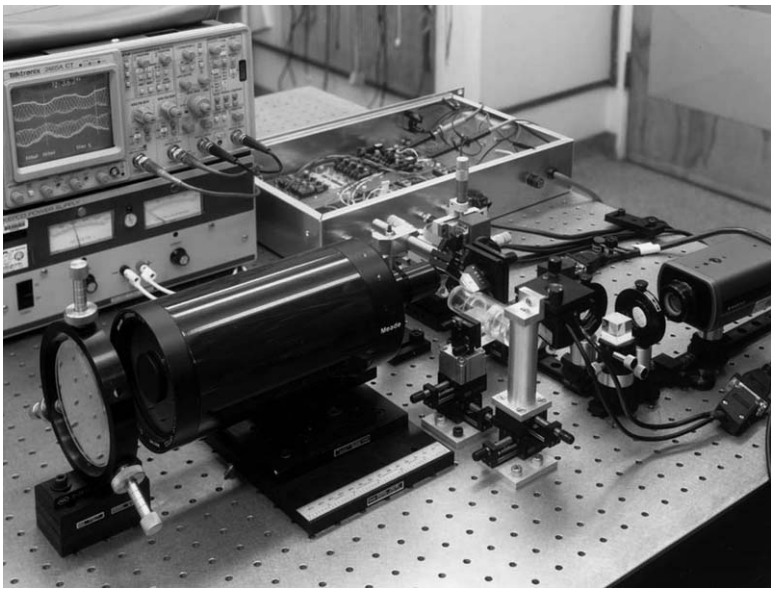


Fig. 1-6. Integrated optical communications test bed.

JPL has also investigated, as a hybrid technique, tracking of a weak uplink laser beacon signal used in conjunction with inertial sensors (e.g., accelerometers) to measure the vibrational components of the spacecraft. This approach is much more promising and allows the weak uplink beacon to be integrated longer to determine the spacecraft absolute attitude, while the inertial sensors permit compensation for the higher-frequency vibrational components [51,53,54].

1.3.4 Detectors

Another crucial component technology is that of detectors. Both detectors for optical-communication data extraction and detector arrays for spatial acquisition and tracking are needed. For data channel detection, the detector used in the multi-bit/photon demonstration was an RCA 31034C PMT. However, this and similar PMTs suffered from two problems. First, their quantum efficiencies at the primary candidate operational wavelengths were too low (typically less than 1 percent). Second, the tubes had such high gains that nominal background and/or strong signal levels would likely cause output currents that exceeded the anode plate current limitations. Clearly some other kind of detector was needed. An alternative is to use an avalanche photodiode detector (APD). Normally, APDs are operated in a mode where a bias voltage up to, but not exceeding, the spontaneous avalanche breakdown voltage is applied. The higher the voltage, the higher the gain, but also the higher the rate of spontaneous dark-count-generated detection events. Furthermore, the output resulting from the avalanche gain of the detected signal had a high variance, resulting from random multiplication gains through the photodiode's lattice structure. Although often higher in quantum efficiency, such detectors were not suited for detection of single photons.

In 1985, JPL began looking at APDs that were biased beyond the avalanche breakdown voltage. In this case, the gains would be high enough to detect single photon arrival events. Under normal conditions this would result in a constant avalanche condition due to thermally generated carriers in the photodiode. However, by cooling the APD nominally down to about liquid-nitrogen temperatures, the thermal carrier generation process could be significantly suppressed. That would leave the photodiode detector ready to trigger a massive avalanche, but with most of the thermally generated false detections eliminated. The result would be an optical detector that operated similarly to the way a Geiger counter works on radioactive detection events. To verify this, a test setup was created, and APD detectors were tested under single-photon input level conditions. Greater than 30 percent quantum efficient detection of single photons was demonstrated [55–60].

One of the problems identified in these "Geiger-mode" detectors was that after a triggered event occurred, whether from an incident signal or background

photon, or from a residual thermally generated carrier, the avalanche process would have to be stopped (or quenched). One way to quench these avalanches was to place a resistor in series with the APD. When an avalanche would start, the voltage drop across the resistor would reduce the voltage across the APD to below the avalanche breakdown voltage, thus stopping the avalanche. However, the resistance of the load resistor, coupled with the junction capacitance of the APD, resulted in a relatively large resistance–capacitance (R–C) time constant, thus overly limiting the bandwidth of the detection system. An alternate approach was to build an active quenching circuit that would rapidly trigger an electronic voltage interrupt. Unfortunately, such circuits were difficult to design and operate at that time.

More recently, work has been done on operating commercially available APDs at voltages just under the avalanche breakdown voltage. Since the voltage is high, the gain, and hence detectivity, is also high. But, since the detector is operated below avalanche breakdown, the detector does not lock up in a sustained avalanche, and the output resembles an amplified version of the input. Additionally, by cooling the detector, the resulting dark count rates can be minimized. Single-photon detection efficiencies greater than 30 percent have been demonstrated [61].

The other major detector needed is a detector array for the spatial acquisition and tracking system on the spacecraft. This detector is used to track the location of a beacon signal from the intended receiving location and often a portion of the outgoing transmit beam signal for precision beam pointing. The detector must have a large-enough field of view to cover the attitude uncertainty of the host spacecraft (often several milliradians [mrad]), yet produce final spatial resolution measurements that are a small fraction of a transmitted beamwidth (resolutions well below a μrad). Furthermore, the detector must be read out fast enough to compensate for higher-frequency vibrations on the spacecraft that would cause excessive beam jitter. Conventional charge-coupled device (CCD) detector arrays have adequate field of view (FOV) and resolution, but the typical frame rates (10–30 Hz) are inadequate to follow higher-frequency jitter components. A significant amount of effort was then directed toward windowed CCD arrays. With a windowed array, only small regions (typically 10 × 10) around the desired spatial tracking points need to be read out; after which, the rest of the array signal can be dumped and the next image taken. By windowing, the repeat time to the desired tracking points (after acquisition has occurred) can be fast enough to track even the higher-frequency jitter components. In the future, even more efficient tracking detectors will be possible with the use of active pixel sensor (APS) detector arrays. With APS detectors, the signals from the windowed regions of interest will not need to be read off the detector chip. Instead, it will be possible to process the signals into real tracking information via on-chip complementary metal oxide semiconductor (CMOS) processing.

1.3.5 Filters

On the receiving end of the link, narrow-band filters will be required before the detectors, especially if daytime reception on the ground is to be used [62]. Narrow transmission bandwidths will eliminate much of the background light interference, but the throughput efficiencies must be high to avoid causing significant loss to the desired signal. Multi-dielectric filters are the commonly used filters, but they are limited in how spectrally selective they can be and still have adequate throughput.

One filter investigated in this category is the Fraunhofer filter [63]. In the solar spectrum, there are narrow regions where the solar energy is trapped by certain elements in the Sun's photosphere. These are regions of the solar spectrum where the Sun is effectively dark (or at least not so bright). By selecting a laser line that corresponds to a Fraunhofer line, and then using an interference filter matched to that line, communications can take place with significantly lower background interference levels. One of the laser wavelengths of early interest was that of a frequency-doubled Nd:YAG laser at 532 nm. Several spectral dips exist in the solar spectrum near 532 nm.

To achieve really narrow passbands (less than 1 nm) it is necessary to use filters that are based on atomic transitions in materials. Atomic resonance filters (ARFs) can produce sub-nanometer bandwidths. However, these filters cannot be used in front of acquisition and tracking systems since the filtering operation relies on the absorption of a photon at one wavelength and the corresponding emission of another at a new wavelength. The creation of the new photon is dependent on the energy absorption of the input photon, but its angular direction is not preserved. To get around this, work was done in the early 1990s on the development of filters that produced polarization rotations as a result of the anomalous dispersion shifts of certain pumped gasses. Two versions were studied: the Faraday anomalous dispersion optical filter (FADOF) and the Stark-shifter anomalous dispersion optical filter (SADOF) [64,65]. Both filters work by passing polarized light into an atomic cell. If the input light is precisely on resonance with the excited gas in the cell, the input light will undergo a polarization rotation due to the anomalous dispersion of the gas. Light that is not precisely on resonance (i.e., background light) will pass though the cell but without the polarization rotation. By placing a crossed polarizer at the output of the cell, only the on-resonance light is allowed to pass. Furthermore, since the light is not absorbed and then re-emitted, the angular direction of the on-resonance light is preserved. A diagram of the SADOF filter is shown in Fig. 1-7.

1.3.6 Error Correction Coding

The last, but by no means the least, component technology to be discussed is optical coding. As mentioned earlier, the original multi-bit/detected-photon

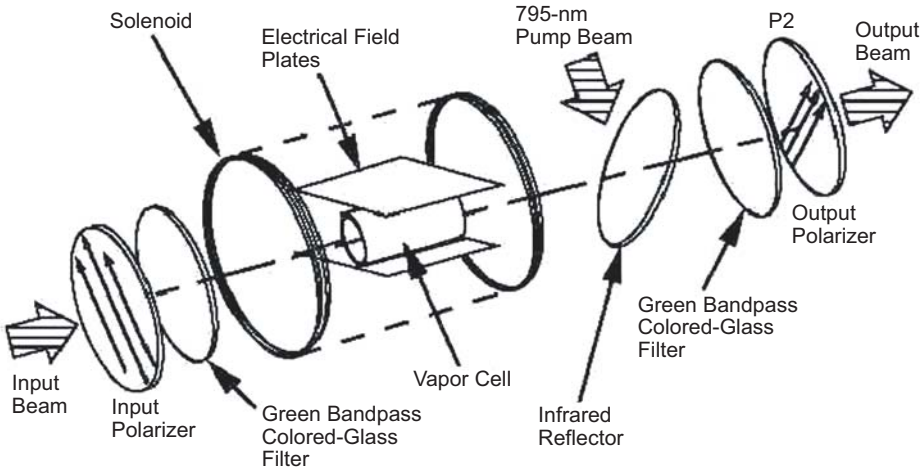


Fig. 1-7. Optical path through a SADOF filter.

demonstration used high-order PPM modulation (256-PPM) with a high alphabet (8-bit alphabet) RS code. The RS alphabet was matched to the PPM modulation since each 8-bit character would specify which of the 256 pulse locations would be used for that character. The prevailing belief was that the higher the order of the PPM modulation, the better the performance, provided the modulation was used with a matching-alphabet RS code. However, as the PPM order increased, the matching RS alphabet (and hence code) became much more complex. Furthermore, it was known that if the PPM order was reduced (along with its matching RS code), performance of the link was significantly reduced. This usually forced system designers to consider only high-order PPM modulations, but high-order PPM meant a high value of peak-to-average power level from the laser since the laser's average power was concentrated in a much narrower (and infrequently filled) pulse slot. Laser power limitations became a constraint on how high the order of the modulation could be.

Recent progress has been made in the development of codes that can relax the need for higher-PPM formats, and hence the required peak-to-average power levels of the lasers [66–70]. The codes (called accumulator codes) are based on product-coding techniques where simpler codes are combined and then jointly (and iteratively) decoded. One of the benefits is that one can start with a lower-PPM alphabet that is further from the overall channel capacity limit and regain a large portion of the lost performance with coding. Going to higher-order PPM modulations and using a good code over that modulation is still better in terms of performance, but the difference between properly coded lower-order modulations and properly coded higher-order modulations has diminished. Table 1-1 gives a comparison of several different PPM modulation orders and corresponding coding gains from the accumulator codes. Note the higher coding gains for the lower PPM orders.

Table 1-1. Coding gains of accumulator codes (in dB) for various PPM modulations.

PPM Order	2048	256	16	4
Gain relative to RS code	2.25	2.78	4.82	9.08
SNR gap to capacity (dB)	1.26	1.29	1.03	1.08
Optimal constraint length	2	2	3	4
Average iterations required	9	9	7	6

1.4 Flight Terminal Developments

Component technologies by themselves will not constitute an overall subsystem. They cannot be just hooked together and used because the design parameter spaces are large and the interfacing requirements can be tight. Furthermore, even the mounting platform (e.g., optical bench) on which some of the optical components are mounted must often be both thermally and mechanically stable since the design tolerances, due to the short wavelengths of optical signals, are frequently very tight and must be maintained over the temporal and environmental life of the system. Several flight transceiver (terminal) designs have been completed over the last two decades.

1.4.1 Optical Transceiver Package (OPTRANSPAC)

The first flight terminal system design was the Optical Transceiver Package (OPTRANSPAC) study conducted it 1984 [71]. It was a contracted study with McDonnell Douglas Corporation and leveraged their prior work for the United States Air Force on the Airborne Flight Test System and subsequent development activities for the Defense Support Program's (DSP's) planned Laser Crosslink System (LCS). The design had independent detectors for spatial acquisition, spatial tracking, and uplink data detection. The design was being performed as a pre-project study for a possible flight demonstration on the Cassini deep-space mission to Saturn. The OPTRANSPAC system design had a 28-cm telescope, had a 400-mW (frequency-doubled Nd:YAG) laser, and would return a 100-kbps communication flow from Saturn to a 10-m-diameter Earth-orbiting receiving aperture. The mass and power consumption estimates were 52 kg and 57 W, respectively, and the terminal occupied a volume of approximately 0.1 m^3. At the time, the mass was considered too much for the Cassini mission to fly as a mission-enhancement demonstration, so the full-scale development was not continued. However, the OPTRANSPAC study results were used as a basis for the IOCTB development mentioned earlier. A sketch of the OPTRANSPAC terminal design is shown in Fig. 1-8.

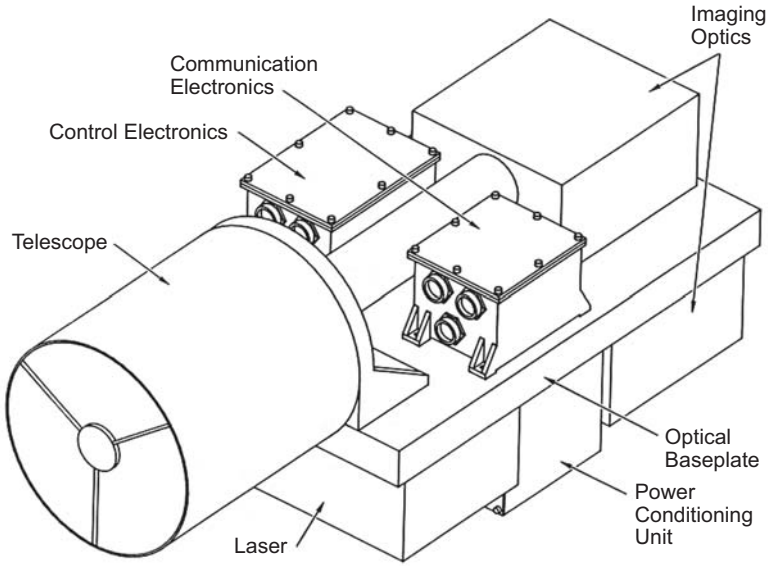


Fig. 1-8. Isometric drawing of the OPTRANSPAC flight terminal design.

1.4.2 Optical Communications Demonstrator (OCD)

One of the attributes of the OPTRANSPAC design was that it used separate detectors for acquisition, for tracking, and for beam point-ahead (the offset angle needed to lead the Earth-station receiver when there is relative cross-velocity between the two ends of the link). Additionally, there were separate fine-steering mirrors to implement the necessary beam centering and offset functions. Detectors and steering mirrors are primary optical system components, but they usually need secondary elements (e.g., focusing lenses and beam-folding mirrors) to make them work properly. All these components must be precisely held on thermally stable structures in space terminals. This meant that if the basic design of a flight terminal had a lot of primary components, then the overall complexity, mass, and cost of the terminal would be much higher due to all the elements (primary components, secondary components, and supporting infrastructure) required to make the end system function properly. By realizing this relationship, it was conversely realized that if the number of primary components could be reduced, then the number of secondary components would also be decreased, as would the requirements for the supporting structure. This realization led to the basic design of the Optical Communications Demonstrator (OCD) [72–76].

The fundamental design for the OCD is shown in Fig. 1-9. The OCD concept works as follows. A beacon signal is sent to the flight terminal from the intended receiving terminal. That signal is received on the flight terminal by its telescope (depicted by just a lens in the diagram for simplicity). The telescope

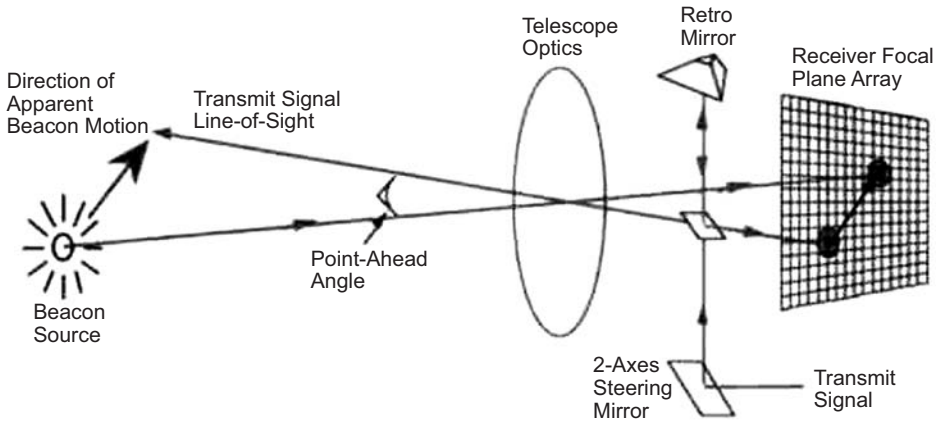


Fig. 1-9. Concept diagram for the Optical Communications Demonstrator (OCD).

collects the beacon signal and focuses it to a point on the receiver focal-plane array in the terminal. The location of this spot on the array represents the direction from the received beacon signal relative to the telescope's axis (the center of the array). The array size determines the field of view of the telescope and is large enough to cover the initial pointing uncertainties of the telescope (often defined by the attitude control dead-band limit cycle of the spacecraft). No overt effort is made to center the received beacon signal on the focal-plane array. This just represents the knowledge of the direction to the receiver. The modulated laser signal that is to be returned by the flight terminal to the ground receiver is coupled (via optical fiber) into the OCD assembly and initially strikes a two-axis steering mirror. After reflecting off the steering mirror, it passes up to a dichroic beam splitter that reflects almost all the signal out of the telescope. However, there is a small amount of signal that passes through the dichroic beam-splitter and progresses upward to a retro-reflector. The retro-reflected signal returns to the backside of the beam-splitter where it is directed toward and focused onto the focal-plane array. This spot location represents the direction of the outgoing laser signal relative also to the telescope axis. The vector difference between the focused beacon signal and the focused residual of the laser transmit signal on the focal plane represents the angular difference between the received and transmitted directions and is independent of the axis of the telescope. (The actual axis of the telescope is common and drops out in the vector difference.) Now, as stated earlier, there is a need to implement a point-ahead angle to the transmitted beam. This can be done by simply monitoring the vector difference between the two focused spots and making sure that it represents the needed point-ahead angle (which can be easily calculated given the orbital predicts and the nominal spacecraft-orientation information.

Note that in the OCD design there is only a single focal-plane detector array and a single two-axis steering mirror (primary components). Since the number of primary components has been reduced, there will be a corresponding reduction in the number of secondary components (not all are shown), and hence, there is a simplification of the entire structure. This design is called a minimum-complexity design, and it was specifically invented to decrease the complexity, mass, and cost of the flight design.

The OCD was implemented in a laboratory-qualified form/fit/function realization based on this minimum complexity design. Figure 1-10 shows the resulting implementation. The terminal has a 10-cm-diameter telescope, and the entire unit is about the size of a loaf of bread. The basic design assumes that the OCD is body-mounted to the spacecraft and that coarse pointing of the terminal is accomplished by the attitude orientation of the spacecraft. This is realistic for many deep-space missions, and in fact, the same is usually done with the conventional RF antennas. In those cases where pointing independent of the spacecraft orientation is required, such as is more common with Earth-orbital missions, then the OCD terminal can be mounted on a separate two-axis gimbal or used in conjunction with an external two-axis controlled-steering flat mirror. Figure 1-11 shows the OCD on a two-axis gimbal. Although the OCD was designed as a laboratory-qualified terminal, it has been used as the basis for many proposed flight designs. The basic terminal design would allow kilobits per second data returns from Pluto distances and multiple gigabits per second data returns (with significant excess link margins) from Earth-orbital distances to the ground. Additionally, the OCD terminal has also been taken into the field for a series of 46-km mountain-peak-to-mountain-peak demonstrations (see Section 1.10.4).

1.4.3 Lasercom Test and Evaluation Station (LTES)

The characteristics of a flight terminal engineering model cannot be evaluated without also developing a test infrastructure for measuring its

Fig. 1-10. Basic OCD unit.

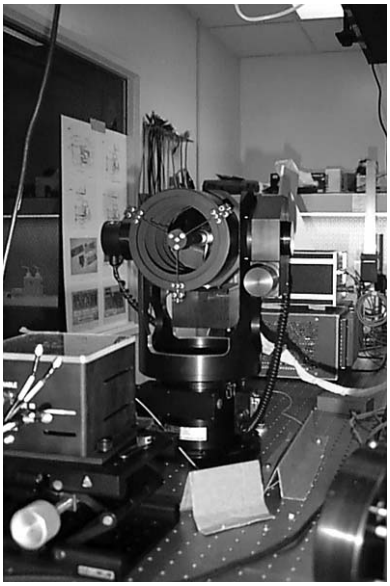


Fig. 1-11. OCD on gimbal.

performance. Accordingly, the Lasercom Test and Evaluation Station (LTES) was developed. The LTES was viewed as a more general-purpose test station, so it was required to operate over a broad wavelength band (from 0.5 μm to 2.0 μm). The LTES can provide a calibrated beacon signal for use with the flight-engineering model under test, and it can receive and analyze the laser signal from that engineering model. Tests that can be made include spatial acquisition and tracking performance parameters, detection of transmitted data, and measurement of transmitted power levels. Although designed originally as a test infrastructure for the OCD, the LTES was first used to perform selected tests on a Ballistic Missile Defense Organization-developed optical communications flight terminal that flew on the Space Technology Research Vehicle 2 (STRV-2) mission. The LTES was sensitive enough to determine that the flight unit would have significant beam wander and misalignment of the parallel transmitting lasers as a function of the terminal warm-up temperature. Figure 1-12 is a photograph of the LTES.

1.4.4 X2000 Flight Terminal

The next major flight terminal design effort was undertaken as part of the X2000 program. The X2000 program was initiated to fill the gap of major needed technology developments required for future missions. It was recognized that because of the NASA shift to faster-better-cheaper (FBC) missions, the technology developments that had customarily been developed as part of the former “flagship” missions would no longer be possible.

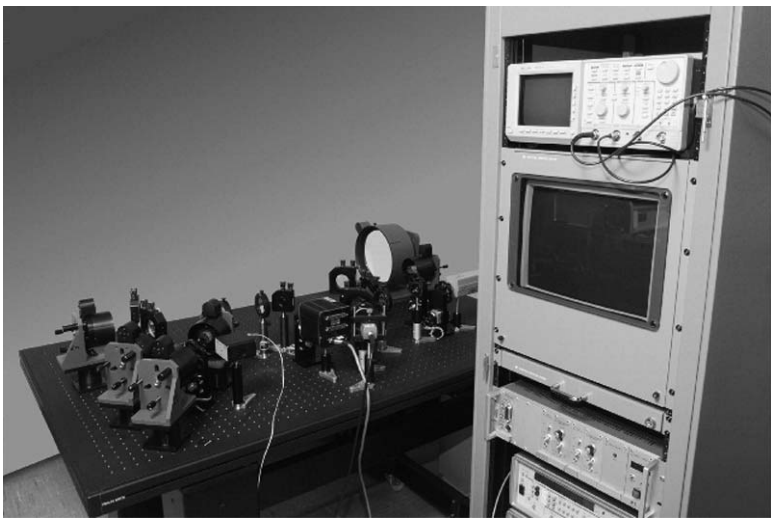
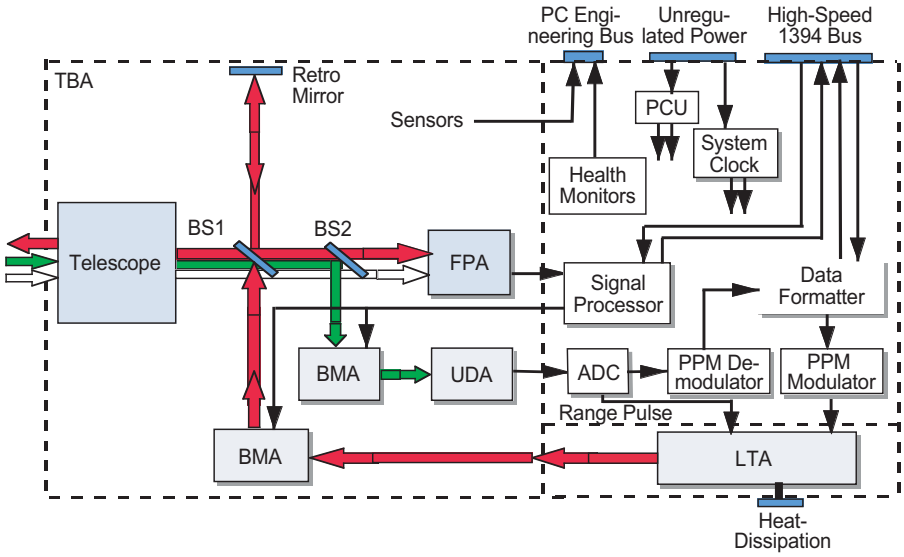


Fig. 1-12. Photograph of the LTES.

Development of an optical communications flight-qualified engineering model terminal for a proposed Europa Orbiter mission was the second largest planned development in the program.

A block diagram of the design is shown in Fig. 1-13, and it had several features that had not been included in the OCD development. First, the diameter of the telescope was increased to 30 cm. This was based on the successful development of the silicon carbide telescope mentioned earlier. Second, it was realized that the basic structure of the OCD contained a telescope and a focal-plane array, the two primary components in an imaging camera. Third, an uplink command detector path was added. Since this detector was a high-speed detector, it could also serve as an uplink ranging detector for an optical turn-around ranging system. Thus, the requirements for the X2000 design included dual-use as a science imaging camera and as an uplink reception capability for command and ranging. Furthermore, the proposed Europa Orbiter mission study team was considering the use of a laser altimeter. It was realized that the optical communication telescope and high-speed uplink detector could also be used as the laser altimeter return signal receiver/detector. A computer-aided design (CAD) drawing of the X2000 terminal preliminary design is shown in Fig. 1-14.

The X2000 optical communications development proceeded to the point of a concept design. However, budget pressures in the rest of the X2000 program ultimately caused cancellation of all X2000 developments except for the primary element, a spacecraft computer/avionics system. The optical communications terminal design had been progressing well, but the application time frame for the technology was considered to be far enough to accommodate



ADC = Analog-to-Digital Converter
BMA = Bi-Axis Mirror Assembly
FPA = Focal Plane Array
LTA = Laser Transmitter Assembly
PCU = Power Conditioning Unit
PPM = Pulse Position Modulation
TBA = Telescope & Beam-Optics Assembly
UDA = Uplink Detector Assembly

Fig. 1-13. Architectural design of the X2000 flight terminal.

the delayed development. The system and analysis to that date were captured in a study report [77].

1.4.5 International Space Station Flight Terminal

Another flight development program came along a few years later. In 1996, a NASA call was released for payloads that could be demonstrated on the International Space Station (ISS). The objective was to use the ISS as an engineering center for such demonstrations. A proposal was written, and in 1997 the proposal was selected for development. The program was funded under the ISS Engineering Research and Technology (ISSERT) program that also funded a number of other attached payload developments. The terminal design was based on the OCD architecture, and the plan was to transmit at least 2.5 Gbps from a terminal mounted on the ISS external nadir-pointing truss to the ground. The flight terminal would be developed and integrated with the ISS express pallet, for subsequent transport in the Space Transportation System (STS or Space Shuttle) to the ISS. The ground terminal would be a new Optical Communications Telescope Laboratory (OCTL) that was already funded under the Deep Space Mission System (DSMS) Technology Program (see Section 1.12.1). The operational concept included an uplink beacon from OCTL to the ISS mounted terminal. The flight terminal would spatially lock onto that beacon

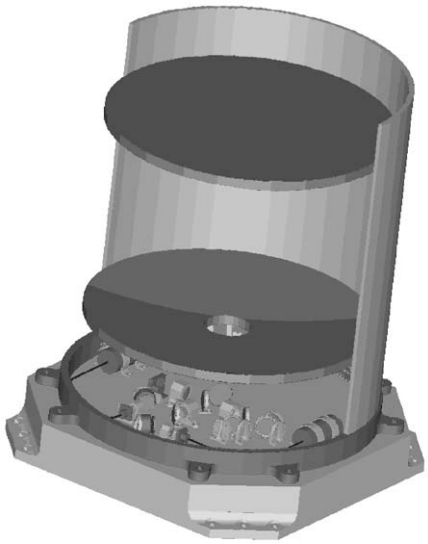


Fig. 1-14. CAD drawing of the X2000 flight terminal design.

and transmit down a pseudo-random coded data stream. Early in the program development, the ISSERT management at NASA Johnson Space Center realized that the optical communication terminal represented a valuable resource that would likely be underutilized. As a result, they initiated a change order to provide an optical-fiber transfer line from the interior of the ISS to the optical communication terminal location on the external truss. This would allow real data to be sent over the optical link to the ground.

Figure 1-15 shows the location on the ISS external truss where the terminal was to be located. Unfortunately, as was the case in the X2000 development, budget pressures were heavy here as well. The program progressed through Phase A and had just completed its preliminary design review (PDR) when budget pressures related to the building of the core ISS resulted in cancellation of all attached payload developments, including the optical communication terminal.

1.5 Reception System and Network Studies

Studies and system designs have also been done on the Earth-reception side of deep-space optical communications links. Both ground-based and Earth-orbital receivers have been studied, although the ground-based receivers appear to be the most realistic for the time being. However, that may change over time if access to space becomes more routine and less costly.

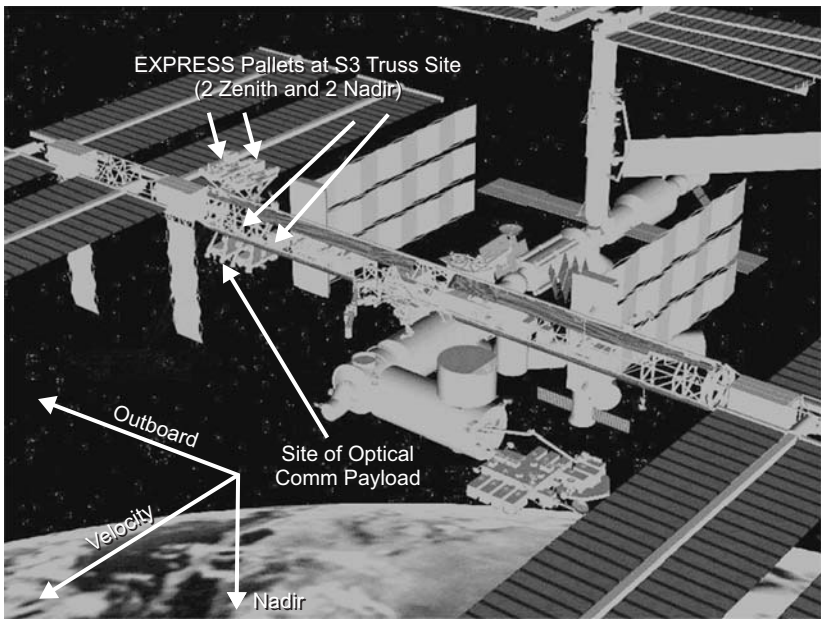


Fig.1-15. ISS showing planned site for the optical communication terminal.

1.5.1 Ground Telescope Cost Model

The first serious look at the definition of a ground receiving system was done in 1986 and involved establishing a cost-versus-performance model for ground-based telescopes [78,79]. The study started with a set of data on existing RF, solar concentrator, and optical astronomical telescopes. When the costs of those telescopes were plotted as a function of diameter, it was noticed that the costs could be modeled as

$$C = \alpha D^x$$

where C was the cost (in \$M), D was the diameter of the telescope (in meters), and x was a value that varied between 2.4 and 2.8 (taking 2.6 as a nominal value). The value of α was dependent on the inverse of the telescope's focused blur circle diameter "F" and was approximately given by

$$\alpha = 10^5 / F$$

(the actual expression is given in [78]). Next, the performance of a communication link was calculated as a function of telescope diameter and blur circle using a reference transmitter and a set of background conditions. Since the cost and the performance could each be calculated based on the same two parameters (diameter and blur circle), then the cost of the telescope could be

plotted as a function of communication performance with telescope diameter and blur circle diameter as parameters. Upon optimizing over diameter and blur circle, one then had a plot of optimized-cost-versus-communications performance.

Figure 1-16 shows the results of this analysis but extrapolated to a worldwide network. The analysis showed that the knee of the cost curve occurred at about 18 dB of improvement over the reference X-band link. The values of the optimized parameters in this region were a 10-m-diameter telescope and a blur circle that was larger (less precise) than the diffraction-limited focus. This result was intuitively satisfying since it was known that 10-m-diameter diffraction-limited telescopes could be quite expensive but that similar-sized solar concentrators were much less expensive. Since non-diffraction-limited telescopes were essentially photon buckets, then this also meant that direct detection of the received signals could be used, and there would not be a need to compensate for atmospheric turbulence-induced phase fluctuations.

1.5.2 Deep Space Optical Reception Antenna (DSORA)

Given the insight afforded by the cost-modeling effort, a series of studies was conducted to define, analyze, and estimate the cost factors for various realizations of a 10-m-diameter photon bucket [80–88]. These generally went under the name of Deep Space Optical Reception Antenna (DSORA). Early in the process, it was realized that some form of sunshade would be important if the system was to be used in the daytime, especially if that use was directed

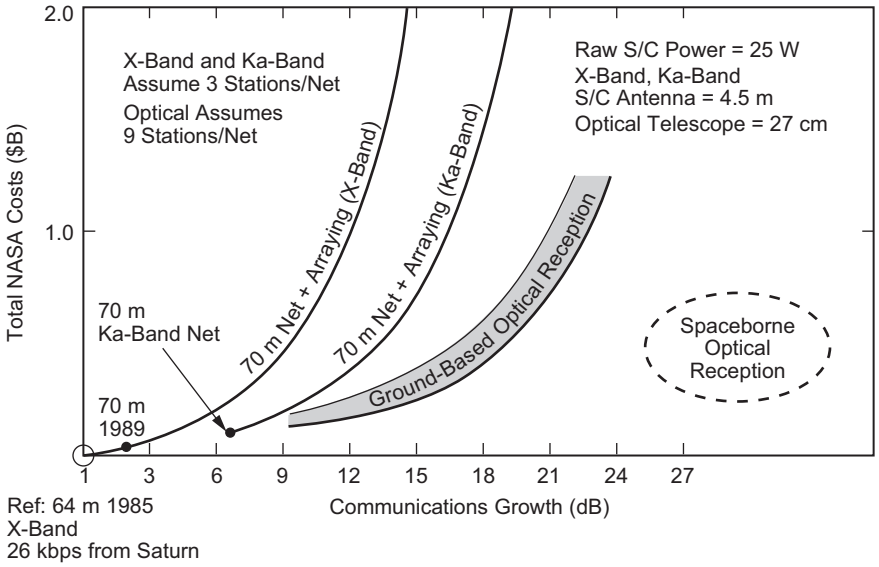


Fig. 1-16. Optimized cost model for Ka-band, X-band, and optical reception.

anywhere close to the Sun. Several sunshield concepts were explored, including the use of an external tube outside, but connected to, the dome. The favored approach was an "integral" sunshield that followed the "soda straw bundle" concept. The idea was to collect together a set of hexagonal tubes that had cross-sections the same sizes and shapes of the primary mirror segments. These would be placed over the primary collector surface, but with the lower portions of the central "tubes" shortened so that the ray paths from the primary mirror segments would not be blocked from getting to the secondary mirror. Since the length/diameter (L/D) ratio of each "tube" was large, the telescope/sunshield could point much more closely to the Sun without direct sunlight hitting the primary mirror surface. A drawing of the DSORA with an integral sunshade is shown in Fig. 1-17.

The remaining challenge, however, was the fact that the tubes became good collectors of solar radiation (heating), and there was concern that unacceptably large turbulence would result. Several concepts, including the use of "expanded metal" (similar to that used in window screens), were considered to mitigate this effect.

1.5.3 Deep Space Relay Satellite System (DSRSS) Studies

In 1992, and after examining various DSORA concepts, it was decided to look more carefully at the possibility of an orbiting reception station rather than a ground-based station. Two study contracts were let for a Deep Space Relay Satellite System (DSRSS), one to Stanford Telecommunications and the other to TRW. Actually, both studies were part of a periodic look at space-based

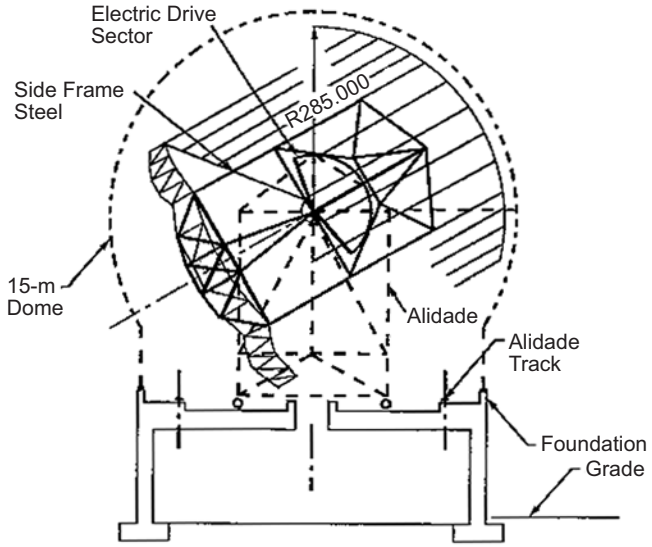


Fig. 1-17. Cross-section view of DSORA with integral sunshade.

reception for deep-space signals in general and were partially a re-examination of a concept for an orbiting RF-receiving station that had been done 14 years earlier [89]. Furthermore, the study statements of work required that the contractors look at both RF and optical reception stations to see which was the most attractive. The study contracts were for a 12-month duration, but at the mid-contract reviews, both contractors reported that the RF-reception system would not be competitive and requested that they concentrate the remainder of their efforts on the optical-system definition studies. The design proposed by Stanford Telecommunications was an on-orbit erectable optical telescope of 10 m in diameter [90]. TRW provided designs for two orbiting receivers, one a remotely deployable 10-m telescope that used direct detection, and a smaller 4-m diffraction-limited receiver that used coherent detection [91]. Based on cost and risk assessments, they did not recommend the smaller coherent system over the larger direct detection one. Although both designs would allow operation without having to worry about cloud blockages or daytime atmospheric scattered light, the cost estimates for these designs were not competitive with a network of redundant (i.e., spatially diversified to increase weather availability) ground stations. As a result, the orbital network approach was placed on the back burner for reconsideration at some time in the future, most likely as a possible second-generation capability to augment an earlier ground-reception network infrastructure.

1.5.4 Ground-Based Antenna Technology Study (GBATS)

In parallel with the DSRSS studies, JPL performed an updated study on ground-based optical receivers. The study, dubbed Ground-Based Antenna Technology Study (GBATS), considered both the details of the design for a 10-m optical reception ground station as well as the overall operational network architecture using such stations as element nodes [92]. The design of the 10-m telescope consisted of a segmented primary aperture with active control of the primary segments (to control low-bandwidth aperture distortions caused by gravity loading, thermal distortions, and wind buffeting). Furthermore, a collapsible dome structure similar to an existing United States Air Force 3.5-m telescope was included. For the network architecture, it was necessary to consider spatial-diversity reception from the beginning to circumvent cloud-cover outages.

Two fundamental architectures were considered. The first consisted of three clusters of three optical telescopes in each of the three current DSN RF antenna regions. This would allow the three-longitude paradigm of the current DSN to continue. However, for spatial diversity benefits, each of the clusters at each longitude would have to be spread out over several hundred kilometers to be in different weather cell regions. This automatically implied a network of nine stations. For the other architecture, the constraint that the stations needed to be

somehow “clustered” around an existing DSN station longitude was removed. This allowed the stations to be located in a pattern where one could act as a redundant neighbor for any of the stations in its neighboring longitudes. Networks of 6, 7, and 8 stations dispersed linearly in longitude around the globe were considered. It was found that the linearly dispersed optical subnet (LDOS) approach, rather than the DSN-centric “clustered” approach, was the more cost effective. Figure 1-18 shows the basic design of an optical station from GBATS, and Fig. 1-19 shows a sample LDOS configuration taken from that study. The GBATS results were used to compare with the DSRSS study conclusions. The cost, performance, and risk comparisons clearly favored the ground-based-network approach.

1.5.5 Advanced Communications Benefits Study (ACBS)

Although a lot of interest was being generated in the area of optical communications, technologies for alternative approaches were also progressing, and it became evident that a comparison of optical communications with an upgraded X-band system, as well as the emerging Ka-band system, was

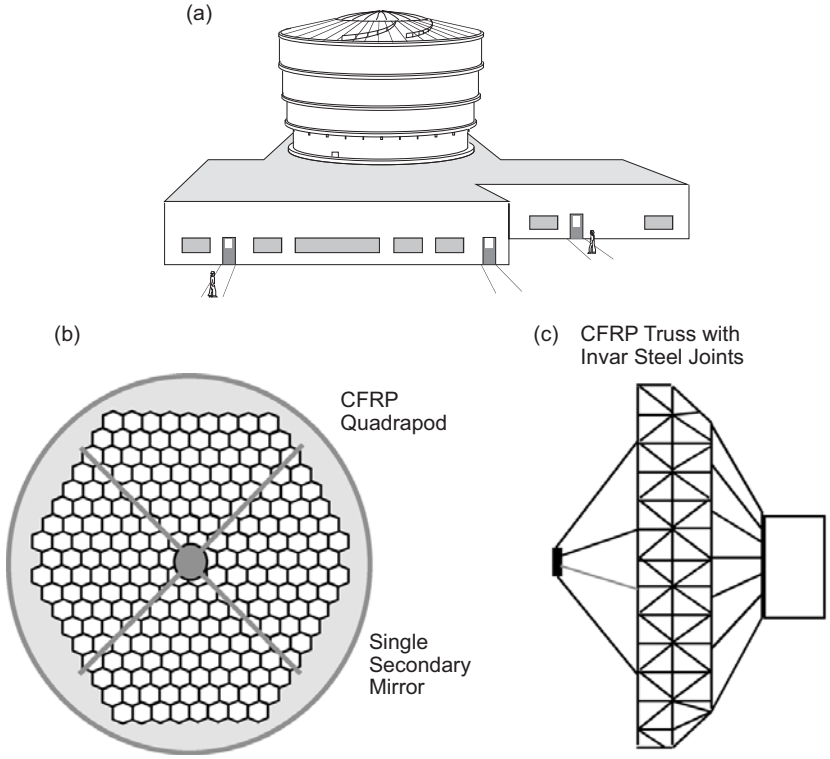


Fig. 1-18. GBATS ground station (a) building with collapsible dome; (b) single secondary mirror of carbon-fiber-reinforced plastic (CFRP); and (c) CFRP truss with invar joints.

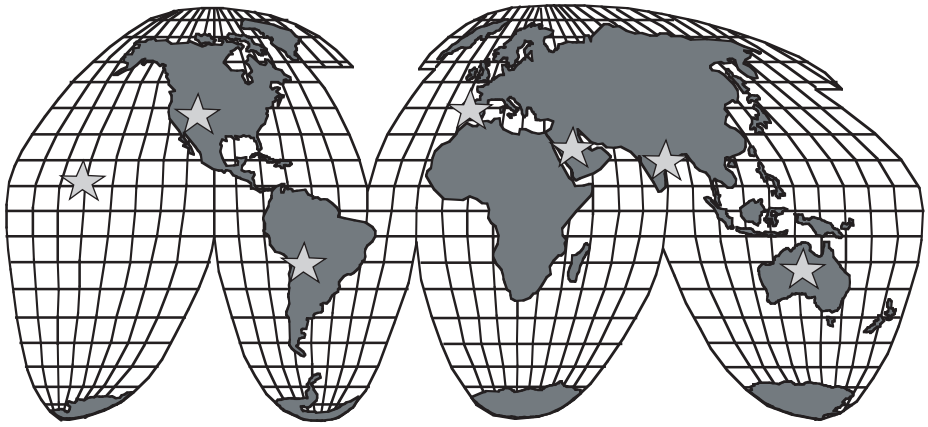


Fig. 1-19. Global map of the seven-station LDOS network.

appropriate. In 1996, an Advanced Communications Benefits Study (ACBS) was initiated [93,94]. The study constraints were to consider reception of signals from Mars at three daily data return volumes (0.1 gigabit (Gb), 1 Gb, and 10 Gb). Enhanced X-band, the emerging Ka-band, and the optical communications end-to-end systems designs were developed, and the overall cost estimates were compared. The cost estimates considered the cost of mass and power on the spacecraft, the non-recurring and recurring costs of the spacecraft terminals, and the cost of the ground infrastructure (recognizing that there was already a DSN infrastructure in place). As part of this study, a more detailed assessment of the design and the cost estimates for optical ground infrastructure were developed. The cost estimates were then evaluated based on total initial investment cost and the recurring spacecraft costs over some number of missions. Figure 1-20 shows the result of the study for the 10-Gb/day data volume design point. The results show that the initial investment required for the optical system is higher than either the X-band or the Ka-band approaches, but that investment cost is recovered after 5–8 missions because of the lower recurring costs of the optical systems.

1.5.6 Earth Orbit Optical Reception Terminal (EOORT) Study

In 1998, there was another examination of the orbiting optical reception approach. Building on the DSRSS study results, the Earth Orbit Optical Reception Terminal (EOORT) study considered two orbital configurations, one a 7-m optical photon bucket receiver and the other a 4-m coherent receiver [95]. The reasoning was that these values more closely compared in terms of performance with the 10-m ground-based designs that were being considered. Team X at JPL performed studies of these two configurations. Again it was found that the larger but noncoherent reception approach was less expensive,

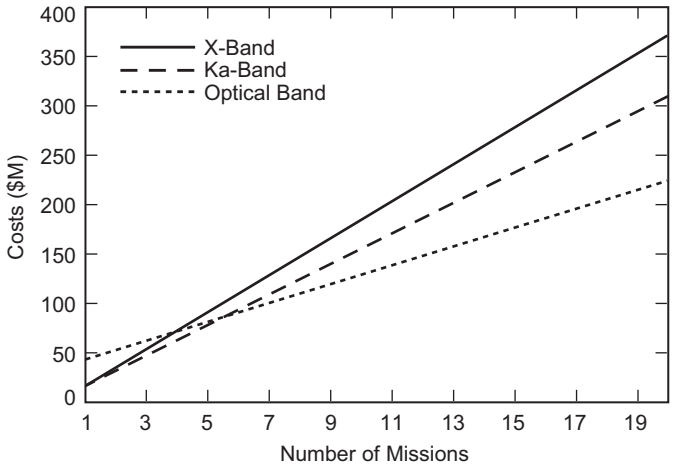


Fig. 1-20. Life-cycle cost comparisons of expanded X-band, Ka-band, and optical communications as a function of number of user missions.

but that the economic comparisons with the ground-based approach still favored the ground.

1.5.7 EOORT Hybrid Study

The real attraction of an orbiting receiver is that one does not need to worry about infrequent (but still possible) cloud-cover blockages during mission critical events. Still, for the bulk of the time, a diversified ground network provides perfectly adequate coverage and can do so at a fraction of the cost. This led to the idea that a hybrid approach might be a good solution. In 1998 a follow-on study to the EOORT study was conducted [96]. This study considered a reduced number of ground stations that could receive signals from the spacecraft much of the time (i.e., when not obscured by clouds), but it also had an orbital system with a small (70-cm in the study) telescope to guarantee that at least critical data reception could be received even when the relevant ground stations were clouded out. The solution looked attractive, although there would be a more challenging signal design required. In particular, the signal would have to be decodable with only the 70-cm aperture for the critical data, but it would provide the remainder of the science data if one or more of the ground station connections were available. How this would work operationally will require more investigation.

1.5.8 Spherical Primary Ground Telescope

One cost-effective approach for a ground receiver is to use a primary telescope aperture made up of spherical segments. Such segments are easy to fabricate, and hence, they can save significant costs for the station. However,

spherical segments also cause spherical aberrations that can further blur the telescope's focused energy (and hence cause the reception system to be more susceptible to background light interference). A recent design [97–99] took such a structure and matched it with a clamshell spherical aberration corrector. This looks like a very promising way to reduce costs of the ground station without sacrificing the performance of the system. Figure 1-21 shows a sketch of the optical station. Figures 1-22 and 1-23 show the optical ray trace diagrams for the segmented spherical primary mirror and the clamshell corrector, respectively.

1.5.9 Space-Based versus Ground-Based Reception Trades

Studies have continued on the definition and associated cost estimates for ground-based reception stations as well as space-based alternatives. There are certainly advantages to being above the Earth's atmosphere for signal reception. At the same time there are many advantages to having the reception

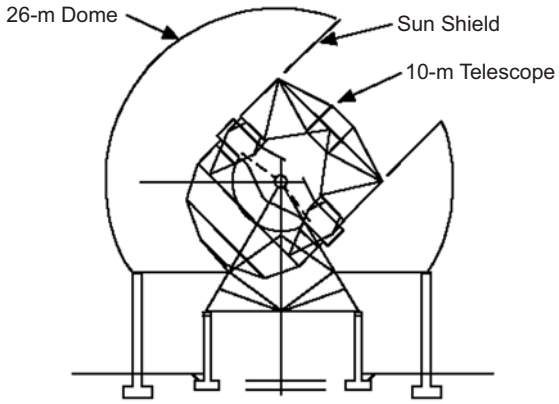


Fig. 1-21. Telescope and dome.

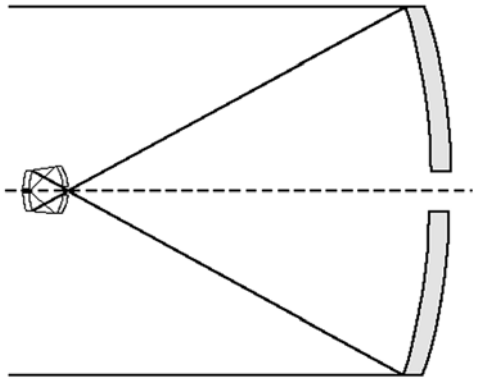


Fig. 1-22. Spherical primary/corrector.

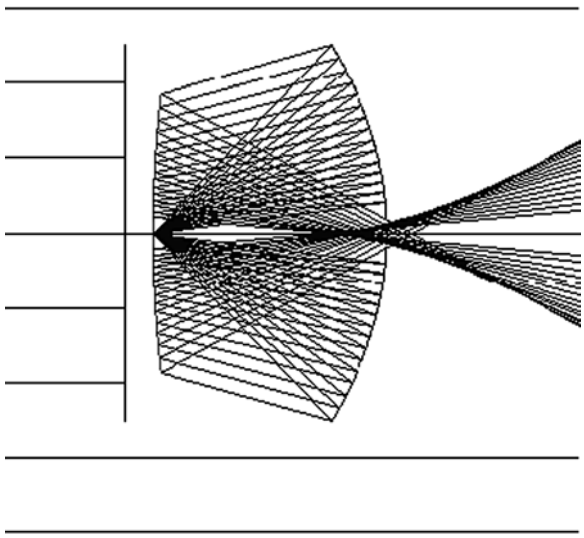


Fig. 1-23. Detailed ray trace of clamshell corrector.

infrastructure on the ground. At the present time, it appears that ground-based reception is the more appropriate choice. If one looks simply from an implementation cost standpoint, the cost estimates for an Earth-orbiting optical communications reception station are approximately the same as the costs required to implement an entire worldwide and spatially diversified ground network. This is comparing the ground net to just a single orbiting receiver. But, a single orbiting reception station would represent a single-point network failure risk that would take significant time to replace (months at a minimum if a standby spare spacecraft was available to years if a replacement development had to be restarted). This says that at least two orbital stations at a minimum would be required, if not more. This heavily weighs the balance toward the ground-based option. Furthermore, orbital receivers would have much shorter lifetimes, and they could not be maintained or upgraded like ground stations can be. Upgrades to new capabilities or wavelengths could not be phased in as easily as they could be on the ground, but would require engineering them in one of the next major block upgrades of the orbiting optical receiving terminal spacecraft series.

Continuous upgrades have been extremely important in the past history of the current DSN. For example, Fig. 1-24 shows the normalized (to Jupiter distance) increase in capability as a function of time that the DSN has achieved. Approximately 12 orders-of-magnitude of improvement in capability have occurred over the past 40 years. Most of this increase has been a result of transitions to higher communications frequencies (i.e., wavelength decreases). These changes have been implemented both on the user mission spacecraft and

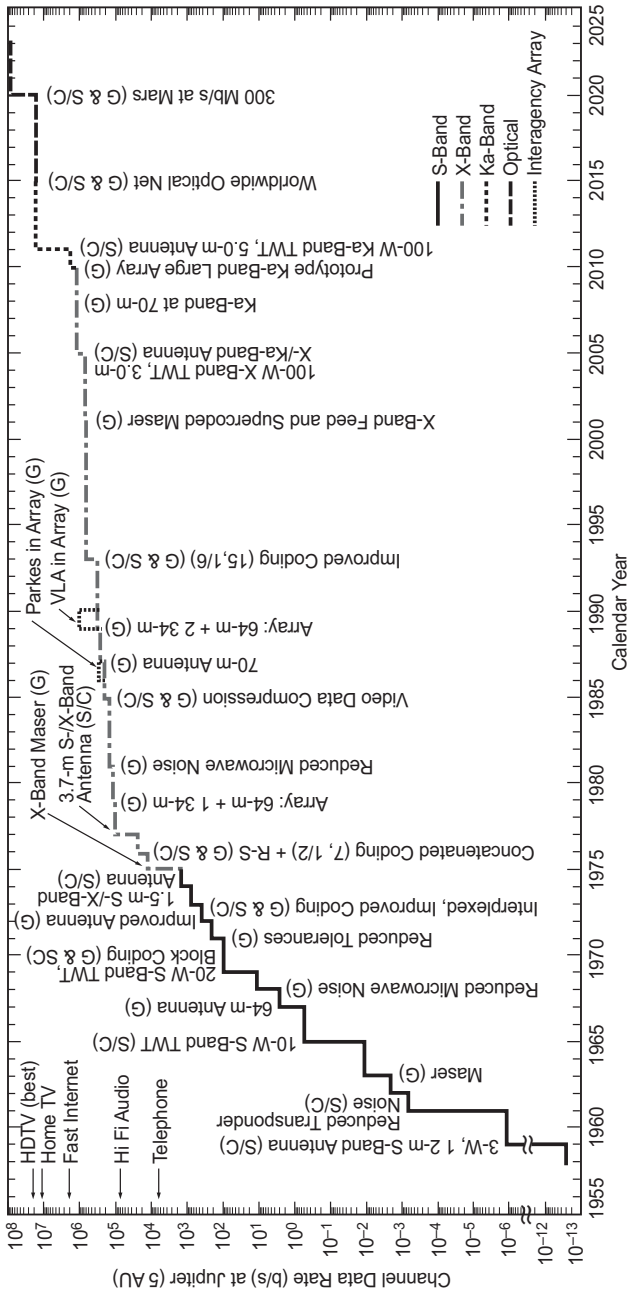


Fig. 1-24. Growth of normalized deep space communications capacity by improved ground (G) or spacecraft (S/C). (Note: major jumps in capability result from frequency changes.)

on the DSN. Having the reception systems accessible on the ground has been a key factor in enabling such a significant growth in capability. Had the DSN been in space, it would have severely restricted such growth to just the major orbital block upgrades, and would have resulted in increased technical risk for those block upgrades. Also, a 40-year longevity would have required a number of full orbital replacement cycles because of limited orbiting station lifetimes. It is clear that operational performance is a key factor in making future system architectural decisions, but the ultimate determining factor will be the overall life-cycle cost of the architectural options and will include all of these factors.

1.6 Atmospheric Transmission

In order to design a deep-space-to-ground optical communications link, it is necessary to understand the losses that will occur as the signal propagates through the atmosphere [100,101]. Both cloud blockages and atmospheric molecular absorption will impede the signal. Understanding the statistics on these losses is crucial so that the requirements for diversified reception (i.e., number of stations) and the resulting communications reliability can be determined. Molecular absorption is based on the percentages of different molecules in the atmosphere, and this effect can be reasonably well predicted using software tools developed over many years by the United States Air Force Research Laboratory (AFRL). As long as the wavelength used for communications does not lie on or very near a strong atmospheric absorption line, the clear-weather link attenuation is relatively constant. Clouds, on the other hand, occur much more randomly and can result in total extinction of the optical signal.

To assess cloud-cover statistics, JPL first obtained cloud-cover statistics taken from the Geostationary Operational Environmental Satellite (GOES) system. These statistics were in the form of cloud-cover contour maps provided by the University of Wisconsin [102], and clearly showed that clear skies are much more likely in the southwestern U.S. However, when comparing statistics, it was also evident that the sum of the clear-sky and cloudy-sky probabilities for a given spot was less than 1. The remaining probability mass is a result of partial cloudy conditions. Furthermore, it is important to know what defines clear or cloudy. Thin cirrus clouds may not show up on a satellite images as clouds, but they still result in some, albeit not always large, attenuation of the signal.

Realizing the need for more detailed statistics on atmospheric throughput, JPL created a program to make in situ measurements of the atmospheric throughput attenuation. To accomplish this, three atmospheric visibility monitoring (AVM) observatories were built and deployed in the southwestern U.S. [103–111]. One is located at Table Mountain, California, a JPL astronomical observatory site near the town of Wrightwood. The second AVM

observatory is located at the DSN's Goldstone Deep Space Communications Complex north of Barstow, California. The third is located on Mount Lemmon in Arizona.

Each AVM observatory contains a 25-cm telescope, a detector array, and several spectral filters on a filter wheel. The system is housed in a roll-off roof enclosure that is connected to a weather-sensing suite. The system operates autonomously, both day and night, to gather atmospheric throughput data by monitoring stars and measuring the stellar intensity on the ground in six spectral bands. By comparing the measured intensities of stars with the above-the-atmosphere values for those stars, the atmospheric throughput can be measured. The weather-sensing tower monitors for conditions at the site that would make telescope observation unsafe (i.e., high winds, rain/snow, excess humidity), and if such conditions are sensed, the enclosure roof and fold-down south-facing wall will close. Any time the enclosure is closed, or the system is not able to detect a star, the resident computer declares that the sky was totally cloudy. Otherwise, the observatory makes measurements of the stellar intensities and records them on the computer. Data are routinely transmitted back to JPL for processing and statistics generation. Figure 1-25 shows the AVM observatory at Table Mountain Facility (TMF) at Table Mountain, California. Figure 1-26 is a sample of a cumulative probability plot derived from the data. The horizontal axis is atmospheric attenuation (in decibels), and the vertical axis is cumulative probability. The two plots correspond to measurements made at two different wavelengths. For example, based on these

Fig. 1-25. Photograph of the AVM site at Table Mountain, California.

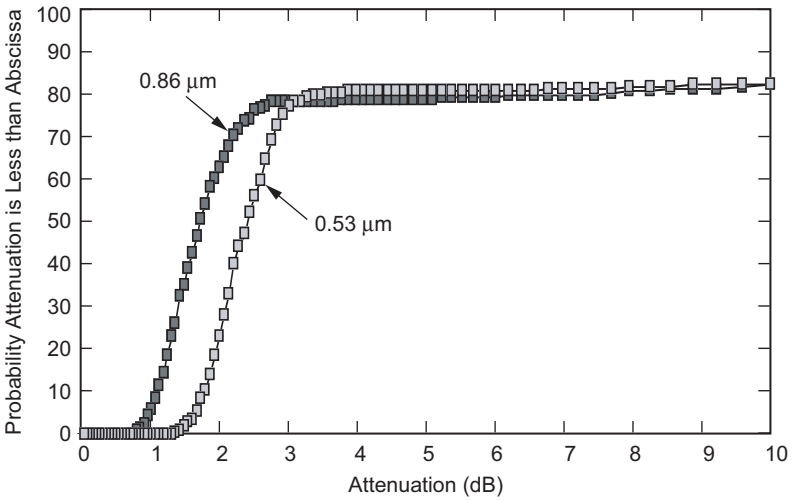


Fig. 1-26. Sample of cumulative atmospheric throughput at TMF (May-June 1996).

data, if an attenuation of 2 dB or less is required, optical reception at a wavelength of 0.53 μm will be possible about 20 percent of the time, but if a wavelength of 0.86 μm is used, the link is available about 60 percent of the time.

More recent studies of station availabilities have been developed through a contract with Northrup Grumman-TASC Corp. using updated satellite observation data [112].

1.7 Background Studies

Background light, whether from outside the atmosphere of the Earth or from within it, can become a severe limitation on system performance. Distant background light includes reflected sunlight from planetary surfaces, integrated starlight, and zodiacal light [113–115]. Of these, reflected planetary light is usually the only one of significance. Additionally, there can be scattered light interference, even for spaceborne receptions systems. Such light depends on many configuration parameters and on the orientation of the telescope relative to the Sun. For ground-based reception in the daylight, sunlight scattered by the atmosphere is the limiting factor for the receiver. Models have been in existence for many years for the level of this background source [116–119]. The actual levels of daytime background depend on many factors (including the telescope axis angle off the Sun, the amount of atmosphere through which the telescope is viewing, and the amount of particulate matter in the atmosphere).

The amount of background light that interferes with the signal detection or acquisition/tracking/pointing processes is also a function of the quality of the telescope. As noted above, photon-bucket (i.e., non-diffraction-limited)

telescopes are of strong interest because they are much less expensive to build and maintain than precise diffraction-limited telescopes. But the larger blur circles associated with photon buckets mean that they will admit more dimensions of background light when the source of that light is distributed over space (particularly daytime skylight and background planetary light). Telescopes of a given aperture size will intercept the same amount of optical signal (which basically comes from a point source). That signal can be detected on those telescopes, provided the detectors at the focal planes are large enough to encompass the "blurred" focused spots. But the larger fields of view of larger detectors mean the detectors also capture more dimensions of the distributed background light. Thus, when considering trades for the precision of the telescope optics, it is often a trade between the background noise susceptibility of the system and the cost of the system.

1.8 Analysis Tools

Analysis of optical communications links is facilitated by the availability of a good set of link-performance prediction tools. The earliest JPL studies were based on the photon-counting channel where ideal Poisson-channel statistics dominated. Simple, first-order link calculations could be done by just considering the geometric aspects of the transmitted beam, the size and quality of the receiving aperture, and any prevailing sources of background noise [120,121]. More detailed calculations are facilitated by the use of specially designed computer programs. The first such program was called OPTI [122,123]. It was very basic, considered only ideal photon-counting detection, and ran on a 286-microprocessor personal computer. Later, a modified version of OPTI (called TOLER) was developed that had more detection options and included a very important feature to account for parameter tolerances [124]. In classical RF deep-space link designs, it is important to not only know the designed values of the link parameters, but to perform an analysis that can justify the amount of link margin required as well. Too little link margin means that if adverse tolerances or conditions mount up together, there might not be enough signal strength to adequately close the link. Too much link margin means that the system is over-designed and that unnecessarily large mass, power, or size (which all translate to increased cost) penalties were imposed on the system. Given the difficulty of deep-space communications and the tight budgets of the space program, managing these margins proactively is important.

As the optical communications program evolved, it became clear that there were many more modulations, codes, laser types, detectors, and background noise sources that needed to be included. Accordingly, a program called Free-Space Optical Communications Analysis Software (FOCAS) was developed [125]. Through several new-release updates, it has been the workhorse of the

optical communications group for the past 15 years. Recently, the program has been ported to a web-based platform for a more general user community.

1.9 System-Level Studies

1.9.1 Venus Radar Mapping (VRM) Mission Study

In 1983 the first JPL mission application system study was done. It involved the Venus Orbiting Imaging Radar (VOIR) mission, later to be known as the Venus Radar Mapping (VRM) mission, and then finally as the Magellan mission. The purpose of the mission was to map the surface of Venus using an imaging radar. The spacecraft used a Voyager 3.7-m antenna, and the operational concept of the basic mission was to use that antenna as the transmit/receive antenna while radar mapping, and then to rotate the spacecraft to point the antenna toward Earth for relaying the captured radar data back to the DSN at X-band. To transmit data to Earth while mapping the planet would require a second antenna separately articulated on the spacecraft so it could point to Earth. This was clearly not practical. However, a much smaller but more capable self-contained and gimbaled communications terminal would allow simultaneous operation and would not require spacecraft attitude rolls for each communications pass. This seemed to be a natural application for an optical communications terminal. A study was conducted and concluded that a 98-kg flight terminal could return 4 Mbps from Venus to a 5-m ground receiver (the Mount Palomar, California, 5-m telescope was used in the study as a reference receiver). A drawing of the optical communications terminal is shown in Fig. 1-27, and a sketch of the spacecraft with the attached optical communications terminal is shown in Fig. 1-28. Although very desirable from a mission operational perspective, it was felt that the development of the flight terminal was too premature to meet the requirements for the mission launch schedule, so the decision was made not to pursue the development.

1.9.2 Synthetic Aperture Radar-C (SIR-C) Freeflyer

In 1994, the Space Shuttle flew the third flight of a synthetic aperture radar mission (SIR-C) in Earth orbit. Because of the success of the flight, there was a strong interest in a long-duration orbital mission called the SIR-C Freeflyer, but there was concern about how the data could be handled. On the SIR-C Shuttle flight, the data were recorded on magnetic tapes. To operate the SIR-C instrument on a long-duration flight mission, a high-data-rate link from the spacecraft to the ground would be required. Here again, optical communication was a natural solution. A link from Earth orbit to the ground, using the 10-cm-diameter OCD terminal on the spacecraft and a 1-m-diameter telescope on the ground, could easily transfer multiple gigabit/second (Gbps) data streams.

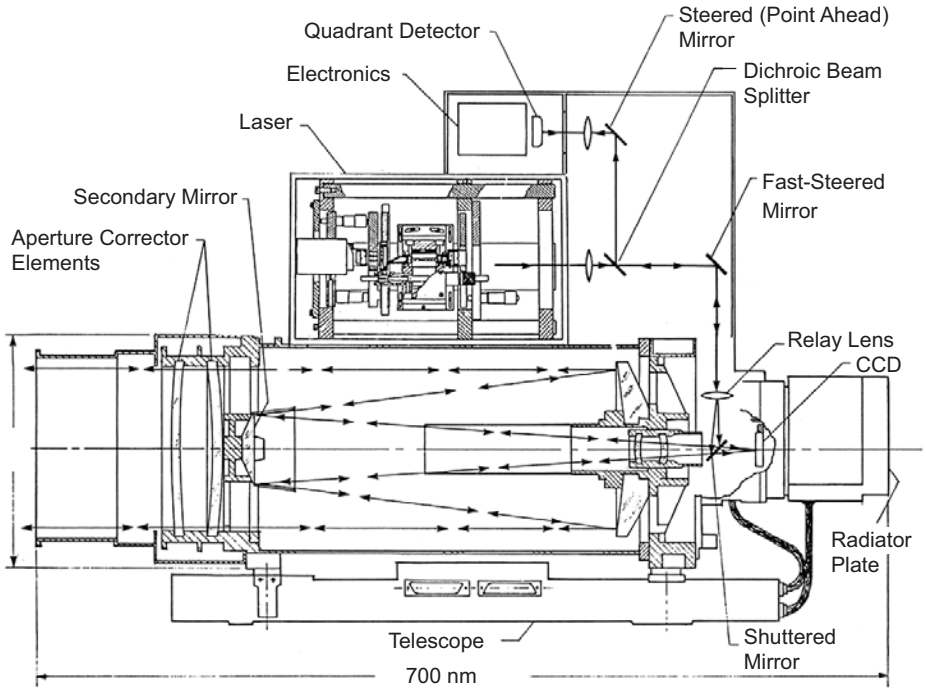


Fig. 1-27. Design of the VRM optical communication terminal.

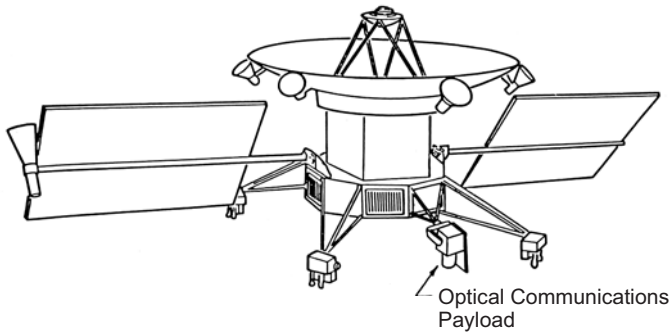


Fig. 1-28. VRM (Magellan) spacecraft with proposed optical terminal.

However, there was still concern that the optical communications technology was not adequately mature.

1.9.3 ER-2 to Ground Study

Given the experiences with the earlier OPTRANSPAC study targeted for Cassini and the earlier VRM flight terminal study, it was felt that some form of precursor flight demonstration of the optical communications capability would be needed. The targeted precursor was an aircraft-to-ground demonstration.

Initially, it was thought that the SR-71 aircraft (of which NASA had three) would be the best platform. However, after visiting the aircraft at NASA Dryden, and then viewing the ER-2 (NASA version of the U-2) aircraft at NASA Ames Research Center, it was concluded that the ER-2 could accommodate the payload much more easily. A detailed study of the use of the ER-2 aircraft with a flyable version of the OCD optical terminal was conducted [126]. The terminal would be mounted in the ER-2 "Q-bay" and would have to be operated without direct operator control.

The study also included provisions for an even earlier flight on the NASA DC-8 aircraft where operators could intervene if needed. Interest was strong in such a demonstration, but the interest dissipated when the ISSERT flight demonstration program awards (see Section 1.4.5) were announced, and it was felt that an air flight demonstration was no longer needed since a space flight demonstration program had already started.

1.9.4 Thousand Astronomical Unit (TAU) Mission and Interstellar Mission Studies

One of the attractions of optical communications is that the transmitted beam from the spacecraft diverges (i.e., dilutes) at a very slow rate as the beam propagates through space. This is of particular interest to really long-distance missions where the $1/R^2$ propagation losses can be enormous. An opportunity to see the real benefits of optical communications came up when JPL performed a study for a Thousand Astronomical Unit (TAU) mission (1 astronomical unit = mean Sun–Earth distance = 150 million km). Strong interest existed in such a mission because it was believed that the Oort cloud resided there and was essentially the factory for producing our Solar System's comets. Furthermore, such a mission was viewed as an excellent interstellar precursor mission. A study of the TAU mission was conducted, and it included an optical communications terminal for the return of data [127,128]. The design being considered would enable 20 kbps from a distance of 1000 AU. A drawing of the payload portion of the TAU spacecraft, which contains the optical communications terminal, is shown in Fig. 1-29.

The basic approach used in the TAU mission study was later applied and extended in a study for a full-interstellar mission to Alpha Centauri. Data rates up to 10 bps were projected from Alpha Centauri at 4.3 light years (271,000 AU or 40 trillion km) [129].

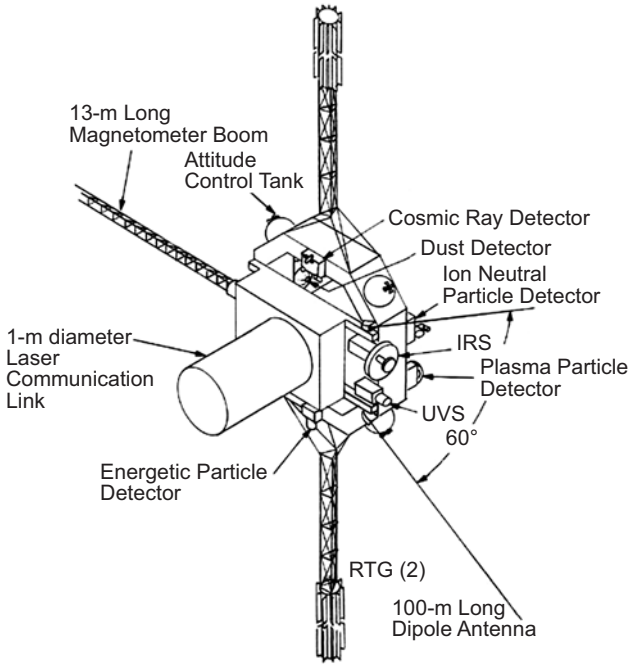


Fig. 1-29. TAU mission spacecraft design showing the optical terminal.

1.10 System-Level Demonstrations

1.10.1 Galileo Optical Experiment (GOPEX)

Although there were disappointments associated with the tight budgets and consequent cancellations of flight system developments, there were nevertheless some very successful and highly rewarding system-level demonstrations accomplished. One of these was performed in December of 1992 and involved the Galileo spacecraft. Recall that the similarities between an optical communications terminal and an imaging camera are quite high. In particular, a telescope with a focal-plane array constitutes a large part of the acquisition and tracking system for an optical communication terminal. Using this fact, a demonstration was conceived that used the imaging camera on the Galileo spacecraft to do an uplink optical communication demonstration.

The Galileo mission design included a trajectory that consisted of two returns to the vicinity of the Earth for trajectory-change gravity assists before traveling out toward Jupiter. Shortly after the second gravity assist in December 1992, a demonstration of uplink laser transmission to the spacecraft was conducted. The idea was to transmit pulsed laser signals from two optical ground telescopes. One was from the 60-cm telescope at TMF near

Wrightwood, California. The other was from the 1.5-m telescope at the Starfire Optical Range at Kirtland Air Force Base (AFB) in Albuquerque, New Mexico. The primary objective was to demonstrate that, based on spacecraft trajectory predicts and local stellar-mount calibrations of the ground telescopes, an uplink laser signal (simulating an uplink beacon) could be successfully transmitted to the spacecraft. The transmissions were done when the ground telescopes were in darkness. The reception of those laser signals was done by pointing the Galileo high-resolution imaging camera back to the Earth and using it as an optical communications receiver. During the demonstration, the Galileo camera shutter was opened, and the camera was scanned across the Earth in a direction parallel to the Earth's day–night boundary line. By doing this, individual laser pulses from a given ground telescope would fall on different pixels in a straight line across the camera's CCD focal-plane array. Transmissions from the two ground telescopes could be distinguished because they would appear as separated lines of dots at different lateral (i.e., Earth longitude) positions on the camera's focal-plane array. Additionally, the two uplink lasers had different pulse repetition frequencies (20 Hz and 30 Hz), so their spatial periods were different along their associated lines. Once the scan had traversed the angle necessary to cause the focused spots to move fully across the focal-plane array, the camera shutter was closed, and the integrated captured image (consisting of smeared Earthshine beyond the day–night terminator and the vertical lines of dots in the dark region of the array) was transmitted to the ground over the standard X-band RF communications link for processing at JPL's Mission Image Processing Laboratory. The resulting images were then analyzed to determine the uplink detection performance.

The demonstration, dubbed Galileo Optical Experiment (GOPEX) was conducted over a narrow window of eight nights, and one of those nights was a non-demonstration night because the spacecraft was scheduled to support other activities. Successful uplink detections occurred on all seven of the actual demonstration nights [130–138]. On certain nights both stations were available, and their signals were detected. On other nights, either one of the two stations was not operating due to poor weather or the signal transmitted by it was not detected due to excessive cloud attenuation. Still, the fact that successful detections were achieved on each night showed that open-loop predict-based uplink pointing was possible. It also showed the value of spatial diversity to circumvent weather outages.

Figure 1-30 shows a sample image from one of the experiment days. Clearly evident in the image are smeared Earthshine and the two rows of laser-pulse detected dots. Note the different pulse detection periods associated with the individual stations. The first demonstration night occurred when the spacecraft was 600,000 km from the Earth. Successful detections on the last demonstration night were when the spacecraft was 6,000,000 km from the Earth. One of the key findings of the demonstration was that although the

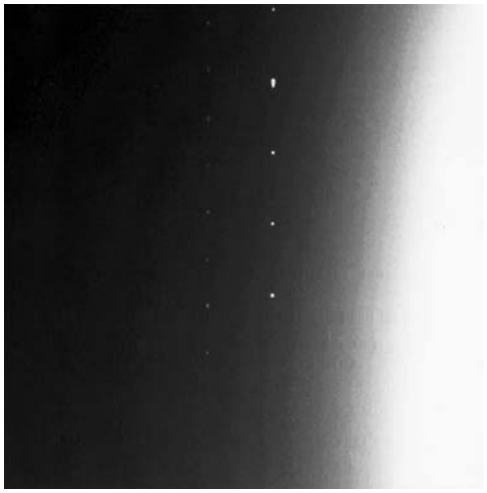
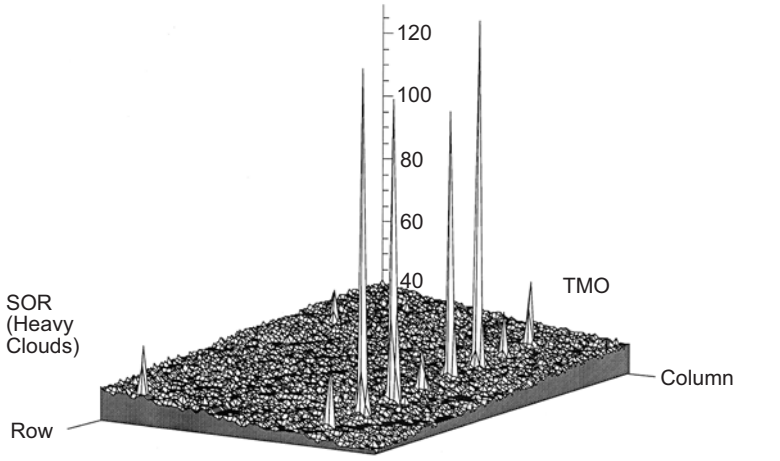


Fig. 1-30. Earth image from Galileo spacecraft showing laser pulses (vertical rows of points).

pointing of the uplink beam was accurate enough to intercept the spacecraft, the scintillation on the uplink beam caused by atmospheric turbulence was quite severe. Figure 1-31 shows an intensity distribution of the detected pulses on one of the nights. This figure made it clear that some form of atmospheric turbulence mitigation would be needed to provide a stable intensity beacon signal at the spacecraft.

1.10.2 Compensated Earth–Moon–Earth Retro-Reflector Laser Link (CEMERLL)

The intensity variations of the uplink generated increased interest in adaptive optics (AO) for mitigating some of the effects of the atmospheric turbulence. The AFRL had just installed a new 3.5-m telescope at Starfire. Furthermore, they had an AO system already connected to their 1.5-m telescope at that facility. Through another cooperative arrangement with AFRL, a Compensated Earth–Moon–Earth Retro-reflector Laser Link (CEMERLL) demonstration was initiated. Laser signals were transmitted from the 1.5-m telescope toward the corner cube arrays that were left on the Moon by the Apollo astronauts. The 3.5-m telescope was used to collect the retro-reflected return signals. Initially, the uplink transmission did not use the AO system. Later the AO system was turned on with the aid of an artificial laser-beam-induced guide star. No discernible signal was detected when the AO system was not used. This was due to the fact that the atmospheric turbulence was breaking up the uplink signal and causing enough scintillation that the returned signal was too weak. When the AO system was engaged, significant return signals were detected at times [139].



SOR = Starfire Optical Range TMO = Table Mountain Observatory

Fig. 1-31. Typical pulse-height variations for laser signals received by Galileo spacecraft at a range of 2.2×10^6 km.

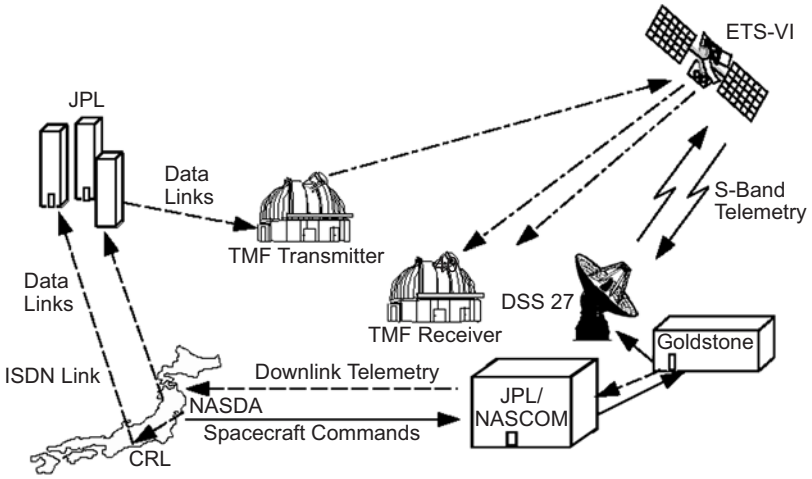
1.10.3 Ground/Orbiter Lasercomm Demonstration (GOLD)

In 1995 another space–ground optical communications opportunity came along. The Japanese had launched the ETS-VI spacecraft, an engineering test satellite, and on it was mounted a small laser communication terminal. The spacecraft was to be parked at GEO over Tokyo, and space–ground demonstrations were to be conducted between the satellite and a 1.5-m telescope at the Tokyo-based Communications Research Laboratory (CRL), a facility of the Japanese Ministry of Posts and Telecommunications. However, after getting into the elliptical transfer orbit required to reach GEO, the GEO-stabilizing rocket motor on the spacecraft failed. This left the spacecraft in the transfer orbit in which it reached to GEO-height altitudes over many countries. Through negotiations between NASA and the Japanese space agency (NASDA), an agreement was reached to do a cooperative space–ground demonstration using their spacecraft and ground telescopes at JPL's TMF.

Because of its orbit, the spacecraft passed over TMF at high altitude approximately every third night. In November of 1995, and after studying the details of the spacecraft and implementing the necessary equipment at Table Mountain (including a 14.5-W argon-ion laser), the Ground/Orbiter Lasercomm Demonstration (GOLD) operational phase commenced [140–145]. Every third night a 4–6 hour pass occurred, during which both uplink and downlink transmissions were made. The operational mode was for the ground station to send up a beacon (argon-ion laser) signal to the spacecraft. If the spacecraft saw the beacon, it would send down a laser signal using the beacon as a pointing reference. Once two-way beacon tracking was established, data modulation (at

1 Mbps) could be imposed on the uplink, on the downlink, or in a turn-around uplink–downlink mode. Sensors were located on the spacecraft to monitor many of the terminal operational parameters, including the uplink detected power levels. Similar instrumentation was installed at the ground. Uplink transmissions were accomplished using the TMF 0.6-m telescope that had been used for the GOPEX demonstration. Downlink signal reception was done at a neighboring 1.2-m telescope at TMF. Spacecraft terminal telemetry and performance data were transmitted via radio link to the DSN and then via NASA Communications (NASCOM) ground circuits to JPL before forwarding to NASDA and CRL. Processed data from those telemetry streams were then forwarded to TMF. The time delay in receipt of the processed spacecraft telemetry data at TMF was about 15 seconds. Having near-real-time data concerning the spacecraft terminal's performance was extremely helpful when conducting the demonstration activities. A diagram showing the end-to-end demonstration data flow is shown in Fig. 1-32.

The demonstration was conducted every third night from November 1995 through May of 1996, except for nights when the weather was bad at TMF. Actually, as the 6-month demonstration phase progressed, the times of the demonstration passes became later and later, until the end of the pass was well into the daytime. This provided experience with the additional effects of daytime sky-background interference. On almost all of the days, two-way lock-up of the links was achieved. And, on many of those days, long periods of solid signal strengths were observed. However, on other days the signals were



CRL = Communications Research Laboratory (Japan)
ISDN = Integrated Services Digital Network (NASA)
NASDA = National Space Development Agency (Japan)

Fig. 1-32. Data flow diagram for the GOLD demonstration.

observed to fade in and out. It was very difficult to diagnose all the variations, but it was generally believed that these variations were caused by attitude fluctuations of the spacecraft coupled with imprecise spacecraft adjustments for those fluctuations. Even at times when the signal strengths were steady, the bit error rates measured on the links were not always stable. The reasons for such variations were not isolated. Then again, there were also long periods when error-free communication was achieved. Figure 1-33 shows a sample of the demodulated data stream during one such occasion when the data reception quality was very good.

One of the key lessons learned from the GOPEX demonstration was that the uplink beacon often contains significant amounts of scintillation due to atmospheric turbulence (if not corrected). In the GOLD demonstration, there was no AO system installed. However, one of the objectives of the demonstration was to show that a multiple-beam uplink beacon signal could significantly reduce that scintillation. If a single beam is transmitted up through the atmosphere, the atmospheric turbulence will break that beam up into smaller beam segments that independently move around in angle due to local changes in the atmospheric refractive index. When these beam segments overlap at the spacecraft target, they can either combine in-phase or out-of-phase. In-phase events will cause a surge in power whereas out-of-phase events will cause severe signal fades. Such interference effects can give rise to very large fluctuations in the signal as seen at the spacecraft. However, if the beams are broken up into a bundle of beams before they enter the telescope, and if they are caused to be noncoherent relative to one another, then any overlapping of the beams at the spacecraft will result in an addition of the powers of the two beams. Furthermore, the probability of a really deep scintillation fade is very much reduced since it only occurs if all of the component beams simultaneously deflect away from the direction of the spacecraft (a much more unlikely event).

To implement this, the signal from the argon-ion laser was split using proportional beam splitters into two beams. Then one beam was delayed

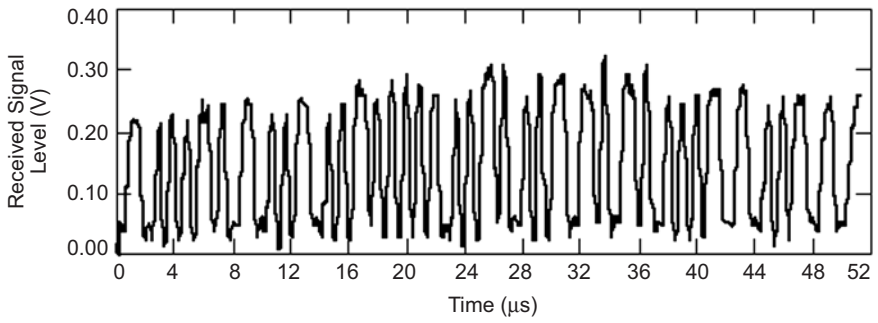


Fig. 1-33. Demodulated data patterns during the GOLD demonstration.

relative to the other so that it was outside the coherence length (and hence independent) of the other beam. Then each beam was sent up to different sub-aperture portions of the telescope primary mirror. Later, each of the two beams was in turn split (and delayed) so that a noncoherent four-beam uplink resulted. Figure 1-34 is a photograph of the uplink signal leaving the telescope when the four-beam transmission was used. Figure 1-35 shows uplink intensities as measured at the spacecraft for a single uplink beam of full power, a dual-beam uplink with half the power in each sub-beam, and finally a four-beam uplink with each beam containing 1/4 the power. As is clearly evident, the two-beam uplink has somewhat reduced the occurrence of really deep signal fades, whereas the four-beam uplink has very significantly reduced those uplink signal strength fades. Also shown in Fig. 1-35 are the histograms of the measured signal strengths at the spacecraft where these effects are even more evident.

1.10.4 Ground–Ground Demonstrations

Additional system-level demonstrations have taken place between ground stations. These have involved use of the 0.6-m telescope at Table Mountain and the Optical Communications Demonstrator engineering model terminal. Table Mountain is located at 2272 m (7400 ft) elevation in the eastern San Gabriel Mountains. The terrain drops off rapidly to the east, reaching down into the Cajon Pass, and then rises again at the western edge of the San Bernardino Mountains. A U.S. Forest Service tower is located at Strawberry Peak, a mountain peak on the eastern side of the pass. The 46-km line-of-sight path between Table Mountain and Strawberry Peak is an ideal place to conduct simulated space-to-ground optical links since most of the path is high above the

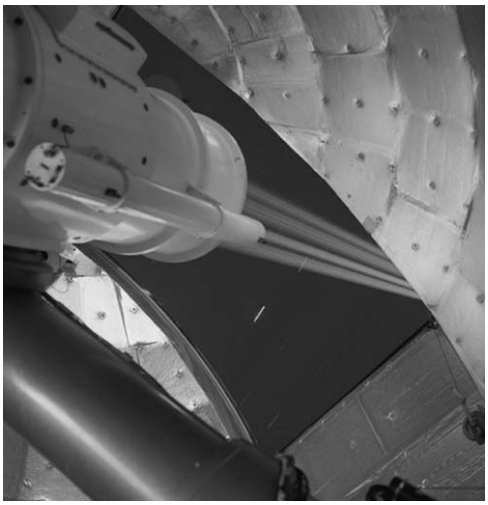


Fig. 1-34. Photograph of the four-beam uplink.

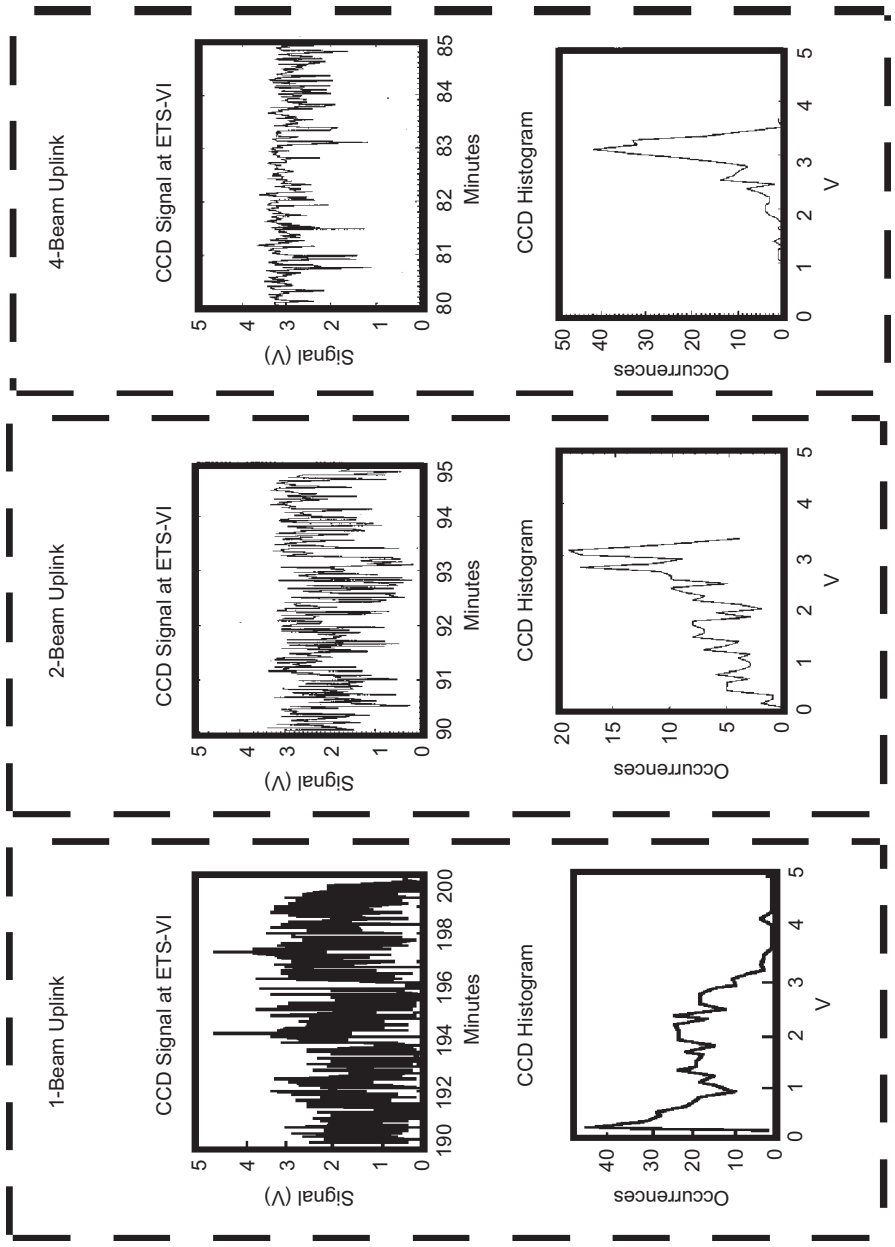


Fig. 1-35. Uplink signal intensities and intensity histograms for one-, two-, and four-beam uplinks.

ground and minimizes the horizontal path turbulence. The turbulence, although much less than that of most horizontal paths of the same length, is more characteristic of the worst case for typical space-to-ground links. A cross section of the path between these mountain peaks showing the height of the beam above the ground is provided in Fig. 1-36. A view of Table Mountain from the Strawberry Peak site is shown in Fig. 1-37.

Three separate experimental campaigns were conducted from June 1998 through September of 2000 [146–149]. The first two of these concentrated on

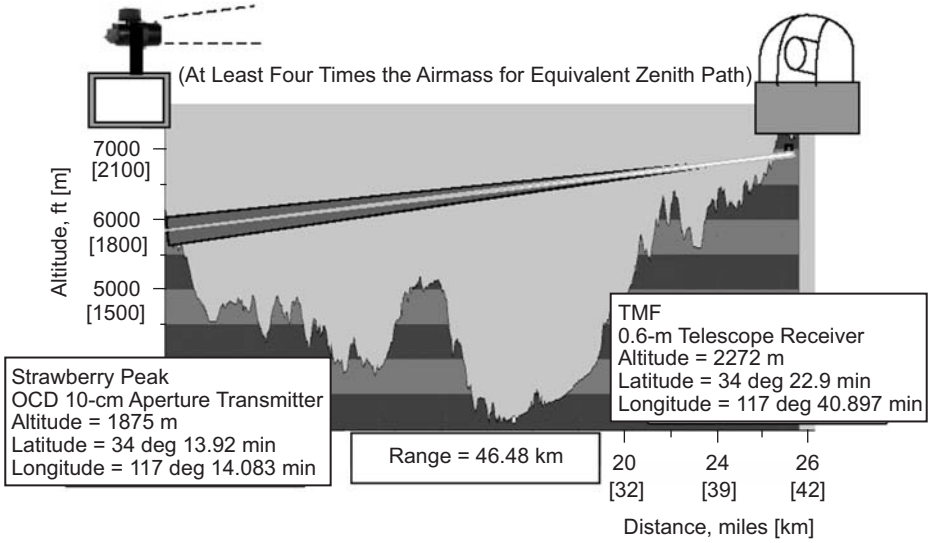


Fig. 1-36. Cross section of the path between TMF and Strawberry Peak.

Fig. 1-37. View of TMF from Strawberry Peak

characterizing the atmospheric turbulence and beam-wandering parameters for the link. The third focused on the performance of the OCD spatial acquisition and tracking system performance as well as the transmission of 400-Mbps data across the link. Measurements made during the campaign showed that uncompensated tracking jitter of ±1.1 μrad in the horizontal direction and ±2 μrad in the vertical direction could be achieved.

1.11 Other Telecommunication Functions

In standard RF deep-space communications systems, the signal is routinely used for more functions than just communications. Both navigation parameters and scientific measurements are integrated into the signal design for the link. For navigation, two-way ranging to the spacecraft, as well as the Doppler shift of the coherently tracked carrier signal, provide navigational tracking measurements. Two-way ranging can provide both distance to the spacecraft and (by differencing subsequent measurements) the velocity component of the trajectory. Doppler tracking, in conjunction with the Earth's rotation about its axis, provides measurements of the angular location of the spacecraft in the sky. These angular measurements are often augmented by delta differential one-way ranging (DDOR) fixes for added angular measurement accuracy. DDOR uses pairs of intercontinentally spaced DSN tracking stations to make interferometric measurements of the arriving RF signals, first from the spacecraft, and then from an angularly close radio source (a distant quasar). By differencing these measurements, many of the error sources (e.g., station clock offsets, uncertainties of the station locations, and atmospheric delays) common to both measurements are cancelled out, leaving highly precise measures of the spacecraft location relative to the reference quasar positions. Angular measurement accuracies of spacecraft locations are often made to the 5-nrad accuracy level in this way.

Conventional RF communications system signals are also used to make "radio science" measurements. As the signal propagates back from deep space, its phase, amplitude, and/or polarization may be altered by things encountered in the propagation path. For example, as a spacecraft passes behind a planet, the return signal can be altered as it passes through the atmosphere of that planet. Another example is the phase jitter that can be imparted on the RF signal by charged-particle fluctuations from the solar wind as the signal passes close to the Sun.

1.11.1 Opto-Metric Navigation

It is desirable to be able to accomplish other functions with an optical communications signal as well, especially navigation measurements. An easy navigation measurement to make is that of two-way ranging. To accomplish this, a laser pulse can be sent on the uplink (i.e., on the beacon or on the

command uplink) and detected on the spacecraft. The detected pulse is then used to trigger a pulse from the downlink laser. By measuring the time delay from the time the uplink pulse was generated to the time the downlink pulse is detected, and if the delays in the electronics of the spacecraft and ground systems are constant and calibrated, then the two-way path delay and hence distance can be calculated. This approach was included in the X2000 terminal development mentioned previously.

Angular measurements can also be made with optical signals [150–155]. Recall that for RF systems, multiple ground stations are used to interferometrically make angular measurements of the spacecraft location relative to the quasar reference catalog. With a ground-based optical telescope, the signal from the spacecraft can be focused onto a focal-plane detector array. The spot on the focal plane represents the location of the spacecraft, relative to the unknown axis of the telescope. However, if simultaneously the light from a stellar object is also collected, its energy will be concentrated onto a different location on the focal-plane detector. This spot represents the location of the star relative to the unknown telescope axis direction. Since both spots are on the same focal-plane detector at the same time, the vector difference between the two spot locations represents the angular offset of the spacecraft signal relative to the (catalog) star, irrespective of the actual axis of the telescope. Thus, for optical spacecraft-signal tracking, equivalent angular measurements to the RF interferometric systems can be made. In fact, they only require a single ground telescope. Furthermore, whereas the current RF DDOR technique requires independent measurements of the horizontal and vertical components of angles, an optical tracking telescope can make measurements in both directions at the same time. This technique has been used in optical astrometry, and it has yielded angular offsets between star pair measurements to accuracies of 5–10 nrad [156].

1.11.2 Light Science

In principle, scientific measurements should also be possible using optical communications signals, although more work needs to be done with the scientific community to fully develop these claims. Such measurements will not duplicate the specific measurements made today with RF systems, but will undoubtedly be complementary to them. One such example is the optical equivalent to RF occultation measurements. As a spacecraft moves behind a planet, the signal from its optical transmitter also passes through the planetary atmosphere. Measurements of the perturbations of that signal might reveal attributes of that atmosphere. Furthermore, the optical signal fluctuations will likely be much more sensitive to higher-altitude components of the atmosphere than are detectable with radio signals. As optical communications technology develops, and the use of the technology on future missions becomes more

likely, it is believed that the scientific community will begin to recognize the scientific opportunities of this form of “Light Science.”

1.12 The Future

This past experience base provides a springboard for many of the planned activities of the future. These are both developmental activities as well as some exciting system demonstrations. Many of these will use the infrastructure already created, whereas others will result in the development and validation of new systems, tools, and techniques.

1.12.1 Optical Communications Telescope Laboratory (OCTL)

One of the key infrastructure elements recently created is the Optical Communications Telescope Laboratory (OCTL) [157–159]. Located at JPL’s TMF, OCTL will be the main ground support facility for a number of planned free-space optical communications demonstrations. Although there are a number of telescopes already at TMF, they are not well suited for use in the emerging set of planned demonstrations. Most of the current telescopes have inadequate space in their focal planes to accommodate the optical and electronic systems needed for planned future demonstrations, and none of the telescopes was designed for use during the daytime. The OCTL telescope is a 1-m-diameter diffraction-limited telescope that has a coudé focus and four coudé instrument rooms. Separate demonstrations can be set up in each of the coudé rooms. The telescope axis can be connected to one of these rooms by a coudé-room fold mirror (designated as M7). This will allow the telescope to be used while other demonstration setups are being installed in other rooms. Furthermore, the telescope is designed to operate within its diffraction-limited wavefront error tolerances down to solar offset angles of 30 percent. Although its wavefront errors will be degraded at smaller solar angles, the thermal control system will allow it to function at even smaller solar angles. Additionally, the telescope mount is capable of precision tracking of low-altitude satellites. This will allow it to support demonstrations that are relevant to near-Earth applications as well as deep-space applications. A photograph of the OCTL facility is shown in Fig. 1-38. A picture of the telescope assembly, with a clear view of the thermal control veins, is shown in Fig. 1-39.

1.12.2 Unmanned Arial Vehicle (UAV)–Ground Demonstration

One of the planned early demonstrations involves optical communications from an uncrewed airborne vehicle and the ground. Funded by the United States Missile Defense Agency, this activity will fly a modified version of the OCD terminal called the Optical Communications Terminal (OCT) on a Predator B unmanned aerial vehicle (UAV). The OCT will be outfitted with a

Fig. 1-38. The OCTL facility.

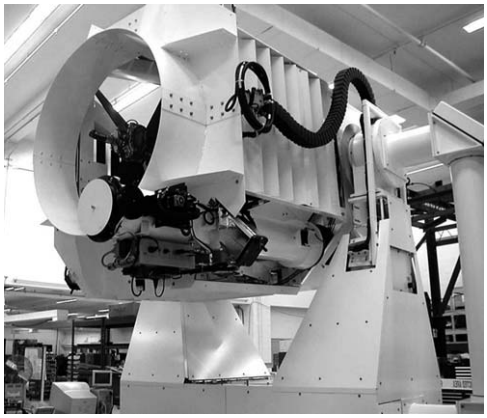


Fig. 1-39. The OCTL telescope.

1550-nm laser transmitter and will transmit at 2.5 Gbps from the UAV to the OCTL telescope. Reference [160] describes the UAV–Ground Demonstration program.

1.12.3 Adaptive Optics

One of the key technologies to be validated in the OCTL will be AO. As mentioned above, turbulence in the atmosphere can cause significant beam wander and intensity fluctuations on uplink beacon or command signals sent to distant spacecraft. Additionally, turbulence causes broadening of focused signal energy at the focal planes of ground-based receive telescopes. This broadening can drive the requirements for the detectors and result in increased susceptibility to background light interference. Under internal funding, JPL has been examining the use of AO techniques for optical communications.

AO techniques have been used in the past and are becoming commonplace on many astronomical observatories. Indeed, the JPL work is building directly

on experiences gained by implementing AO on the Palomar and Keck telescopes [161–165]. These techniques have been used to sharpen images and enable astronomers to distinguish closely spaced celestial objects. Figure 1-40 is a diagram of the AO system on the Palomar 5-m telescope. It is located at the telescope's Cassagrain focus of the telescope, photographically shown in Fig. 1-41.

However, unlike astronomy observatories, optical communication ground stations must also operate in the daytime. This exacerbates the levels of turbulence that must be accommodated. Additionally, the overall objective is different for optical communications relative to astronomical observations. For astronomy, the objective is to increase the sharpness of images so that the finest

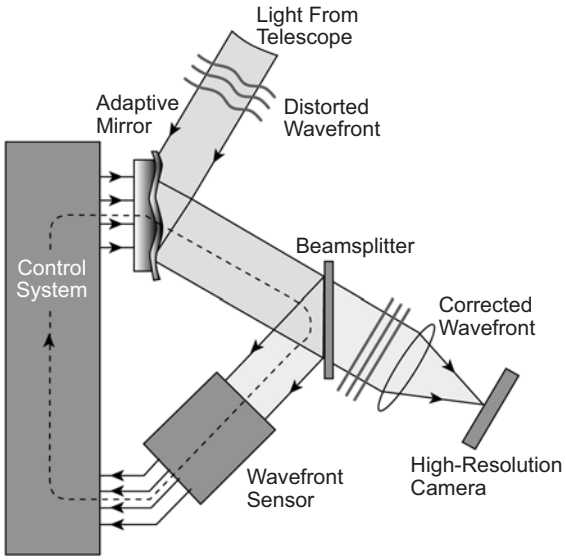


Fig. 1-40. Palomar AO layout.

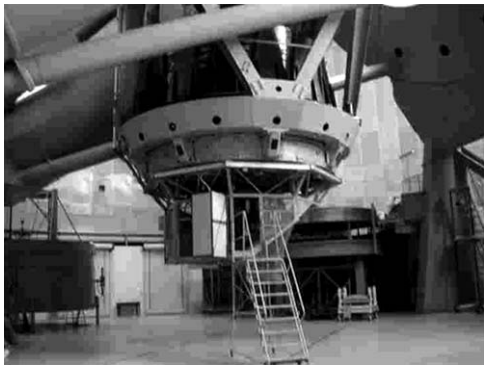


Fig. 1-41. Palomar Cassegrain focus.

details can be observed. If in the process there is a loss of signal energy, then that loss can be made up by just observing longer. For communications, the signal energy devoted to a given data bit is fixed, and must be conserved as much as possible. Thus, the optimization function for an optical communication AO system is to minimize the overall field of view (to minimize the amount of background light admitted) while maximizing the amount of desired signal energy captured (for the most robust signal detection) [166].

For uplink beacon and command links, the multi-beam transmission technique mentioned in the GOLD demonstration can be used to reduce some of the beam intensity fluctuations. Increased transmit power is easier to generate on the ground so that the beams can reach further into space. But, to reach even farther, or to reduce the uplink power requirements for a given distance, uplink AO will be beneficial. Such systems will likely rely on artificially generated laser guide-star calibrators to accomplish the uplink signal adaptation [167,168]. Both downlink signal-to-noise-ratio improvement and uplink beam-adaptive pre-distortion techniques can be validated using OCTL.

1.12.4 Optical Receiver and Dynamic Detector Array

Work is also continuing on the development of optical communication receivers to work at the focal plane of the reception telescopes. Such receivers must process electronic signals received from the downlink signal detectors and process them to extract the data modulation and the required temporal synchronization signals. Synchronization includes the recognition and dynamic tracking of slot synchronization (i.e., tracking the boundaries of the short intervals into which actual pulses could be placed, but which, due to the PPM modulation, are infrequently actually there) and PPM word synchronization (the places in the sequence of slots that define the beginning of a PPM symbol and from which the receiver must measure to extract the data bits associated with a given pulse). Additionally, higher levels of synchronization (such as frame or packet synchronization) may have to be accomplished in the receiver (or if not there, then certainly in the subsequent processing systems). Data detection assumes that synchronization has already been achieved and involves optimally processing the signals output from the detector to convert them to data bits that can be fed into the channel decoding system for error correction. Often the receiver is also required to provide soft (i.e., confidence) information to the decoder to improve its error-correction performance.

Another approach for efficient signal reception is an adaptive processing receiver that combines the photo-detection process and the front-end functions of the electronic signal processing [169–174]. This approach uses a focal-plane detector array whose outputs are combined through a weighting network. As the energy of the received signal is concentrated on the focal-plane array, local “hot spots” occur due to atmospheric turbulence. This receiver senses the

regions on the focal plane where the signal is the strongest and weights those contributions more heavily than regions where the signal is weaker. Such a receiver can be used as an electronic form of AO, or it can be used in conjunction with an AO system to further compensate for residual turbulence-induced beam fluctuations. A photograph of such a detection and processing system is shown in Fig. 1-42.

1.12.5 Alternate Ground-Reception Systems

Work is also under way to assess alternate architectures for ground-based reception telescopes. In 2001 a JPL internally funded study was started to examine the use of a collection of smaller telescopes to act effectively as a single large telescope. Initial results indicate that, for ground-based reception systems, arrays of small telescopes, each with its own focal-plane detector system, can be an attractive alternative to large single-aperture-reception telescopes, especially if each telescope includes a focal-plane detector array for electronically tracking the atmospheric turbulence-induced "hot spots" [175–178]. Given these preliminary findings, an experimental program was initiated to validate the projections of such an array. A JMI Inc. 63-cm-diameter New Technology Telescope (NTT) was procured, and initial tests have begun. Although the results are promising, the ultimate conclusions will depend on a thorough understanding of the performance characteristics of both large single-aperture and arrayed smaller-aperture telescope architectures, as well as complete life-cycle-cost analyses of both approaches. A depiction of a reception array using 63-cm NTTs is shown in Fig. 1-43.

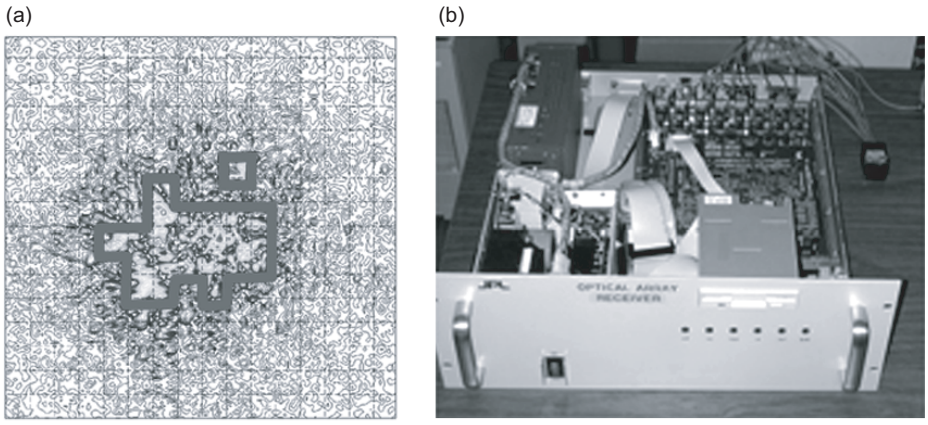


Fig. 1-42. 16×16 focal-plane array showing (a) higher-intensity pixels and (b) the 16×16 dynamic signal combiner.

Fig. 1-43. A large-aperture telescope made from an array of 63-cm telescopes.

1.13 Mars Laser Communication Demonstration

Over the past 25 years the key question has been, “When will optical communications be mature enough to use operationally?” The activities above have contributed significantly to the development and understanding of the technology, but there have been many major parallel development successes worldwide as well. In 1998, the European Space Agency launched an experimental optical communications terminal on the Satellite Probatoire d’Observation de la Terre (SPOT 4) French Earth observation satellite [179]. In 2001, they launched the companion terminal on the ARTEMIS satellite and demonstrated 50-Mbps links from the SPOT 4 in LEO to ARTEMIS in GEO [180]. The Japanese are also developing a terminal similar to the SPOT 4 terminal and hope to accomplish a similar demonstration to ARTEMIS [181]. The United States Defense Department has had a very successful flight demonstration involving the Geosynchronous Lightweight Technology Experiment (GEOLite) spacecraft. These and many other past studies, developments, and demonstrations have resulted in serious consideration of optical communications for the next generation of transformation communications systems. This intensified interest in the technology has led NASA to formulate a major initiative to demonstrate optical communication for deep-space applications.

To accomplish this, the Mars Laser Communications Demonstration (MLCD) project has been started. MLCD is a joint project of NASA Goddard Space Flight Center (GSFC), JPL, and Massachusetts Institute of Technology (MIT) Lincoln Laboratory. The basic objective is to demonstrate optical communications from Mars to the Earth at data rates from 1 through 10 Mbps, depending on specific conditions [52,182]. Indeed, under some conditions and if a 10-m-diameter ground-receiving telescope were available, data rates in the region of 100 Mbps would even be possible. The MLCD project will fly an optical communications terminal on the Mars Telesat Orbiter (MTO) that is planned for launch in 2009. An uplink laser signal will be transmitted to the spacecraft to act as a pointing reference beacon. The spacecraft terminal will track the beacon uplink and return a PPM-modulated downlink laser signal to the ground. Additionally, the flight terminal will have an inertial reference unit that will allow the spacecraft terminal to track a lower-power uplink beacon signal. This will simulate an even longer link and show that inertial reference systems can take up part of the tracking burden that relied on strong beacons in the past.

Several options are being assessed for the ground system demonstration support. Single large-aperture telescopes, both existing telescopes and possibly new ones, are being studied. Additionally, an array architecture is also being studied. The specific choices for this demonstration will depend on both technical performance and overall project cost constraints. But whatever the choices are, the demonstration promises to be a key milestone in the development of optical communication for deep space. Furthermore, the studies and developments associated with the MLCD project, when combined with thorough future mission load analyses and long-term infrastructure life-cycle-cost analyses, will provide valuable data for the subsequent definition and justification for the optical portions of the future Interplanetary Network [1].

1.14 Summary of Following Chapters

The remaining chapters of this book describe various aspects of deep-space optical communication systems in detail. They will provide both theoretical and practical considerations required in the realization of these future systems.

Chapter 2 begins with an overview of the end-to-end optical communications system, and then segues into some of the key design drivers for such links. One of these is the choice of the operating wavelength (equivalent to frequency selection in RF communication). Then, the Link Design Control Table (LDCT) is introduced, and a sample LDCT is given.

Chapter 3 presents the key attributes of the atmospheric channel. These include the effects of beam propagation, weather availability of links, the various background noise sources that degrade the optical channel, and how these factors affect choices for a potential optical deep-space network.

Chapter 4 summarizes efficient modulation and coding systems. The chapter begins with a discussion of the statistical models that describe the various methods of optical signal detection. Next, it presents descriptions of candidate modulation schemes. Given the detector statistics and the specific modulation schemes, the uncoded link performances of some of the more promising modulations with the candidate detectors are described. Given the modulation scheme and the kind of detector used, the resulting channel can be characterized in terms of its channel capacity. This is done for several of the more promising combinations of the two. Then, Chapter 4 describes codes that can be used over the resulting channels defined by the corresponding modulation and detection schemes. These codes are key to achieving data rates that approach channel capacity limits. Finally, it addresses and compares the performance of these coded systems.

Chapter 5 covers the key systems associated with an optical communications flight terminal. This begins with a section on the acquisition, tracking, and pointing (ATP) subsystem. Methods used to remove the errors associated with spacecraft platform jitter and to resolve uncertainties in the absolute direction needed for the transmitted beam are described. Next is a subsection describing laser transmitters. It covers both the methods of generating coherent laser energy and the systems used to modulate that energy. This leads to a subsection on the opto-mechanical subsystem that connects the laser and the ATP subsystems to the flight-unit telescope. It is important to understand both how these can be connected and the error sources that can impact the resulting system performance. Chapter 5 ends with a discussion of issues, challenges, and techniques associated with space qualification of the resulting flight terminal.

Chapter 6 describes the Earth reception terminal for a deep-space optical link. This chapter begins by considering the various types and possible locations of receiving telescopes. Next, the types of photo-detection schemes are described. This leads to a subsection on the remaining subsystems in the optical communications receiver that provide spatial acquisition and tracking, temporal synchronization, and demodulation/detection.

The final chapter provides concluding remarks and a look at the future prospects and expected applications of deep-space optical communications.

The challenges associated with deep-space communication are truly monumental, and the need for significantly increased deep-space communications capacity continues to grow. One of the promising technologies for overcoming these challenges and enabling substantial channel capacity growth is optical communication. This book provides a description of the systems required to achieve these gains and is the product of the collective knowledge of the past quarter century of effort in this field. In the future it is anticipated that optical communication will enable new kinds of deep-space missions to be flown that have heretofore been impractical to consider because

of the difficulty in returning the required data. The resulting scientific discoveries will surely be awesome.

References

[1] W. J. Weber, R. J. Cesarone, D. S. Abraham, P. E. Doms, R. J. Doyle, C. D. Edwards, A. J. Hooke, J. R. Lesh, and R. B. Miller, "Transforming the Deep Space Network Into the Interplanetary network," Paper IAC-03-U.4.01 *54th International Astronautical Congress*, Bremen, Germany, September 29–October 3, 2003.

[2] J. R. Pierce, "Optical Channels: Practical Limits with Photon Counting," *IEEE Transactions on Communications*, vol. COMM-26, pp. 1819–1821, 1978.

[3] J. R. Lesh, J. Katz, H. H. Tan, and D. Zwillinger, "2.5-Bit/Detected Photon Demonstration Program: Description, Analysis, and Phase I Results," *The Telecommunications and Data Acquisition Progress Report 42-66, September and October 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 115–132, December 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[4] W. K. Marshall, "A PPM Modulator and Demodulator for the 2.5-Bit/Detected Photon Demonstration," *The Telecommunications and Data Acquisition Progress Report 42-68, January and February 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 50–54, April 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[5] J. Katz, "2.5 Bit/Detected Photon Demonstration Program: Phase II and III Experimental Results," *The Telecommunications and Data Acquisition Progress Report 42-70, May and June 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 95–104, August 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[6] H. H. Tan, "A Statistical Model of the Photomultiplier Gain Process With Applications to Optical Pulse Detection," *The Telecommunications and Data Acquisition Progress Report 42-68, January and February 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 55–67, April 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[7] J. R. Lesh, "Power Efficient Optical Communications for Space Applications," *Proceedings of 1982 International Telemetering Conference*, San Diego, California, pp. 109–113, September 28–30, 1982.

[8] J. R. Lesh, "Optical Communications Research Program to Demonstrate 2.5 Bits/Detected Photon," *IEEE Communications Magazine*, vol. 20, pp. 35–37, November 1982.

[9] R. J. Stokey and P. J. Lee, "Approximation to the Probability Density at Output of a Photomultiplier Tube," *The Telecommunications and Data Acquisition Progress Report 42-73, January–March 1983*, Jet Propulsion Laboratory, Pasadena, California, pp. 36–39, May 15, 1983. http://ipnpr.jpl.nasa.gov/progress_report/

[10] W. K. Marshall, "Detection and Symbol Synchronization for Multiple-Bit Per Photon Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-80, October–December 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 24–30, February 15, 1985. http://ipnpr.jpl.nasa.gov/progress_report/

[11] V. A. Vilnrotter, E. R. Rodemich, and H. H. Tan, "A Synchronization Technique for Optical PPM Signals," *The Telecommunications and Data Acquisition Progress Report 42-87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 24–31, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[12] R. J. McEliece and L. R. Welch, "Coding for Optical Channels With Photon-Counting," *The Deep Space Network Progress Report 42-52, May and June 1979*, Jet Propulsion Laboratory, Pasadena, California, pp. 61–66, August 15, 1979. http://ipnpr.jpl.nasa.gov/progress_report/

[13] L. D. Baumert, R. J. McEliece, and H. Rumsey, Jr., "Coding for Optical Channels," *The Deep Space Network Progress Report 42-49, November and December 1978*, Jet Propulsion Laboratory, Pasadena, California, pp. 70–77, February 15, 1979. http://ipnpr.jpl.nasa.gov/progress_report/

[14] R. J. McEliece, "Interleaved Block Codes for the Photon Channel," *The Telecommunications and Data Acquisition Progress Report 42-60, September and October 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 98–102, December 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[15] R. J. McEliece, "Practical Codes for the Photon Channel," *IEEE Transactions on Information Theory*, vol. IT-27, no. 4, pp. 393–398, July 1981.

[16] R. J. McEliece, "The R_0 Parameter for Optical Communication Using Photon Counting," *The Deep Space Network Progress Report 42-53, July and August 1979*, Jet Propulsion Laboratory, Pasadena, California, pp. 62–65, October 15, 1979. http://ipnpr.jpl.nasa.gov/progress_report/

[17] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *The Telecommunications and Data Acquisition Progress Report 42-60, September and October 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 68–76, December 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[18] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The Capacity of the Photon Channel," *IEEE Transactions on Information Theory*, vol. IT-27, no. 1, pp. 61–77, January 1981.

[19] H. H. Tan, "Capacity of a Multimode Direct Detection Optical Communication Channel," *The Telecommunications and Data Acquisition Progress Report 42-63, March and April 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 51–70, June 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[20] J. R. Lesh, "Capacity Limit of the Noiseless, Energy-Efficient Optical PPM Channel," *The Telecommunications and Data Acquisition Progress Report 42-67, November and December 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 54–58, February 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[21] J. R. Lesh, "Capacity Limit of the Noiseless, Energy-Efficient Optical PPM Channel," *IEEE Transactions on Communications*, vol. COMM-31, no. 4, pp. 546–548, April 1983.

[22] P. J. Lee, "Capacity and Cutoff Rate of (M + 1)-ary Decision Rules for Noisy M-ary Optical PPM Channel," *The Telecommunications and Data Acquisition Progress Report 42-72, October–December 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 43–50, February 15, 1983. http://ipnpr.jpl.nasa.gov/progress_report/

[23] J. Hamkins, "The Capacity of Avalanche Photodiode-Detected Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–19, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[24] R. G. Lipes,, "Pulse-Position-Modulation Coding as Near-Optimum Utilization of Photon Counting Channel with Bandwidth and Power Constraints," *The Deep Space Network Progress Report 42-56, January and February 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 108–113, April 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[25] D. Zwillinger, "Maximizing Throughput Over an Average-Power-Limited and Band-Limited Optical Pulse Position Modulation Channel," *The Telecommunications and Data Acquisition Progress Report 42-62, January and February 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 81–86, April 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[26] S. J. Dolinar, Jr., "A Near-Optimum Receiver Structure for the Detection of M-ary Optical PPM Signals," *The Telecommunications and Data Acquisition Progress Report 42-72, October–December 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 30–42, February 15, 1983. http://ipnpr.jpl.nasa.gov/progress_report/

[27] P. J. Lee, "Analysis of a Coded, M-ary Orthogonal Input Optical Channel With Random-Gain Photomultiplier Detection," *The Telecommunications and Data Acquisition Progress Report 42-79, July–September 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 107–113, November 15, 1984. http://ipnpr.jpl.nasa.gov/progress_report/

[28] D. Divsalar and F. Naderi, "Performance of an Optical Relay Satellite Using Reed-Solomon Coding Over a Cascaded Optical PPM and BPSK Channel," *The Telecommunications and Data Acquisition Progress Report 42-70, May and June 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 117–131, August 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[29] D. Divsalar, R. M. Gagliardi, and J. H. Yuen, "PPM Demodulation for Reed-Solomon Decoding for the Optical Space Channel," *The Telecommunications and Data Acquisition Progress Report 42-70, May and June 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 47–59, August 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[30] R. J. McEliece and L. Swanson, "An Easy-to-Implement Coding Scheme for Multifrequency PPM," *The Telecommunications and Data Acquisition Progress Report 42-77, January–March 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 57–63, May 15, 1984. http://ipnpr.jpl.nasa.gov/progress_report/

[31] S. A. Butman, J. Katz, and J. R. Lesh, "Practical Limitations on Noiseless Optical Channel Capacity," *The Deep Space Network Progress Report 42-55, November and December 1979*, Jet Propulsion Laboratory, Pasadena, California, pp. 12–14, February 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[32] S. A. Butman, J. Katz, and J. R. Lesh, "Bandwidth Limitations of the Noiseless Optical Channel Capacity," *IEEE Transactions on Communications*, vol. COMM-30, pp. 1262–1264, May 1982.

[33] R. J. McEliece, E. R. Rodemich, and A. L. Rubin, "The Practical Limits of Photon Communication," *The Deep Space Network Progress Report 42-55, November and December 1979,* Jet Propulsion Laboratory, Pasadena, California, pp. 63–67, February 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[34] M. Ross, J. Abernathy, J. Wolf, P. Freedman, G. Matassov, and J. D. Barry, "Space Optical Communications with the Nd:YAG Laser," *Proceedings of the IEEE,* vol. 66, pp. 319–344, March 1, 1978.

[35] J. Katz, "High Power Semiconductor Lasers for Deep Space Communications," *The Telecommunications and Data Acquisition Progress Report 42-63, March and April 1981,* Jet Propulsion Laboratory, Pasadena, California, pp. 40–50, June 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[36] J. Katz, "Phase-Locking of Semiconductor Injection Lasers," *The Telecommunications and Data Acquisition Progress Report 42-66, September and October 1981,* Jet Propulsion Laboratory, Pasadena, California, pp. 101–114, December 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[37] J. Katz, "Power Efficiency of Semiconductor Injection Lasers," *The Telecommunications and Data Acquisition Progress Report 42-66, September and October 1981,* Jet Propulsion Laboratory, Pasadena, California, pp. 94–100, December 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[38] H. Hemmat, "Single Longitudinal Mode Operation of Semiconductor Laser Arrays With Etalon Control," *The Telecommunications and Data Acquisition Progress Report 42-86, April–June 1986,* Jet Propulsion Laboratory, Pasadena, California, pp. 66–69, August 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[39] J. Daher, "Preliminary Results Toward Injection Locking of an Incoherent Laser Array," *The Telecommunications and Data Acquisition Progress Report 42-84, October–December 1985,* Jet Propulsion Laboratory, Pasadena, California, pp. 26–34, February 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[40] J. Katz, "Power Combining of Semiconductor Lasers: A Review," *The Telecommunications and Data Acquisition Progress Report 42-70, May and June 1982,* Jet Propulsion Laboratory, Pasadena, California, pp. 87–94, August 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[41] J. Katz, "Phase Control and Beam Steering of Semiconductor Laser Arrays," *The Telecommunications and Data Acquisition Progress Report 42-68, January and February 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 42–49, April 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[42] J. Katz, "Electronic Beam Steering of Semiconductor Injection Lasers," *The Telecommunications and Data Acquisition Progress Report 42-68, January and February 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 25–41, April 15, 1982. http://ipnpr.jpl.nasa.gov/progress_report/

[43] D. L. Sipes, Jr., "Highly Efficient Nd:YAG Lasers for Free-Space Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-80, October–December 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 31–39, February 15, 1985.

[44] D. L. Sipes, "Highly Efficient Neodymium: Yttrium Aluminum Garnet Laser End-Pumped by a Semiconductor Laser Array," *Applied Physics Letters*, vol. 44, no. 2, pp. 74–76, July 15, 1985.

[45] D. L. Sipes, *Method and Apparatus for Efficient Operation of Optically Pumped Laser*, U.S. Patent 4,710,940, Washington, District of Columbia, December 1, 1987.

[46] W. K. Marshall, K. Cowles, and H. Hemmati, "Performance of Efficient Q-Switched Diode-Laser-Pumped Nd:YAG and Ho:YLF Lasers for Space Applications," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 168–173, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[47] D. L. Robinson and M. D. Rayman, "Calculations of Laser Cavity Dumping for Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 174–179, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[48] H. Hemmati and J. R. Lesh, "Environmental Testing of a Diode-Laser-Pumped Nd:YAG Laser and a Set of Diode Laser Arrays," *Proceedings of SPIE Symposium on Space Sensing, Communications, and Networking*, paper 1059-22, January 1989.

[49] C. C. Chen, "Effect of Earth Albedo Variation on the Performance of a Spatial Acquisition Subsystem Aboard a Planetary Spacecraft," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 202–211, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[50] G. G. Ortiz and S. Lee, "Acquisition, Tracking, and Pointing Using Earth Thermal Images for Deep Space Optical Communications," *2003 IEEE LEOS Annual Meeting Conference Proceedings, Tucson Arizona*, vol. 1, pp. 83–84, October 27, 2003, IEEE, Piscatawny, New Jersey, 2003.

[51] S. Lee, G. G. Ortiz, J. W. Alexander, A. Portillo, and C. Jeppesen, "Accelerometer-Assisted Tracking and Pointing for Deep Space Optical Communications," *IEEE Conference Proceedings*, Big Sky, Montana, vol. 3, pp. 31559–31564, March 10–17, 2001.

[52] B. Moision and J. Hamkins, "Deep-Space Optical Communications Downlink Budget: Modulation and Coding," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–28, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[53] S. Lee, J. W. Alexander, and G. G. Ortiz, "Sub-Microradian Pointing for Deep Space Optical Telecommunications Network," presented at *19th AIAA International Communications Satellite Systems Conference*, Toulouse, France, 2001.

[54] G. G. Ortiz, A. Portillo, S. Lee, and Juan Ceniceros, "Functional Demonstration of Accelerometer-Assisted Beacon Tracking," in *Free-Space Laser Communication Technologies XIII, Proceedings of SPIE*, vol. 4272, pp. 112–117, 2001.

[55] J. Katz, "Detectors for Optical Communications: A Review," *The Telecommunications and Data Acquisition Progress Report 42-75, July–September 1983*, Jet Propulsion Laboratory, Pasadena, California, pp. 21–38, November 15, 1983. http://ipnpr.jpl.nasa.gov/progress_report/

[56] H. H. Tan, "Avalanche Photodiode Statistics in Triggered-Avalanche Detection Mode," *The Telecommunications and Data Acquisition Progress Report 42-79, July–September 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 69–80, November 15, 1984. http://ipnpr.jpl.nasa.gov/progress_report/

[57] D. L. Robinson and D. A. Hays, "Photon Detection With Cooled Avalanche Photodiodes: Theory and Preliminary Experimental Results," *The Telecommunications and Data Acquisition Progress Report 42-81, January–March 1985*, Jet Propulsion Laboratory, Pasadena, California, pp. 9–17, May 15, 1985. http://ipnpr.jpl.nasa.gov/progress_report/

[58] D. L Robinson and B. D. Metscher, "A Cooled Avalanche Photodiode With High Photon Detection Probability," *The Telecommunications and Data Acquisition Progress Report 42-87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 41–47, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[59] D. L. Robinson and R. D. Metscher, "Photon Detection with Cooled Avalanche Photodiodes," *Applied Physics Letters*, vol. 51, no. 19, pp. 1493–1494, November 9, 1987.

[60] C.-C. Chen, "Effect of Detector Dead Time on the Performance of Optical Direct-Detection Communication Links," *The Telecommunications and Data Acquisition Progress Report 42-93, January–March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 146–154, May 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[61] A. Biswas and W. H. Farr, "Laboratory Characterization and Modeling of a Near-Infrared Enhanced Photomultiplier Tube," *The Interplanetary Network Progress Report 42-152, October–December 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, February 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[62] K. Shaik, "Spectral Filters for Laser Communications," *The Telecommunications and Data Acquisition Progress Report 42-106, April–June 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 93–101, August 15, 1991. http://ipnpr.jpl.nasa.gov/progress_report/

[63] E. L. Kerr, "Fraunhofer Filters to Reduce Solar Background for Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 48–55, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[64] B. Yin and T. M. Shay, "The Stark Anomalous Dispersion Optical Filter: The Theory," *The Telecommunications and Data Acquisition Progress Report 42-118, April–June 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 14–21, August 15, 1994. http://ipnpr.jpl.nasa.gov/progress_report/

[65] B. Yin, L. S. Alvarez, and T. M. Shay, "The Rb 780-Nanometer Faraday Anomalous Dispersion Optical Filter: Theory and Experiment," *The Telecommunications and Data Acquisition Progress Report 42-116, October–December 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 71–85, February 15, 1994. http://ipnpr.jpl.nasa.gov/progress_report/

[66] K. Kiasaleh and T.-Y. Yan, "T-PPM: A Novel Modulation Scheme for Optical Communication Systems Impaired by Pulse-Width Inaccuracies," *The Telecommunications and Mission Operations Progress Report 42-135, July–September 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–16, November 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[67] M. Srinivasan, "Receiver Structure and Performance for Trellis-Coded Pulse-Position Modulation in Optical Communication Systems," *The Telecommunications and Mission Operations Progress Report 42-135, July–September 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, November 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[68] J. Hamkins, "Performance of Binary Turbo-Coded 256-ary Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–15, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[69] J. Hamkins, J. S. Dolinar, and D. Divsalar, "Optical Channel Capacity Sensitivity," *The Telecommunications and Mission Operations Progress Report 42-143, July–September 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–16, November 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[70] B. Moision and J. Hamkins, "Constrained Coding for the Deep-Space Optical Channel," *The Interplanetary Network Progress Report 42-149, January–March 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1-29, May 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[71] S. G. Lambert, *Design and Analysis Study of a Spacecraft Optical Transceiver Package, Final Report – JPL Contract 957061*, McDonnell Douglas Corp., St. Louis, Missouri, August 19, 1985.

[72] J. Lesh, C.-C. Chen, and H. Ansari, *Lasercom System Architecture with Reduced Complexity*, U.S. Patent Number 5,517,016, Washington, District of Columbia, May 1996.

[73] M. Jeganathan, A. Portillo, C. S. Racho, S. Lee, D. M. Erickson, J. Depew, S. Monacos, and A. Biswas, "Lessons Learned from the Optical Communications Demonstrator (OCD)," *Free-Space Laser Communications Technologies XI, Proceedings of SPIE,* G. S. Mecherle, ed., vol. 3615, pp. 23–30, 1999.

[74] M. Jeganathan and S. Monacos, "Performance Analysis and Electronics Packaging of the Optical Communications Demonstrator," *Free-Space Laser Communications Technologies XI, Proceedings of SPIE,* G. S. Mecherle, ed., vol. 3226, p. 33, 1998.

[75] C. Racho and A. Portillo, "Tracking Performance Analysis and Simulation of the Digital Pointing System for the Optical Communication Demonstrator," *The Telecommunications and Mission Operations Progress Report 42-136, October–December 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, February 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[76] C. Racho and A. Portillo, "Characterization and Design of Digital Pointing Subsystem for Optical Communication Demonstrator," *Free-Space Laser Communication Technologies XI, Proceedings of the SPIE*, G. S. Mecherle, ed., vol. 3615, no. 15, pp. 250–261, January 1999.

[77] *X2000 Optical Communications Subsystem: Baseline Design Document for Europa-Orbiter Mission*, JPL D-30264 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1998.

[78] J. R. Lesh and D. L. Robinson, "A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes," *The Telecommunications and Data Acquisition Progress Report 42-87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 56–64, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[79] D. L. Robinson and J. R. Lesh, "A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes," *Proceedings of SPIE OE Lase 87* (Optical Technologies for Space Communications, Los Angeles, California, January 15–16, 1987, paper 756-20), vol. 756, pp. 130–134, January 1987.

[80] V. A. Vilnrotter, *Optical Receivers, Using Rough Reflectors*, JPL Publication 85-25, Jet Propulsion Laboratory, Pasadena, California, May 1, 1985.

[81] E. L. Kerr, "Strawman Optical Reception Development Antenna (SORDA)," *The Telecommunications and Data Acquisition Progress Report 42-93, January–March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 97–110, May 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[82] E. L. Kerr and C. W. DeVore, "Shutters and Slats for the Integral Sunshade of an Optical Reception Antenna," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 196–201, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[83] E. L. Kerr, "An Integral Sunshade for Optical Reception Antennas," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 180–195, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[84] K. Shaik "Parameter Uncertainties for a 10-Meter Ground-Based Optical Reception Station," *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp.118–127, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[85] K. Shaik, "A Two-Telescope Receiver Design for Deep Space Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-101, January–March 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 114–120, May 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[86] B. D. Clymer, "Active Optics for a Segmented Primary Mirror on a Deep-Space Optical Receiver Antenna (DSORA)," *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 128–134, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[87] B. D. Clymer, "Reflected Sunlight Reduction and Characterization for a Deep-Space Optical Receiver Antenna (DSORA)," *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 110–117, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[88] D. Wonica, "Research and Development Optical Deep Space Antenna Sizing Study," *The Telecommunications and Data Acquisition Progress Report 42-118, April–June 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 115–124, August 15, 1994. http://ipnpr.jpl.nasa.gov/progress_report/

[89] J. A. Hunter, *Orbiting Deep Space Relay Station Final Report*, JPL Publication 79-30, vols. 1, 2, 3, Jet Propulsion Laboratory, Pasadena, California, April 1, 1979.

[90] *DSRSS Concept Design and Technology Report*, JPL Contract 958734 Final Report, Stanford Telecommunications (TR93062), Reston, Virginia, July 7, 1993.

[91] *Deep Space Relay Satellite System Study Final Report*, JPL Contract No. 958733, TRW Federal Systems Division, Redondo Beach, California, September 20, 1993.

[92] K. Shaik and M. Wilhelm, "Ground Based Advanced Technology Study (GBATS): Optical Subnet Concepts for the DSN," JPL D-11000, Release 1 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 5, 1994.

[93] H. Hemmati, K. Wilson, M. K. Sue, L. J. Harcke, M. Wilhelm, C.-C. Chen, J. R. Lesh, Y. Feria, D. Rascoe, F. Lansing, and J. W. Layland, "Comparative Study of Optical and Radio-Frequency Communication Systems for a Deep-Space Mission," *The Telecommunications and Data Acquisition Progress Report 42-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–33, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[94] J. W. Layland, "Comparative Deep-Space Link Performance," *The Telecommunications and Mission Operations Progress Report 42–138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–12, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[95] K. E. Wilson, M. Wright, R. Cesarone, J. Ceniceros, and K. Shea, "Cost and Performance Comparison of an Earth-Orbiting Optical Communication Relay Transceiver and a Ground-Based Optical Receiver Subnet," *The Interplanetary Network Progress Report 42-153, January–March 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–12, May 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[96] B. Levitt, K. Wilson, T. Roberts, R. Cesarone, and R. Hastrup, *Hybrid Optical DSN Architecture: Interleaved PPM Concept*, JPL D-30264 (internal document), Pasadena, California, November 1, 2004.

[97] J. V. Sandusky, D. J. Hoppe, and M. J. Britcliffe, "Deep-Space Optical Reception Antenna (DSORA): Aperture Versus Quality," *The Telecommunications and Mission Operations Progress Report 42-143, July–September 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, November 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[98] M. J. Britcliffe and D. J. Hoppe, "Main-Reflector Manufacturing Technology for the Deep Space Optical Communications Ground Station," *The Telecommunications and Mission Operations Progress Report 42-145, January–March 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–10, May 15, 2001. http://ipnpr.jpl.nasa.gov/progress_report/

[99] M. Britcliffe, D. Hoppe, W. Roberts, and N. Page, "A Ten-Meter Ground-Station Telescope for Deep-Space Optical Communications: A Preliminary Design," *The Interplanetary Network Progress Report 42-147, July–September 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, November 15, 2001. http://ipnpr.jpl.nasa.gov/progress_report/

[100] K. S. Shaik, "Atmospheric Propagation Effects Relevant to Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-94, April–June 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 180–200, August 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[101] K. S. Shaik, "A Preliminary Weather Model for Optical Communications Through the Atmosphere," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September*, Jet Propulsion Laboratory, Pasadena, California, pp. 212–218, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[102] D. P. Wyle, "Cloud Cover Statistics from GOES/VAS Satellite," *Proceedings of SPIE OE/Lase 88*, paper 874-34, Los Angeles California, January 1988.

[103] K. Cowles, "Site Selection Criteria for the Optical Atmospheric Visibility Monitoring Telescopes," *The Telecommunications and Data Acquisition Progress Report 42-97, January–March 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 235–239, May 15, 1989. http://ipnpr.jpl.nasa.gov/progress_report/

[104] K. Cowles, "A Visibility Characterization Program for Optical Communications Through The Atmosphere," *The Telecommunications and Data Acquisition Progress Report 42-97, January–March 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 221–225, May 15, 1989. http://ipnpr.jpl.nasa.gov/progress_report/

[105] D. Erickson and K. Cowles, "Options for Daytime Monitoring of Atmospheric Visibility in Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-97, January–March 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 226–234, May 15, 1989. http://ipnpr.jpl.nasa.gov/progress_report/

[106] K. A. Cowles, "Site Comparison for Optical Visibility Statistics in Southern Arizona," *The Telecommunications and Data Acquisition Progress Report 42-102, April–June 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 57–61, August 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[107] K. Cowles, "Site Comparison for Optical Visibility Statistics in Southern California," *The Telecommunications and Data Acquisition Progress Report 42-105, January–March 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 31–40, May 15, 1991. http://ipnpr.jpl.nasa.gov/progress_report/

[108] K. Cowles and D. Erickson, "Alignment of the Atmospheric Visibility Monitoring Telescope," *The Telecommunications and Data Acquisition Progress Report 42-108, October–December 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 79–83, February 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[109] K. Cowles, "Hardware Design for the Autonomous Visibility Monitoring (AVM) Observatory," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 295–301, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[110] K. Cowles and B. M. Levine, "A Preliminary Optical Visibility Model," *The Telecommunications and Data Acquisition Progress Report 42-119, July–September 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 201–209, November 15, 1994. http://ipnpr.jpl.nasa.gov/progress_report/

[111] B. Sanii, A. Datta, D. Tsiang, J. Wu, and A. Biswas, "Preliminary Results of an Upgraded Atmospheric Visibility Monitoring Station," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–12, August 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[112] *NASA JPL CFLOS Study Final Review, JPL Contract 1248595 Final Briefing*, Northrop Grumman (TASC), Chantilly Virginia, September 26, 2003.

[113] J. Katz, "The Deep-Space Optical Channel: I. Noise Mechanisms," *The Telecommunications and Data Acquisition Progress Report 42-64, May and June 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 180–186, August 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[114] J. Katz, "The Deep Space Optical Channel: II. Wave Propagation Effects," *The Telecommunications and Data Acquisition Progress Report 42-65, July and August 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 53–58, October 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[115] J. Katz, "Planets as Background Noise Sources in Free Space Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-85, January–March 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 13–24, May 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[116] W. L. Wolfe and G. J. Zissis, eds., *The Infrared Handbook*, Chapter 3, Environmental Research Institute of Michigan, Office of Naval Research, U.S. Department of the Navy, Washington, District of Columbia, 1978.

[117] E. E. Bell, L. Eisner, J. Young, and R. A. Oetjen, "Spectral Radiance of Sky and Terrain at Wavelengths between 1 and 20 microns: II—Sky Measurements," *Journal of the Optical Society of America*, vol. 50, no. 12, pp. 1313–1320, December 1960.

[118] L. Eisner, E. E. Bell, J. Young, and R. A. Oetjen, "Spectral Radiance of Sky and Terrain at Wavelengths between 1 and 20 microns: III—Terrain Measurements," *Journal of the Optical Society of America*, vol. 52, no. 2, pp. 201-209, February 1962.

[119] V. V. Vilnrotter, *Background Sources in Optical Communications*, JPL Publication 83-72, Jet Propulsion Laboratory, Pasadena, California, November 15, 1983.

[120] W. K. Marshall and B. D. Burk, "Received Optical Power Calculations for Optical Communications Link Performance Analysis," *The Telecommunications and Data Acquisition Progress Report 42-87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 32–40, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[121] J. R. Lesh, W. K. Marshall, and J. Katz, "A Simple Method for Designing or Analyzing an Optical Communication Link," *The Telecommunications and Data Acquisition Progress Report 42-85, January–March 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 25–31, May 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[122] W. K. Marshall and D. D. Burk, "Received Optical Power Calculations for Optical Communications Link Performance Analysis," *The Telecommunications and Data Acquisition Progress Report 42–87, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 32–40, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[123] *Optical Communications Link Analysis Computer Program, NASA New Technology Report*, NPO-17107, December 16, 1986.

[124] C.-C. Chen, "Design of an Optical PPM Communication Link in the Presence of Component Tolerances," *The Telecommunications and Data Acquisition Progress Report 42-94, April–June 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 170–179, August 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[125] M. Jeganathan, G. S. Mecherle, and J. R. Lesh, "Development of the Free-space Optical Communications Analysis Software (FOCAS)" *Free-space Laser Communications Technologies X, Proceedings of SPIE*, G. S. Mecherle, ed., vol. 3266, pp. 90–98, 1998.

[126] J. R. Lesh, *Optical Communications Air-to-Ground Demonstration: Proposal Study*, JPL D-30265, (internal document), Jet Propulsion Laboratory, Pasadena, California, October 10, 1994.

[127] B. D. Metscher, "An Evaluation of the Communications System for the TAU Mission Study," *The Telecommunications and Data Acquisition Progress Report 42-89, January–March 1987*, Jet Propulsion Laboratory, Pasadena, California, pp. 108–117, April 1987. http://ipnpr.jpl.nasa.gov/progress_report/

[128] K. T. Nock, "TAU—A Mission to a Thousand Astronomical Units," *Proceedings of the 19th AIAA/DGLR/JSASS International Electric Propulsion Conference*, Colorado Springs, Colorado, Paper AIAA-87-1049, May 11–13, 1987.

[129] J. R. Lesh, C. J. Ruggier, and R. J. Cesarone, "Space Communications Technologies for Interstellar Missions" *Journal of the British Interplanetary Society*, vol. 49, pp. 7–14, 1996.

[130] K. E. Wilson, J. R. Lesh, T.-Y. Yan, J. Schwartz, M. D. Rayman, and S. Wee, "GOPEX: A Deep-Space Optical Communications Demonstration With the Galileo Spacecraft," *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 262–277, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[131] M. D. Rayman, "Selection and Observability Tests of GOPEX Reference Stars," *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 278–283, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[132] K. Kiesaleh and T.-Y. Yan, "A Statistical Model for Evaluating GOPEX Uplink Performance," *The Telecommunications and Data Acquisition Progress Report 42-111, July–September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 325–332, November 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[133] R. Q. Fugate, "GOPEX at the Starfire Optical Range," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 255–279, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[134] K. Shaik, D. Wonica, and M. Wilhelm, "Optical Subnet Concepts for the Deep Space Network," *The Telecommunications and Data Acquisition Progress Report 42-115, July–September 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 153–181, November 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[135] K. E. Wilson and J. R. Lesh, "An Overview of the Galileo Optical Experiment (GOPEX)," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 192–204, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[136] J. Yu and M. Shao, "Galileo Optical Experiment (GOPEX) Optical Train: Design and Validation at the Table Mountain Facility," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 236–247, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[137] G. Okamoto and K. Masters, "GOPEX Laser Transmission and Monitoring Systems," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 248–254, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[138] A. Biswas, "Calibration of the Receiver Channel for the GOPEX Precursor Experiments," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 205–212, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[139] K. E. Wilson, P. R. Leatherman, R. Cleis, J. Spinhirne, and R. Q. Fugate, "Results of the Compensated Earth–Moon–Earth Retroreflector Laser Link (CEMERLL) Experiment," *The Telecommunications and Data Acquisition Progress Report 42-131, July–September 1997*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, November 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[140] S. D. Gillam, J. W. Young, and D. R. Sidwell, "JPL Table Mountain Facility Support of the Ground/Orbiter Lasercomm Demonstration," *The Telecommunications and Data Acquisition Progress Report 42-125, January–March 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, May 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[141] M. Jeganathan, K. E. Wilson, and J. R. Lesh, "Preliminary Analysis of Fluctuations in the Received Uplink-Beacon-Power Data Obtained From the GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report 42-124, October–December 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 20–32, February 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[142] K. E. Wilson, "An Overview of the GOLD Experiment Between the ETS-VI Satellite and the Table Mountain Facility," *The Telecommunications and Data Acquisition Progress Report 42-124, October–December 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 8–19, February 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[143] K. Wilson, M. Jeganathan, J. R. Lesh, J. James, and G. Xu, "Results From Phase-1 and Phase-2 GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report 42-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[144] C. Pasqualino, "Characterization of the Avalanche Photo Diode Detector Used for Optical Communications Experiments With the Japanese ETS-VI," *The Telecommunications and Data Acquisition Progress Report 42-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–8, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[145] M. Toyoshima, K. Araki, Y. Arimoto, M. Toyoda, M. Jeganathan, K. Wilson, and J. R. Lesh, "Reduction of ETS-VI Laser Communication Equipment Optical-Downlink Telemetry Collected During GOLD," *The Telecommunications and Data Acquisition Progress Report 42-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–9, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[146] A. Biswas and S. Lee, "Ground-to-Ground Optical Communications Demonstration," *The Telecommunications and Mission Operations Progress Report 42-141, January–March 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–31, May 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[147] A. Biswas and M. W. Wright, "Mountain-Top-to-Mountain-Top Optical Link Demonstration: Part I," *The Interplanetary Network Progress Report 42-149, January–March 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–27, May 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[148] A. Biswas, M. W. Wright, B. Samii, and W. A Page, "45-km Horizontal Path Optical Link Demonstration," *Proceedings of SPIE, Free-Space Laser Communications Technologies XIII*, G. S. Mecherle, ed., vol. 4272, pp. 60–71, 2001.

[149] A. Biswas and M. W. Wright, "Mountain-Top-to-Mountain-Top Optical Link Demonstration: Part II," *The Interplanetary Network Progress Report 42-151, July–September 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–16, November 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[150] B. L. Schumaker, "Apparent Brightness of Stars and Lasers," *The Telecommunications and Data Acquisition Progress Report 42-93, January–March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 111–130, May 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[151] W. M. Folkner and M. H. Finger, "Photon Statistical Limitations for Daytime Optical Tracking," *The Telecommunications and Data Acquisition Progress Report 42-99, July-September 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 90–97, November 15, 1989. http://ipnpr.jpl.nasa.gov/progress_report/

[152] G. W. Null and P. W. Chodas, "Optical Self-Crossing Measurements for Ida Ephemeris Improvement," *The Telecommunications and Data Acquisition Progress Report 42-103, July-September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 87–96, November 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[153] W. M. Folkner and M. H. Finger, "Preliminary Error Budget for an Optical Ranging System: Range, Range Rate, and Differenced Range Observables," *The Telecommunications and Data Acquisition Progress Report 42-101, January–March 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 121–135, May 15, 1990. http://ipnpr.jpl.nasa.gov/progress_report/

[154] G. W. Null, W. M. Owen, Jr., and S. P. Synnott, "Deep-Space Navigation Applications of Improved Ground-Based Optical Astrometry," *The Telecommunications and Data Acquisition Progress Report 42-110, April–June 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 118–127, August 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[155] G. W. Null, W. M. Owen, Jr., and S. P. Synnott, "Systems Analysis for Ground-Based Optical Navigation," *The Telecommunications and Data Acquisition Progress Report 42-111, July–September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 23–40, November 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[156] G. D. Gatewood, "The Multichannel Astrometric Photometer and Atmospheric Limitations in the Measurement of Relative Precisions," *Astronomical Journal*, vol. 94, no. 1, pp. 213–224, 1987.

[157] K. E. Wilson, M. Britcliffe, and N. Golshan, "Progress in Design and Construction of the Optical Communications Laser Laboratory," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–6, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[158] K. E. Wilson, W. T. Roberts, V. Garkanian, F. Battle, R. Leblanc, H. Hemmati, and P. Robles, "Plan for Safe Laser Beam Propagation from the Optical Communications Telescope Laboratory," *The Interplanetary Network Progress Report 42-152, October–December 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, February 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[159] K. E. Wilson, N. Page, J. Wu, and M. Srinivasan, "The JPL Optical Communications Telescope Laboratory Test Bed for the Future Optical Deep Space Network," *The Interplanetary Network Progress Report 42-153, January–March 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–8, May 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[160] G. G. Ortiz, S. Lee, S. P. Monacos, M. W. Wright, and A. Biswas, "Design and development of a robust ATP subsystem for the Altair UAV-to-Ground Lasercom 2.5 Gbps Demonstration," *Free-Space Laser Communications Technologies XV, Proceedings of SPIE*, G. S. Mecherle, ed., vol. 4975, pp. 103–114, January 28, 2003.

[161] M. Troy, R. Dekany, G. Brack, B. Oppenheimer, E. Bloemhof, T. Trinh, F. Dekens, F. Shi, T. Hayward, and B. Brandl, "Palomar Adaptive Optics Project: Status and Performance," *Proceedings of SPIE*, vol. 4007, pp. 31–40, 2000.

[162] T. L. Hayward, B. Brandl, B. Pirger, C. Blackenm, G. E. Gull, J. Schoenwald, and J. R. Houck, "PHARO-A Near-Infrared Camera for the Palomar Adaptive Optics System," *Publications of the Astronomical Society of the Pacific (PASP)*, vol. 113, no. 779, pp. 105–118, January 2001.

[163] T. Truong, G. Brack, M. Troy, T. Trinh, F. Shi, and R. Dekany, "Real-Time Wavefront Processors for the Next Generation of Adaptive Optics Systems: A Design and Analysis," *Proceedings of SPIE*, vol. 4839, pp. 911–922, 2003.

[164] M. A. van Dam and B. A. Macintosh, "Characterization of adaptive optics at Keck Observatory," *Astronomical Adaptive Optics Systems and Applications*, R. K. Tyson and M. Lloyd-Hart, eds., *Proceedings of SPIE*, vol. 5169, pp. 1–10, 2003.

[165] R. Dekany, B. Bauman, D. Gavel, M. Troy, B. Macintosh, and M. Britton, "Initial Concepts for CELT Adaptive Optics," *Proceedings of SPIE*, vol. 4839, pp. 1165–1174, 2003.

[166] K. Wilson, M. Troy, M. Srinivasan, B. Platt, V. Vilnrotter, M. Wright, V. Garkanian, and H. Hemmati, "Daytime Adaptive Optics For Deep Space Optical Communications," *Space Activities and Cooperation Contributing to All Pacific Basin Countries: Tenth International Space Conference of Pacific Basin Countries, ISCOPS*, Tokyo, Japan, December 10–12, P. M. Bainum, L. Furong, and T. Nakajima, eds., American Astronomical Society, pp. 481–492, 2003.

[167] W. Happer, G. J. MacDonald, C. E. Max, and F. J. Dyson, "Atmospheric Turbulence Compensation by Resonant Optical Backscattering from the Sodium Layer in the Upper Atmosphere," *Journal of the Optical Society of America*, vol. A 11, p. 263, 1994.

[168] R. Q. Fugate, B. L. Ellerbroek, C. H. Higgings, M. P. Jelonek, W. J. Lange, A. C. Slavin, W. J. Wild, D. M. Winker, J. M. Wynia, J. M. Spinhirne, B. R. Boeke, R. E. Ruane, J. F. Moroney, M. D. Oliker, D. W. Swindle, and R. A. Cleis, "Two Generations of Laser-Guide-Star Adaptive-Optics Experiments at the Starfire Optical Range," *JOSA A*, vol. 11, no. 1, pp. 310–324, 1994.

[169] V. Vilnrotter and M. Srinivasan, "Adaptive Detector Arrays for Optical Communications Receivers," *IEEE Transactions on Communications*, vol. 50, no. 7, pp. 1091–1097, July 2002.

[170] V. Vilnrotter and M. Srinivasan, "Optical Communications through Atmospheric Turbulence Using Photodetector Arrays," *Free-Space Laser Communication Technologies XIII, Proceedings of SPIE*, vol. 4272, pp. 282– 292, San Jose, January 2001.

[171] V. Vilnrotter and M. Srinivasan, "Adaptive Detector Arrays for Optical Communications Receivers," *The Telecommunications and Mission Operations Progress Report 42–141, January–March 2000*, Jet Propulsion Laboratory, Pasadena, California, May 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[172] M. Srinivasan and V. Vilnrotter, "Avalanche Photodiode Arrays for Optical Communications Receivers," the *Telecommunications and Mission Operations Progress Report 42–144, October–December 2000*, Jet Propulsion Laboratory, Pasadena, California, February 15, 2001. http://ipnpr.jpl.nasa.gov/progress_report/

[173] V. Vilnrotter and M. Srinivasan, "Adaptive Optical Array Receiver: Communicating through Atmospheric Turbulence," *IND Science and Technology News*, JPL 410-62, Jet Propulsion Laboratory, Pasadena, California, issue 14, pp. 14–18, September 2001. http://tmot.jpl.nasa.gov/Program_Overview_Information/IND_Program_News/

[174] M. Munoz and V. Vilnrotter, "Coherent Optical Receiver for PPM Signals Received through Atmospheric Turbulence," *Free-Space Laser Communications Technologies XVI, Proceedings of SPIE*, vol. 5338, pp. 151–163, San Jose, January 2004.

[175] V. Vilnrotter, C.-W. Lau, M. Srinivasan, R. Mukai, and K. Andrews, "An Optical Array Receiver for Deep-Space Communication through Atmospheric Turbulence," *The Interplanetary Network Progress Report 42-154, April–June, 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–21, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[176] V. Vilnrotter, C.-W. Lau, M. Srinivasan, R. Mukai, and K. Andrews, "Optical Array Receivers for Deep Space Communication," *Space Activities and Cooperation Contributing to All Pacific Basin Countries: Tenth International Space Conference of Pacific-Basin Societies;* P. M. Bainum, L. Furong, and Takashi Nakajima, eds., ISCOPS Conference, Tokyo, Japan, December 10–12, 2003, vol. 117, pp. 493–510, 2004.

[177] V. Vilnrotter, C.-W. Lau, and M. Srinivasan, "Optical Array Receiver for Deep-Space Communications," *Free Space Laser Communication Technologies XVI, Proceedings of SPIE*, G. S. Mecherle, C. Y. Young, and J. S. Stryjewski, eds., pp. 163–174, June 2004.

[178] V. Vilnrotter, K. Andrews, and C.-W. Lau, "Conceptual Design of an Optical Array Receiver with Preliminary Experimental Results," *The Interplanetary Network Progress Report 42–156*, Jet Propulsion Laboratory, Pasadena, California, February 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[179] T. Tolker-Nielsen, B. Demelenne, and E. Desplats, "In-orbit Test Results of the First SILEX Terminal," *Free-Space Laser Communication Technologies XI, Proceedings of SPIE*, vol. 3615, pp. 31–42, , San Jose, California, January 1999.

[180] T. Tolker-Nielsen and G. Oppenhauser, "In-Orbit Test Result of an Operational Intersatellite Link between ARTEMIS and SPOT 4," *Free-Space Laser Communication Technologies XIV, Proceedings of SPIE*, vol. 4639, pp. 1–15, San Jose, California, January 2002.

[181] M. Toyoshima, S. Yamakawa, T. Yamawaki, and K. Arai, "Ground-to-Satellite Optical Link Tests Between Japanese Laser Communication Terminal and European Geostationary Satellite ARTEMIS," *Free-Space Laser Communication Technologies XVI*, vol. 4635, pp. 1–15, San Jose, California, January 2004.

[182] A. Biswas and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–38, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

Chapter 2
Link and System Design

Chien-Chung Chen

Laser communications (lasercom) technology offers the potential for significantly increasing in data return capability from deep space to Earth. Compared to the current state of the art radio frequency (RF) communications links, lasercom links operate at much higher carrier frequencies (approximately 200–300 terahertz [THz]) compared to 32 gigahertz (GHz) for state of the art Ka-band deep-space RF links). The use of higher carrier frequencies implies a much smaller diffraction loss (e.g., much narrower beamwidth), which in turn, results in a much higher efficiency in delivering the signal energy. This improved energy delivery efficiency allows an optical link to operate at a lower transmit power and aperture size while still achieving a higher link data rate. Furthermore, unlike RF links where the spectral allocation and available channel bandwidth are tightly regulated due to interference concerns, the optical link is highly directional and virtually free of spectral constraints.

Although the lasercom system offers the potential for a small aperture high-data-rate transmission system, implementation of the lasercom system demands design considerations not commonly required for RF communications systems. This is principally because of the narrow transmission beamwidth of the optical signal. In order to efficiently deliver the signal and to reduce to probability of pointing-induced signal fades, the transmitter pointing error typically needs to be maintained within a small fraction of the transmit beamwidth. For a typical size aperture being considered for near-Earth and deep-space lasercom missions, the transmit beamwidth is typically on the order of a few microradians, and the required pointing accuracy is a small fraction of a microradian. The flight lasercom terminal must achieve this pointing accuracy

in the presence of spacecraft platform jitter and attitude control deadband, both of which can be several orders of magnitude larger than the required pointing accuracy.

Over the last two decades, a number of lasercom flight demonstrations have been flown to demonstrate the technical feasibility of using modulated laser signals for high-rate data transport over free space. These flight experiments, mostly conducted with aircraft and spacecraft in the Earth vicinity, have demonstrated the technical feasibility of establishing and maintaining two-way precision beam pointing between transmit and receive terminals, and the capability of maintaining high-rate data links through the free-space optical channel. These flight experiments also led to the development of high-power space-qualified laser transmitters, optics, and precision beam-pointing hardware, as well as the resulting increase of NASA interest for further exploring the feasibility of using laser communications for deep-space missions.[1]

Even though these previous flight experiments established the feasibility of lasercom systems for near-Earth applications, deep-space missions can impose significant challenges such that a straightforward scaling of the near-Earth lasercom system architecture to deep-space distances would lead to unacceptable link performance. These differences come primarily from the longer link distance involved. The distance covered by the Mars mission ranges from two thousand times (Mars at closest approach) to ten thousand times (Mars at solar conjunction) the distance from Earth to geosynchronous Earth orbit (GEO). The longer link distance translates into larger aperture, higher power, and greater receiver sensitivity requirements for the deep-space link. Pointing and tracking a narrow signal from deep-space distances are also significantly more difficult due to the large link distance and long round-trip light time (RTLT). Additionally, deep-space missions need to handle a wide range of operating conditions and trajectory constraints. For example, solar conjunction outages for GEO satellites typically last for tens of minutes, whereas for planetary missions the solar conjunction outage can last from several days to several weeks, depending on how closely the optical system can operate to the Sun near its optical boresight. Because of the higher launch costs

[1] The need for deep-space optical communications has been articulated in the NASA 2003 Strategic Plan [1] as a "New Effort Building Block" under the "Communications Technological Barrier" for "providing efficient data transfer across the solar system." The Strategic Plan identifies optical communications as necessary to "vastly improve communication to transform science capability, with a first demonstration from Mars." NASA's Science Mission Directorate expressed the need for optical communications as "the Optical Communications Initiative will demonstrate critical space and ground technologies in this decade and perform a flight demonstration of high-data-rate communication from Mars in the 2010 timeframe."

and longer mission lifetimes, deep-space missions generally place a premium on mass and power of the flight terminal, and have a more stringent mission reliability requirement. Finally, unlike RF system designs, where a well-defined ground network can be used to help define the flight terminal requirement, no such infrastructure exists for the deep-space optical network. As a result, system designers will need to evaluate design drivers for both the deep-space equipment and the Earth terminals in order to arrive at the proper design.

Given the relative complex set of trades required to define the deep-space lasercom system, the purpose of this chapter is to provide an overview of the major design drivers for a deep-space lasercom system and their implications for flight terminal and ground network design and implementations, and to provide a context for more in-depth discussion in subsequent chapters. These drivers include:

1) Communications link performance,
2) Beam Pointing and Spatial Acquisition,
3) Laser safety,
4) Other considerations such as mass, power, and impact on spacecraft.

2.1 Overview of Deep-Space Lasercom Link

An overview of a generic deep-space lasercom link is shown in Fig. 2-1. The link consists of a lasercom flight terminal aboard the deep-space spacecraft, an optical channel, and one or more Earth terminals. The flight lasercom terminal interfaces with the host spacecraft, which provides power, control, ephemeris and pointing information, and coarse attitude control. The flight lasercom terminal also receives the downlink data stream from the spacecraft and delivers the uplink data to the spacecraft. The functions of the flight lasercom terminal are to:

1) Encode and modulate the downlink information onto an optical carrier,

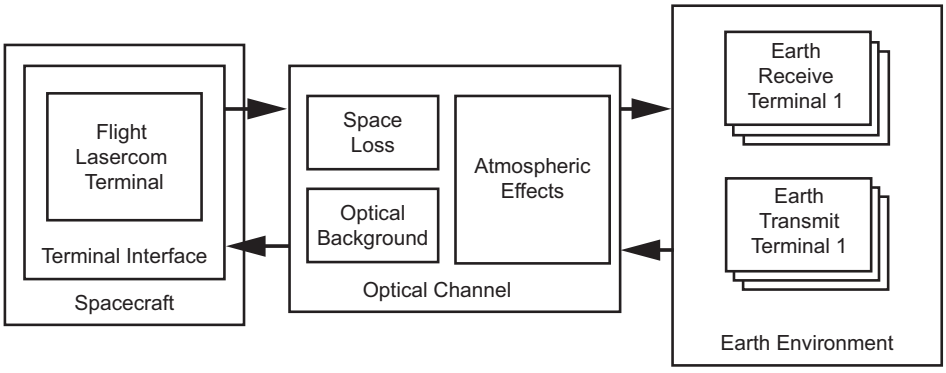


Fig. 2-1. Overview of a deep space lasercom link.

2) Provide an appropriate optical power and transmit antenna gain in order to close the communications link,
3) Acquire the appropriate pointing reference and point the downlink signal at the Earth terminal,
4) Provide suitable pointing stabilization functions against the platform jitter and spacecraft attitude control deadband, and
5) Provide appropriate receiving antenna gain and detection sensitivity to receive uplink data from the Earth terminal.

The signal passes through an optical channel, which adds space loss ($1/Z^2$ loss) to the signal. The optical channel also introduces background noise at the receiving terminal. The major sources of the background noise are the Sun, the Moon, the planets, and bright stars. If the Earth terminal is ground based, the signal also passes through the atmosphere, which introduces additional background (sky irradiance), attenuation, and signal scintillation. In addition to clear weather attenuation, an optical signal passing through the atmosphere can also be severely attenuated by clouds. Effective communications through clouds is not a feasible solution as cloud attenuation can be upwards of tens of dB in some cases (e.g., cumulus nimbus), and appropriate operational workaround needs to be considered as part of the optical link design. Atmospheric scintillation is also an important effect because it breaks up the spatial coherence of the optical signal. As we shall see, this effectively prevents the use of coherent optical reception technique for a ground-based receiver. For an optical uplink, atmospheric scintillation can lead to beam wander and fades at the receiving end, which must be considered when designing an optical uplink.

The optical downlink from the flight lasercom terminal is received by one or more Earth receive terminals. The functions of the Earth receiving terminals are to provide

1) Appropriate receiving antenna gain and sensitivity to receive, demodulate, and decode the optical downlink.
2) Suitable pointing accuracy of the receiving antenna in order to direct the downlink onto the receiving detector while limiting the amount of background signal admitted by the receiving optics.
3) Sufficient spatial diversity to support the mission/link availability requirements.

In addition to the receiving terminals, one or more Earth transmit terminals may also be deployed if either optical uplink communications is required, or the flight terminal pointing acquisition and tracking scheme requires the use of an Earth-based reference beacon to direct the downlink signal. The functions of the Earth transmit terminals are to provide

1) Sufficient optical power, pointing accuracy, and directivity in order to deliver the required uplink signal flux at the flight lasercom terminal for uplink communications or for beacon pointing.
2) Sufficient spatial diversity to support the mission/link availability requirements.

The Earth terminal(s) can be either ground based or balloon/aircraft/spacecraft based. The latter can communicate above much or all of the Earth atmosphere, thus having significant operational advantages. However, because of the large aperture required to support the deep-space link, the lifecycle costs for a balloonborne, airborne, or spaceborne terminal are much higher, and the logistics of supporting a flight terminal are significantly more difficult than those of a ground-based terminal. Consequently, most of the studies performed to date have assumed a ground-based Earth terminal. However, as technologies for lightweight optics continue to develop, such terminals may eventually present feasible options. For the remainder of this Chapter, we shall assume that the Earth terminal is ground based. In order to provide a suitable amount of link availability, it is envisioned that a network of ground stations will be required.

2.2 Communications Link Design

The capability to support (and achieve) a very high downlink data rate is the principal benefit for the deep-space lasercom technology. Given the existing capability of the Deep Space Network (DSN) and the relative maturity of RF communications technology at X-band (8 GHz) and Ka (32 GHz) band, the deep-space lasercom technology needs to achieve a significant data-rate advantage over the existing RF implementation before it can be seriously considered for future missions.

A useful metric for comparing the end-to-end communications link performance is the data rate-distance square product. The current state-of-the-art near-Earth lasercom system supports upwards of a 10 gigabits per second (Gbps) link from GEO distance. Using the data rate-distance square product metric, such a system will scale to approximately 100 bits per second (bps) at Mars distance and 0.25 bps at Pluto, as shown in Fig. 2-2; which is grossly inadequate for the deep-space mission requirements.

In contrast, the performance of several currently on-going or near-term deep-space RF communication systems is shown in Table 2-1 and plotted against the state-of-the art optical link performance in Fig. 2-3. It is seen that both the Cassini and the 2005 Mars Reconnaissance Orbiter (MRO) achieved major link performance advantages over the current state-of-the art optical link (i.e., Geolite). In order to be competitive against the RF system performance, significant improvements (>50 dB) in optical link performance are required.

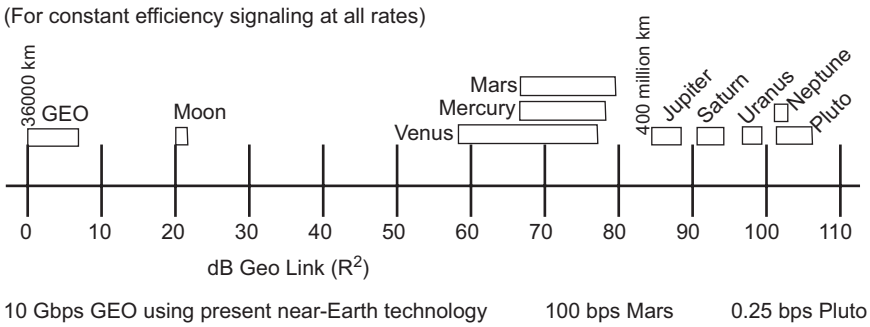


Fig. 2-2. Scaling of lasercom link performance over distance.

Table 2-1. Current RF link performance.

Mission	Communications System	Performance
Cassini	20-W X-band TWT, 4-m HGA	14 kbps at 10 AU
Mars Odyssey	15-W X-band SSPA, 1.3-m HGA	4–110 kbps at 2.6 AU
Mars Reconnaissance Orbiter	100-W X-band TWT, 3-m HGA	500 kbps at 2.6 AU
	35-W Ka-band TWT	300 kbps at 2.6 AU

(AU = astronomical unit [1.496 x 10^{11} m], HGA = high-gain antenna, SSPA = solid state power amplifier, TWT = traveling wave tube)

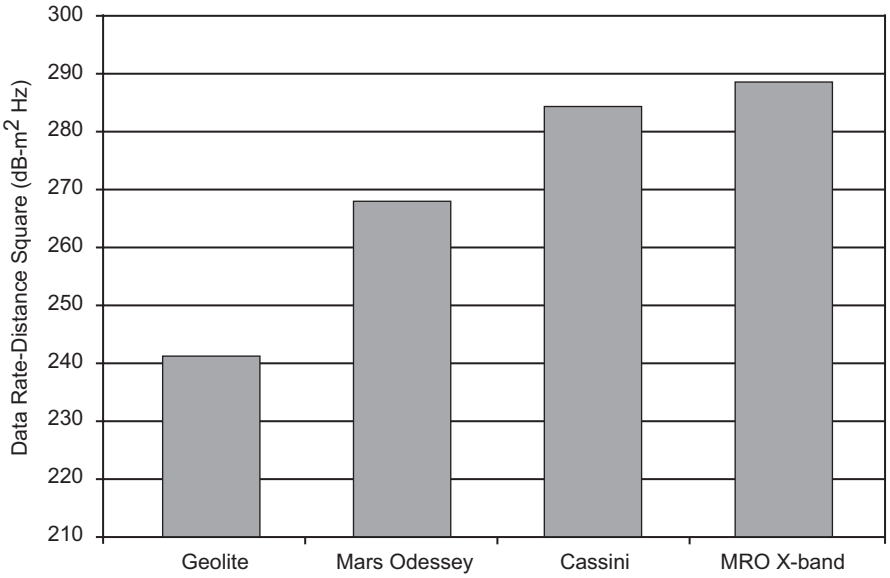


Fig. 2-3. Figures of merit comparison between a current lasercom system and selected RF systems.

Achieving the large performance improvement over current state of the art will require attentions in the following areas:

1) Improving the amount of signal power delivered to the receiver. This will include increasing the amount of transmit power and antenna gains, as well as the efficiency of the optics and pointing performance.
2) Improving the receiver sensitivity, measured in terms of effective delivered bits per received signal photon.

2.2.1 Link Equation and Receive Signal Power

The ability for an optical link to deliver the signal power to the receiver is governed by the link equation, which can generally be written as

$$P_S = P_T\left(\eta_T \eta_A \frac{4\pi A_T}{\lambda_T^2}\right) L_{TP} L_{atm} L_{pol} L_{RP}\left(\frac{A_R}{4\pi z^2}\right)\eta_R \qquad (2.2\ 1)$$

where

P_S is the total signal power at the input to the receiver. For the uplink, this is defined at the input to the optical detector. For the downlink, the receive signal power is defined at the input to the receive optical detector.

P_T is the transmit optical power at the transmit antenna interface.

η_T is the transmit optics efficiency.

η_A is the aperture illumination efficiency of the transmit antennas.

λ_T is the transmit wavelength.

A_T is the aperture areas, respectively.

L_{TP} is the transmitter pointing loss, defined as the ratio of power radiated in the direction of receiver to the peak radiated power. If the transmitter is directly pointed at the receiver, the pointing loss is 0 dB.

L_{atm} is the fractional loss due to absorption of the transmitting medium (e.g., Earth atmosphere and any occluded planet atmospheres)

L_{pol} is the fractional signal loss due to mismatch of the transmit and receive antenna polarization patterns.

A_R is the receive aperture area.

z is the link distance, and the term $\left(A_R / 4\pi z^2\right)$ is the fraction of power that is collected by the receiving aperture if the transmitter is an isotropic radiator.

L_{RP} is the receiver pointing loss, defined as the ratio of receive antenna gain in the direction of the transmitter to the peak receive antenna gain.

η_R is the receiving optics collecting efficiency, defined as the fraction of optical power at the receiving aperture that is collected within the field of view of the receive detector.

Improving the receive signal power, therefore, can be accomplished by the following means:

1) Increasing the transmit power. The most straightforward method of improving the receive signal power is to increase the power at the transmitter since the receive power scales linearly with the transmit power. However, increasing the transmit power also increases the overall system power consumption which, for a deep-space mission, is typically at a premium. Furthermore, the increased power consumption can lead to thermal management issues (increased radiator size and hence mass) for the host spacecraft, as well as reliability concerns.

2) Increasing the transmit aperture. This effectively reduces the transmit beamwidth and hence improves the power delivery efficiency. However, the pointing and tracking of the narrow downlink becomes increasingly more difficult with a narrower downlink. Furthermore, the aperture size is highly correlated with the mass of the transmit terminal and hence cannot be increased indefinitely.

3) Reducing the operating wavelength. Reducing the operating wavelength reduces the diffraction loss of the signal (i.e., reduces the transmit beamwidth). However, the wavelength selection is strongly constrained by the available laser technology, as well as considerations on the receiver sensitivity and detector technology. Furthermore, the transmittance of the atmosphere also depends on the wavelength, as well as the amount of sky background irradiance.

4) Increasing the receiver aperture area. Since the receive signal power scales linearly with the receive aperture area, increasing the receiver aperture area is a relatively simple way to improve the system performance. However, for daytime operations of a receiver inside the Earth's atmosphere, the amount of background noise collected also increases with increasing receiver aperture, and the effective performance improvement does not always scale linearly with increasing aperture area.

5) Reduced pointing loss. Reducing the pointing loss improves the overall signal energy and also reduces the point-induced signal power fluctuation.

6) Improving the overall efficiency, including transmit and receive optical loss, and polarization mismatch losses. This generally requires attention to the optical design. Of particular attention is the transmit optics design. The

transmit aperture illumination efficiency, η_A, depends on the phase and intensity distribution over the aperture. For the general case of a transmit aperture being illuminated by a Gaussian beam, the aperture illumination efficiency can be written as:

$$\eta_A = \frac{2S}{\alpha^2}\left[\exp\left(-\alpha^2\gamma^2\right) - \exp\left(-\alpha^2\right)\right]^2 \tag{2.2-2}$$

where α is the ratio between the aperture diameter and the Gaussian beam $\left(1/e^2\right)$ diameter of the transmit signal, and γ is the obscuration ratio. The term S in Eq. (2.2-3) is known as the Strehl ratio, which is defined as the intensity at the center of the aberrated system to that of an ideal optical system. The Strehl loss is given by

$$S = \exp\left(-(2\pi\sigma / \lambda)^2\right) \tag{2.2-3}$$

where σ is the root mean square (rms) optical path difference, which for smooth optics is approximately 28 percent of the peak-to-valley differences. For a $\lambda/16$ optical system, for example, the Strehl ratio is approximately 86 percent, or approximately a 0.65-dB loss.

2.2.2 Optical-Receiver Sensitivity

In addition to the effective delivery of the signal to the detector, the performance of the optical link also depends on the receiver sensitivity (measured in terms of received photons per bit). Because of the high cost associated with increasing the transmit power and system aperture, improving the receiver sensitivity is an important factor in the deep-space lasercom system design.

Either a coherent receiver or a direct-detection receiver can be used to detect the optical signal. In a coherent optical receiver, the incoming signal is mixed with the output of a strong local oscillator (LO) beam, and the interference between the signal and LO in the combined field is detected using a pair of photodetectors. Figure 2-4 shows a conceptual block diagram of a coherent receiver.

The mixing of the weak signal field and the strong LO field at the front-end of a coherent receiver provides linear amplification and down-converts the optical signal into an electrical output at the intermediate frequency (IF) with gain (usually tens of decibels). With a sufficiently strong LO field, this raises the signal level well above the noise level of subsequent electronics. The sensitivity of the coherent receiver is thus limited by the self noise (i.e., signal shot noise) of the incident signal. Furthermore, because of the spatial mixing process, the coherent receiver is sensitive only to

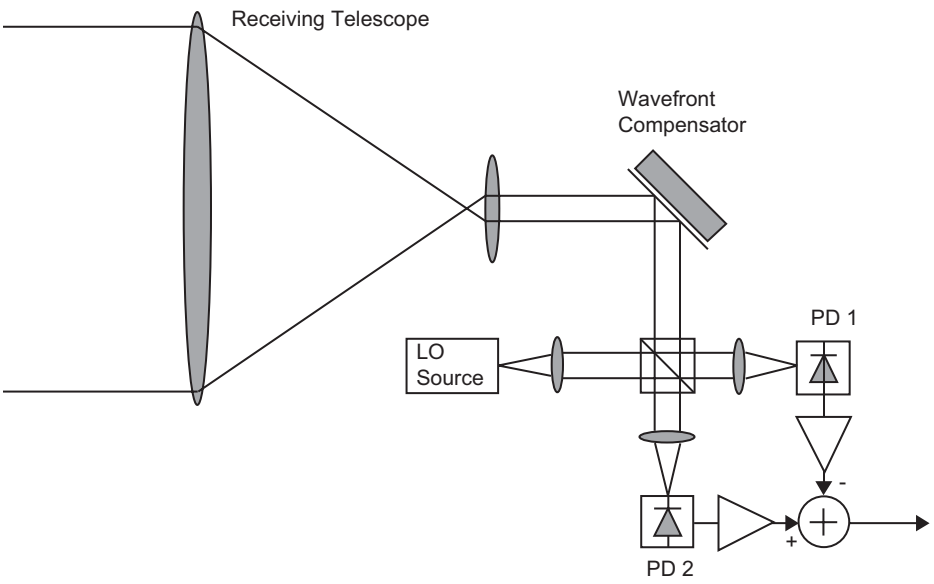


Fig. 2-4. Coherent optical receiver conceptual block diagram.

signal and background noise that falls within the same spatial-temporal mode of the LO. A coherent receiver can, in principle, operate with a very strong background (e.g., with the Sun in the field of view) without significant performance degradation.

The capacity of the coherent optical channel can be written as

$$C_{Coherent} = \left(\log_2 e\right) B \ln\left(1 + \frac{\lambda_S}{B}\right) \approx \left(\log_2 e\right)\lambda_S \qquad (2.2\text{-}4)$$

where λ_S is the rate of detected signal photons, and the last approximation was made in the limit of large signal bandwidth B. Equation (2.2-4) states that the limiting capacity of a heterodyne optical channel is ~1.44 bits per detected photon.

Even though the coherent receiver can in principal provide near-quantum-limited receiver sensitivity, such performance is achieved only through near-perfect spatial-mode matching between the incoming signal and the LO. The added complexity to accomplish the spatial wavefront matching can be very difficult to achieve for a ground-based receiver. This is because the atmosphere effectively breaks up the incident wavefront into a number of coherent cells of sizes approximately the coherence length of the atmosphere r_0. The size of r_0, under typical operating condition, is on the order of 5–30 cm. Although adaptive optics techniques have been developed to partially compensate for the wavefront distortion, effective wavefront correction over the large aperture

diameter envisioned for the deep-space receiver will require an active mirror with a large number of actuators. Because of the complexity of such a system, and because the simpler direct-detection receivers have managed to achieve similar, if not better performance, coherent receivers are not being considered for a ground-based receiver. Instead, the bulk of the development has been focused on the direct-detection receiver.

In a direct-detection receiver, the received optical intensity is detected without extensive front-end optical processing. Figure 2-5 shows a conceptual block diagram of a direct-detection receiver. The incident signal is collected by the receive telescope. A polarization filter followed by a narrowband filter, and a field stop effectively reduces the amount of background noise incident onto the detector.

The capacity of a direct-detection optical link has been studied extensively. When the receiver is capable of detecting individual photons, Pierce [2] first showed that the capacity of the optical channel can be improved by using a modulation format with very high-bandwidth expansion ratios. Subsequent work by Wyner [3] showed that the capacity of a direct detention optical channel in the presence of background can be written as:

$$C = \left(\log_2 e\right)\frac{\lambda_S}{M}\left[\left(1+\frac{1}{\rho}\right)\ln\left(1+\rho\right)-\left(1+\frac{M}{\rho}\right)\ln\left(1+\frac{\rho}{M}\right)\right] \qquad (2.2\text{-}5)$$

where λ_S is the rate of arrival for the detected signal photon (measured in photons/sec), $\rho = \lambda_S / \lambda_B$ is the (detected) peak signal to background power

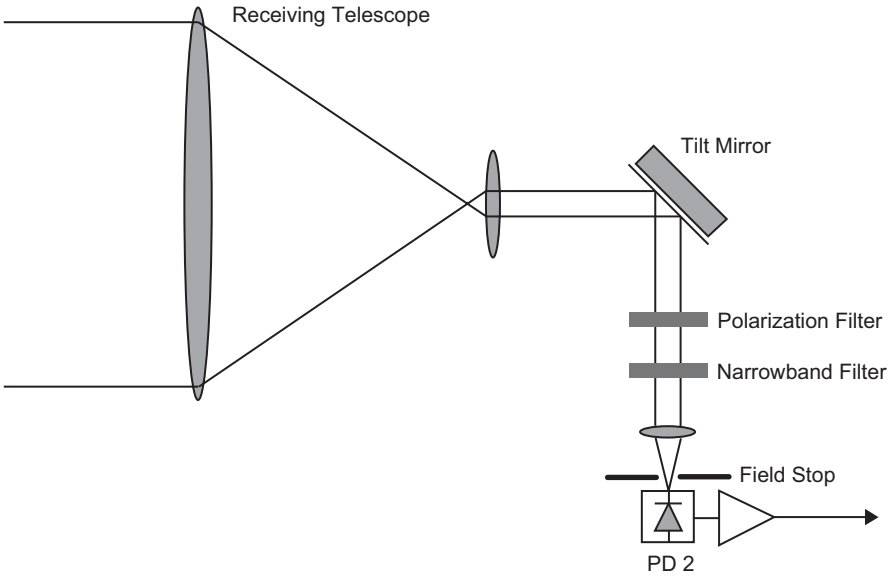


Fig. 2-5. Direct-detection optical receiver conceptual block diagram.

ratio and M is the peak-to-average power ratio of the signal. Figure 2-6 shows a plot of the channel capacity versus the peak-to-average signal ratio for several values of the average signal-to-background noise ratios. It is possible to transmit more than 1 bit/photon at a sufficiently high peak-to-average power ratio [12,13]. In other words, a photon-counting direct-detection receiver can achieve a higher channel capacity than a coherent receiver by using modulation formats that exhibit high peak-to-average power ratios.

Eq. (2.2-5) shows that the capacity of a direct detection optical link using ideal photon-counting detector can be improved by

1) Improving λ_S, or equivalently, increasing the photon detection efficiency for a given receive optical power level,
2) Increasing M, the peak to average power ratio: the performance of the direct detection optical channel can be improved by selecting a modulation format that maintains a high peak to average power ratio,
3) Improving ρ, the signal to noise power ratio by limiting the amount of background optical power detected by the photodetector.

Even though Eq. (2.2-5) was derived from an ideal photon-counting receiver model, the general behavior of the channel capacity remains valid for a wide range of receivers/detectors that are shot-noise limited. That is, the performance of the direct-detection link can be improved by increasing the detector sensitivity, selecting a modulation format with high peak to average power ratio, and reducing the amount of background light detected. Each of these factors is briefly described below.

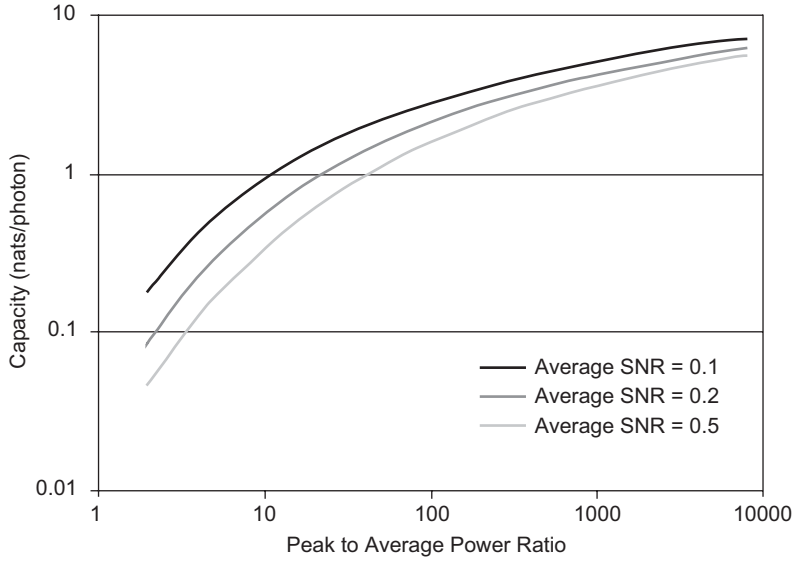


Fig. 2-6. Channel capacity versus peak-to-average power ratio for different signal/background noise ratios.

2.2.2.1 Photon Detection Sensitivity. Improving the photon detection efficiency is an obvious method of improving the channel performance. For a direct-detection receiver, this is generally accomplished by using detectors with internal amplifications, such as avalanche photodiodes (APDs) and photomultiplier tubes (PMTs).

In the limit of a very high amplification gain, the receiver's noise contribution can be ignored, and the receiver is capable of discriminating the individual photon arrival events and counting photons. If the detector contribute negligible amount of dark counts, such a receiver is capable of achieving the channel capacity shown in Eq. (2.2-5). For a more general class of optical receiver that is not capable of discriminating individual photon arrivals, the channel capacity will depend on the noise added by the receiver, including the noise introduced by the amplification process and the thermal noise from the circuit elements. Even if the receiver is not photon-counting, improving the receiver sensitivity can still result in a corresponding increase in the channel capacity. This is accomplished by increasing the detector amplification while controlling the noise introduced by the amplification process (e.g., excess noise) and the thermal/leakage current noise. Refer to Section 6.2 for more detailed discussion of the photon detection.

2.2.2.2 Modulation Format. One practical modulation format to achieve high peak-to-average-power ratio is the M-ary pulse-position modulation (PPM). In an M-ary PPM modulation scheme, each channel symbol period is divided into M time slots, and the information is conveyed through the channel by the time window in which the signal pulse is present. An illustration of the PPM modulation for a simple case of $M = 8$ is shown in Fig. 2-7.

When the transmit laser exhibits a sufficient modulation extinction ratio, the peak-to-average power ratio of an M-ary PPM channel is equal to M, and the capacity of the M-ary PPM channel closely approximates the ideal Poisson channel capacity stated in Eq. (2.2-7). Additionally, when $M = 2^k$, each PPM channel symbol can be mapped directly to a k-bits sequence, thus simplifying the bit-to-symbol mapping problem. For these reasons, except when the

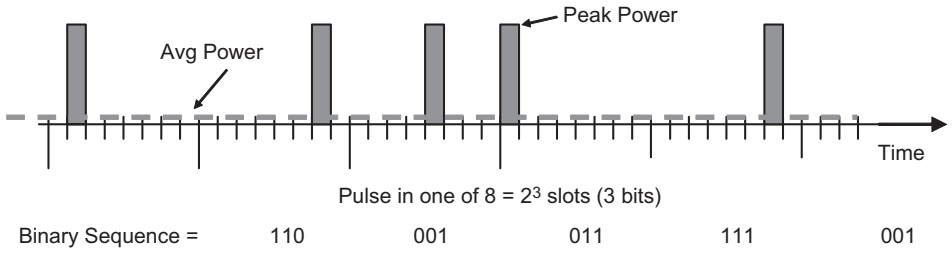


Fig. 2-7. Example of a M-ary PPM modulation with *M* = 8 and straight binary mapping.

transmitter is peak-power limited or when the system is modulation-bandwidth limited, most deep-space optical links analyzed to date had assumed M-ary PPM modulations. [4,5]

2.2.2.3 Background Noise Control. The discussion following Eq. (2.2-5) shows that the performance of the direct detection channel can be improved by reducing the amount of background noise detected by the receiver. For a typical ground based receiver, the sources of background noise include:

1) Diffused (extended) background from the atmosphere, The background irradiance from the extended background can be written as

$$P_{diffuse} = L_\lambda(\theta) A_R \Omega \Delta\lambda \eta_R \tag{2.2-6}$$

where $L_\lambda(\theta)$ represents sky radiance, which is a function of wavelength and solar illumination geometry, A_R is the effective receiver area, Ω is the solid angle field of view in steradians, $\Delta\lambda$ is the optical bandpass, and η_R is the efficiency of the optical receiving system.

2) Planetary or stellar background objects within the receiver field of view. For a point source (e.g., a star) in the receiver field of view, the amount of background power collected by the receiver is written as

$$P_{point} = H(\lambda) A \Delta\lambda \eta_R \tag{2.2-7}$$

where $H(\lambda)$ is the spectral irradiance of the background source, with units of watts per meter squared. micron.

3) In addition to the point sources and extended background sources, another major source of background photons is the scattered light collected by the receive optics. A strong background source near the field of view of the receiver can lead to significant scattering into the receiver field of view. For an optical receiver design with optics under direct exposure to sunlight, the scattering contribution is one of the major background noise sources [6]. The amount of scattered sunlight collected by the receiver can be written as

$$P_{stray} = I_\lambda \Delta\lambda \Omega A \eta_R T(\lambda) BSDF(\theta) \tag{2.2-8}$$

where $T(\lambda)$ represents the atmospheric attenuation and I_λ represents the exo-atmospheric solar constant (0.074 W/cm^{2}μm) and $BSDF(\theta)$ is the bi-directional scatter distribution function as a function of incident angle. The *BSDF* values depend on the surface micro roughness and contamination levels and, in general, they exhibit a power-law dependence to the scattering angle, θ.

In addition to the sunlight scattered off the optical surface, scattered light contribution can also come from scattering off the optomechanical structure inside the optical system. In general, analysis of the scattered light off the mechanical surfaces requires the use of special analytical tools to model the critical surface scattering and the resulting background photon flux. Analysis of the scattered light (other than the optics scattering) is beyond the scope of the current analysis. However, if operation near a bright background source is required, one will need to carefully budget for the scattered background and verify the budget via a series of analytical models and hardware tests.

4) Lastly, the detector itself can contribute "dark currents" which are indistinguishable from the incident photon response. For a well designed system, the contribution of dark current to the overall link budget is generally small.

Background light control is accomplished with a combination of filter, baffle, stops, and masks. For extended background light and out of field stray lights, the amount of background light can be controlled using a field stop that limits the incident light to those from a small angular region around the true direction of the downlink. The diffraction limited field of view of a telescope is approximately $2.44\lambda/D$ which, for a 1 m-class telescope operating at 1 μm, is approximately 2.5 μrad. However, atmospheric turbulence breaks up the incident wavefront into coherent cells with diameters on the order of r_0, the value of which, under typical operating conditions, ranges from a few centimeters to tens of centimeters. The net effect of the turbulence is to redistribute the incident signal energy over an angular region the size of λ / r_0. This effect is shown in Fig. 2-8, which shows the increase in detector area (field of view) required to encompass the downlink energy.

Since $D >> r_0$, a field of view much larger than the diffraction limit is required in order to collect most of the signal energy. Adaptive optics technique can be used to partially mitigate the effect of turbulence at the price of a higher complexity [11]. For the size of aperture being considered for deep-space receivers (several meters), full adaptive optics compensation will require mirrors with upwards of 10^4 actuators.

Another method of controlling the background is to limit the receiver optical bandwidth. This is generally accomplished using a narrowband optical filter. Single optical filters with bandwidths as narrow as 0.05 nm are currently available, and even narrower bandwidth filters have been demonstrated. Finally, the amount of scattered background noise can be controlled by careful control of the surface roughness and cleanliness level on all surfaces that can be directly illuminated by the Sun or by limiting the amount of direct sunlight incident on the optical surfaces.

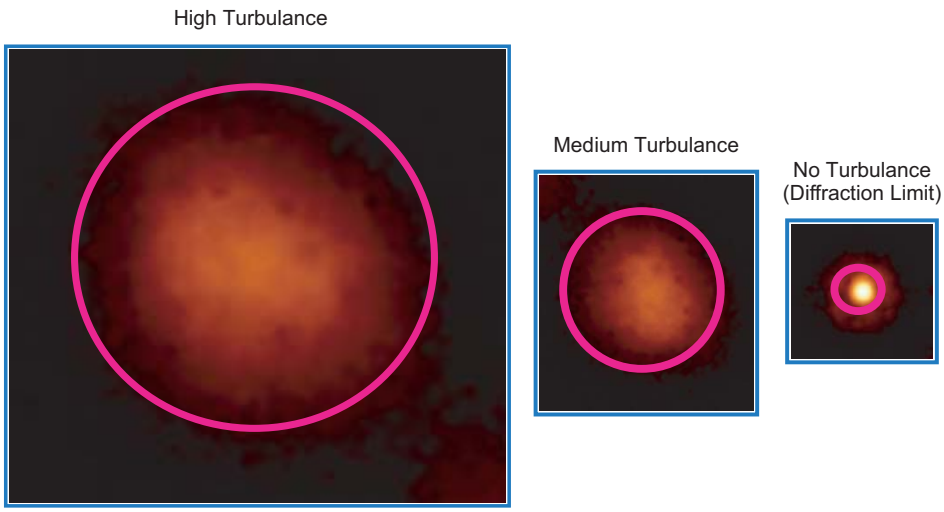


Fig. 2-8. Field-of-view increases induced by turbulence.

2.2.3 Link Design Trades

Design optimization for the optical link is generally accomplished by trading off various design considerations in iterative steps. Some examples of these high level trades include:

2.2.3.1 Operating Wavelength. The operating wavelength of the link is one of the major decisions. This decision is affected by considerations of the following:

1) Link Performance: In general, the antenna gain scales inversely with square of the operating wavelength, and it is more efficient to operate the link at a shorter wavelength. On the other hand, it is easier to maintain the optical quality and high Strehl ratio at longer wavelengths. Because the beamwidth scales inversely with wavelength, it is also easier to maintain pointing and reduce the pointing-induced signal fade at longer wavelength.
2) Availability laser technology and power: Considerations for the laser technology include peak-to-average power ratio, available peak power, electrical-to-optical conversion efficiency, and overall power consumption. Appropriate trades between the available laser technologies, which depend strongly on the operating wavelength, should be conducted to identify the proper design choice.
3) Attenuation and background noise power: The atmospheric loss does not explicitly depend on any link parameter. However, as the attenuation of the atmosphere depends on the absorption and scattering of the signal. The loss will depend on the wavelength choice. The amount of daytime background

noise is also a strong function of the operating wavelength, with a lower day-sky irradiance at a longer operating wavelength.

4) Detector sensitivity: the detector gain, detection efficiency, and excess noise factor determines the sensitivity of the detector in detecting incoming photons. Ideally, one would employ a detector with high-gain, large bandwidth, high efficiency, and low excess noise. However, the availability of such a detector is largely limited by operating wavelength. Silicon detectors, for example, can provide very high-gain bandwidth and low excess noise, but they have very little detection sensitivity at 1.5 μm.

2.2.3.2 Transmit Power and Size of Transmit and Receive Apertures. The power delivery efficiency of the link is proportional to the product of the transmit and receive aperture areas. Consequently, one can trade the size of transmit aperture on the spacecraft, which is typically mass and size constrained, with the size of ground receiver aperture. Furthermore, one can reduce the transmit power requirement by increasing the aperture size. For deep-space missions, the severe mass and power constraints generally lead to a highly asymmetric design. With the flight terminal's transmit power and aperture size limited by the available power and mass margin, a more viable option in improving the system performance is to increase the Earth receive aperture area. While a typical flight terminal has a transmit power of several watts and an aperture diameter of tens of centimeters, the equivalent aperture diameters for the Earth-receiving terminal under consideration generally ranges from a few meters to upwards of tens of meters. Note that since the performance depends on the total area, such an equivalent aperture can be made up from multiple smaller apertures.

The size of the aperture can also affect the pointing performance even though the pointing loss terms, L_{TP} and L_{RP}, do not explicitly depend on the link parameters. Since the beamwidth is inversely proportional to the aperture diameter, larger aperture optics will generally require a tighter pointing accuracy and higher sensitivity toward pointing loss. At the same time, a larger collecting aperture can lead to higher receive signal power and a lower noise equivalent angle.

2.2.3.3 Receiver Optical Bandwidth and Field of View versus Signal Throughput. The link performance can be improved by reducing the amount of optical background. This is accomplished by reducing the optical bandwidth and receiver field of view. Since the optical throughput can depend on the design of the narrowband filter and field of view, appropriate tradeoffs between the narrowband filter bandwidth, receiver field of view, and signal throughput are needed to optimize the link performance.

2.2.3.4 Modulation and Coding. Proper modulation and coding of the optical signal are required to achieve near-capacity performance. A modulation technique with a high peak to average power ratio is needed for the deep-space lasercom system, and optical PPM is generally regarded as an efficient modulation technique of choice. Other modulation techniques with an appropriate peak-to-average power ratio may also be implemented.

Once the modulation format is selected, appropriate channel coding should also be selected. A significant amount of work has gone into the development of channel coding for the optical channel. Earlier work has assumed the use of Reed Solomon (RS) codes that can be naturally mapped to the 2^k-ary alphabet of the PPM symbol. Recently, JPL has proposed the use of a serially concatenated PPM (SCPPM) code for deep-space optical links [9]. These codes can achieve near channel-capacity performance with a photon-counting detector, with a gap to capacity on the order of 0.75–1 dB. An in-depth discussion of the optical modulation and coding can be found in Chapter 4.

2.2.4 Communications Link Budget

As a tool for ensuring that pertinent system parameters related to link performance have been considered, a communications link budget is maintained through the design and built phase of the system development. The link budget is typically represented using a link design control table (DCT), which is a listing of design parameters and the resulting estimated system performance at a specific point in time during the mission. For RF systems, a rigorous and well-established link design procedure exists to calculate the end-to-end link performance and to document the link budget. System designers rely on such a DCT to conduct trade offs between transmit power, aperture size, and other system performance parameters. A similar procedure is used to conduct design tradeoffs for an optical link. Table 2-2 summarizes the typical design parameters that comprise a DCT. An example of a downlink budget from Mars is shown in [10].

2.2.5 Link Availability Considerations

The communications link budget or the DCT is a useful tool in estimating the physical layer link performance (e.g., the link bit error rate). An operational communications link, on the other hand, must also address the issue of link availability. Historically, deep-space RF communications links have achieved an overall link availability of approximately 90 percent. This number includes considerations of station downtime (from equipment failure) and weather-related outages. For RF links, weather-related effects contribute to only a small fraction of the link outages. In contrast, the optical link is much more susceptible to the channel effects, particularly when one end of the link resides within Earth's atmosphere. Additionally, operational constraints of an optical

Table 2-2. Typical design parameters considered in a lasercom design control table.

Link Budget	Parameters
Received signal power	Operating wavelength Link distance Transmit power Transmit aperture area Transmit optics efficiency Transmit Strehl ratio Transmit pointing loss Polarization mismatch loss Receiver aperture area Receive optics efficiency Receiver detector field of view Receiver pointing loss Atmospheric attenuation loss Scintillation-induced loss
Received background power	Receive aperture area Receive optics efficiency Detector field of view Receive optical bandwidth Background spectral irradiance Receive optics scattering behavior Detector dark count
Receiver sensitivity	Detector quantum efficiency Detector noise characteristics • Dark count rate or • Detector Excess and thermal noise Modulation format Coding scheme

link may impose additional link outages. The design of an operational lasercom system, therefore, must address these short-term and long-term outages.

2.2.5.1 Short-Term Data Outages. The optical communications link is susceptible to a number of factors that can contribute short term signal outages, including:

1) Pointing-induced fades: Because of the narrow downlink beamwidth, dynamic pointing error on the downlink can lead to occasional signal fades. The principal sources of this pointing dynamic are the uncompensated platform vibration and the sensor noise that are coupled into the downlink line of sight. During periods of high spacecraft dynamics the uncompensated spacecraft attitude error can also contribute to the pointing-induced signal fade. Depending on the bandwidth of the pointing control subsystem, pointing dynamics-induced fades have a characteristic time constant on the order of several milliseconds to several seconds.

2) Scintillation-induced fades: Atmospheric scintillation can cause variation of received signal power and apparent angle of arrival at time scales on the order of tens of milliseconds. Over the collecting areas typically required for a deep-space optical link, one expects that the effect of downlink scintillation be limited due to aperture averaging effect. On the other hand, the effect of uplink scintillation fades can be quite significant. Even with multiple uplink beams, uplink scintillation fade in excess of 3–6 dB can occasional be observed.
3) Intermittent weather: Intermittent cloud coverage will cause occasional outages of the optical link. For a subsystem design that relies on an uplink laser beacon for pointing the downlink, the occasional cloud outage, if sufficiently long (tens of seconds), can cause the downlink to wander off the desired pointing location. In this case the link availability must account for both uplink and downlink outages. Intermittent weather outages can last from several tens of seconds to days, depending on the site and seasons.
4) Safety-related outages: Safety related outage during aircraft and spacecraft fly-bys can cause uplink outages on the order of several to tens of minutes. If the outage periods exceed the capability of flight terminal to hold its pointing position, then the uplink outage will also translate to downlink pointing outages.

Depending on the outage durations, short-term outages may be addressed using either a data retransmission protocol and/or by interleaving the data over several independent fade periods.

2.2.5.1.1 Signal Fades and Data Interleaving. In the presence of rapid time-varying fades, one can budget a larger amount of link margin to ensure that the probability of a fade with depth exceeding the margin is negligible. Alternatively, for a coded optical link, one can interleave the transmit data such that the signal fade is spread over several code words. A de-interleaver at the receiving end re-assembles the transmit code words. Since the PPM symbols in each codeword might experience a diversity of fades, the occasional deep fades can be effectively corrected by the error-control codes.

In order for the interleaver to be effective, the length of the interleaving period must span a large number of independent fade periods. Due to the high data rate expected for the optical link, interleaving is an effective strategy only for short fades such as those due to pointing error and scintillation fades.

2.2.5.1.2 Retransmission Protocols. A second option to address the occasional signal fade is to rely on the retransmission protocol such as an automated repeat request (ARQ). ARQ schemes are widely used in data communications applications to provide reliable data transmission over an unreliable physical link. Because of the long RTLT involved, simple stop and wait or go-back N

ARQ schemes would likely result in severe bandwidth penalties. Instead, a selective repeat ARQ will most likely be employed. In the selective repeat ARQ scheme, the transmitter continuously transmits the downlink. If any downlink data unit is not acknowledged after a certain period, it is assumed lost and is retransmitted. Alternatively, it is also possible to implement the ARQ scheme in which the corresponding terminal explicitly sends a negative acknowledgement (NACK) signal for the lost frame.

ARQ protocols can be implemented either at the data-link layer or at the transport layer. In either case, a unique sequence number is needed to clearly identify the data unit. The receiving terminal must provide the capability to reorder the downlink frames if they are transmitted out of order (due to a repeat request). Furthermore, In order to implement an effective ARQ scheme, the spacecraft must provide sufficient onboard data storage to buffer the downlink transmission at least over one RTLT. This can drive the data storage requirement on the spacecraft.

2.2.5.2 Weather-Induced Outages. The issue of weather-induced outages is of particular concern for a free-space optical link. For the RF links, the principal effect of weather (other than high wind conditions) is to increase the system noise temperature and link attenuation, and the effect of inclement weather can generally be overcome by increasing the transmit power, or by operating the link at higher link margin. For an optical link, on the other hand, the attenuation due to clouds can be as high as several tens of decibels, and it is generally impractical to provide the link margin necessary to combat cloud-induced signal fade. Consequently, the optical link will generally require cloud-free line of sight (CFLOS) to operate. To achieve the near 90–95 percent availability currently achieved by the RF link will require considerations on the following.

2.2.5.2.1 Weather Availability at the Receiving Site. Selecting the site for the receiving terminal is critical. If the receiver is located above the cloud layer, such as on an high altitude balloon or an orbiting platform, it will be much less susceptible to weather related outages. On the other hand, such a system will have a much higher development and operating cost. Furthermore, as a space-based terminal is much more difficult to service and upgrade, the lifecycle cost of a spaceborne terminal will generally be much higher than the ground-based terminal. For the foreseeable future, therefore, it is likely that the Earth receiver will be located inside the atmosphere, and the location of the ground terminal needs to be carefully selected to minimize the amount of cloud covered days.

The percentage of time a given site can maintain CFLOS with the spacecraft is a function of site location and the season. Some sites also exhibit diurnal variation in cloud coverage. However, single-station weather availability will generally be less than 70 percent, even at outstanding sites such as the southwestern United States. The single-station availability can further

decrease if significant re-acquisition time is required, especially for partly cloudy days.

2.2.5.2.2 Site Diversity with Multiple Ground Stations. Another method of achieving high weather availability is to use site diversity with multiple ground stations. If N stations, located at independent weather cells, are visible from the spacecraft, and each station has a weather availability of p, then the network availability is simply the probability that at least one station has a CFLOS to the spacecraft, and can be written as

$$\text{Network Availability} = 1 - (1 - p)^N \tag{2.2-9}$$

With a large number of ground stations, therefore, one can achieve the required network availability. An example global site placement is shown in Fig. 2-9. With nine sites, each with 67 percent availability, the network can provide a 96 percent availability.

2.2.5.3 Other Long-Term Outages. In addition to weather-related outages, the optical link is expected to experience other long-term outages. One such outage is the solar conjunction (opposition) outage when the Sun-Earth-probe (SEP) or Sun-probe-Earth (SPE) angles are small. At the Earth receiver, low SEP angle implies that the spacecraft is visible when the receiver boresight is close to the Sun. Since the solar radiation is several orders of magnitude stronger than the signal, communications are not possible with the Sun in the field of view for a ground-based direct-detection receiver. However, even when the Sun is not directly in the field of view, scattering due to both the optical surfaces and telescope structure can introduce elevated background levels at small SEP angles to degrade or prevent communications. Furthermore, solar radiation reflected by the telescope can concentrate on the structure and pose a safety hazard on both the facility and the personnel. For the flight terminal, the small SPE angle also implies that the spacecraft's pointing and tracking detector will experience an increase in background noise. This can lead to an increase in pointing error and, at worst case, an inability to detect the Earth image or uplink beacon signal on the focal plane. Good stray-light rejection design is essential to improve the tracking performance at low SPE angle. It should be noted that low SPE angle occurs both during solar conjunction and during opposition. As a result, missions flying optical-communication payloads will likely experience both conjunction and opposition outages; as opposed to RF systems which experience only conjunction outages.

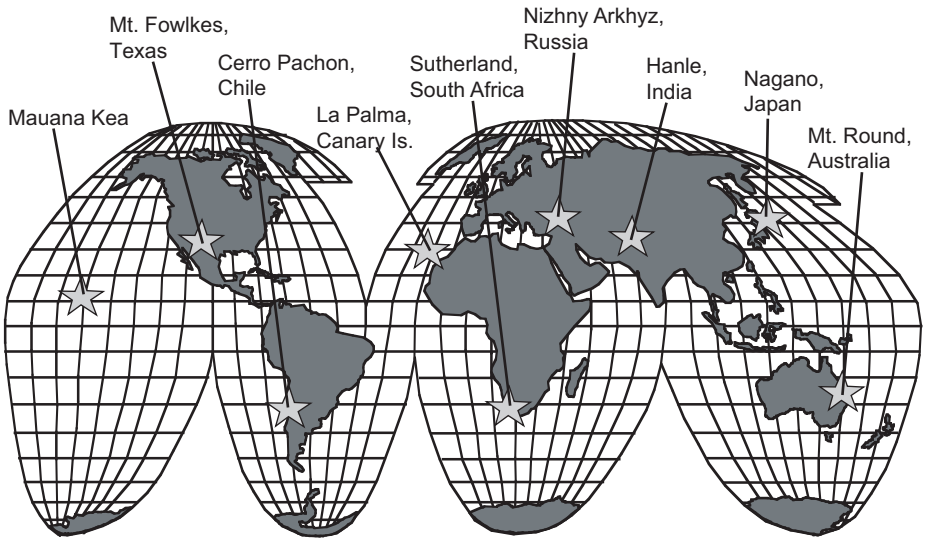


Fig. 2-9. Example multi-site optical network designed for mitigation of weather-induced outages.

In order to limit the conjunction related outages, one must take active measures to limit the amount of scattered sunlight being collected by the detector and to protect the telescope during periods of near-Sun operations. This can be accomplished by (a) employing a solar window to reduce the amount of direct sunlight being collected by the telescope, (b) limiting the amount of direct sunlight on the optical surfaces with baffles and other structures, and (c) controlling the surface quality and contamination level to limit the amount of scattering, and using a combination of Lyot and field stops to limit the out-of-field background.

The solar opposition period also poses a unique challenge to optical communications payloads that rely on the solar-illuminated Earth or the Earth-Moon system for pointing reference. As viewed from the spacecraft, the Earth-Moon system will be barely visible, as only a small fraction of the Sunlit surface is visible from the spacecraft. This reduced photon flux at the receiver can severely limit the tracking frame rate when the solar background interface is most severe. Therefore, even though the Earth receiver will have a favorable background noise condition during solar opposition, the reduced pointing performance at the spacecraft will lead to an effective communications outage.

In addition to the solar conjunction period, long-term laser-communication outages can occur when the spacecraft's attitude is constrained to prevent pointing of the body-mounted flight lasercom terminal at the Earth. Such attitude-constrained periods can occur during nominal mission operations or during periods of spacecraft fault protection. Examples of nominal mission

attitude-constrained periods include the inner cruise period for Cassini when the high-gain antenna (HGA) needed to be Sun-pointed for thermal reasons and the thrusting cruise phase for Deep Space 1 when the spacecraft needed to be pointed along the thrusting vector of the solar electric propulsion system. During attitude-constrained periods, the spacecraft may be prevented from pointing the optical boresight to Earth for very long periods of time. The Cassini inner cruise periods, for example, lasted from October 1997 through February 2000.

For RF systems, mission coverage during these attitude constrained mission phase and during fault protection period is generally accomplished using low gain antennas. Since it is impractical to implement an optical low gain antenna, communications over these attitude constrained mission phase will generally be limited unless the optical system is gimbaled to provide a wide range of coverage or that an auxiliary RF system is used to provide mission coverage during these periods.

2.2.5.4 Critical-Mission-Phase Coverage. A related issue to the long-term mission outage is the requirement for link availability during critical mission phases, such as during orbit insertion burn or during the entry-descend-landing (EDL) mission phase. Coverage during these critical period has been deemed critical due to lessons learned from past mission failures (e.g., Mars Observer and Mars 98 [7]). Unlike weather-related outages that can be overcome by buffering the data onboard the spacecraft, critical-mission-phase coverage will require that the communications link be available at the precise moment of each such maneuver. Given that a ground-based receiver will almost always be susceptible to weather-related outages (unless a space-based receiver is implemented), critical-mission-phase coverage should be accomplished using RF links, and missions flying a lasercom system will generally need to also provide an auxiliary RF link. Such a link may also provide the ability to communicate during any attitude-constrained mission phases, as well as during spacecraft fault protection periods when the ability to precisely point the downlink to Earth using the lasercom terminal may be compromised.

2.3 Beam Pointing and Tracking

Due to the narrow transmit beamwidth, accurate pointing acquisition and tracking are critical to the deep-space laser communications system implementation. For a typical deep-space lasercom terminal, the required pointing accuracy is a small fraction of a microradian. The flight lasercom terminal must achieve this pointing accuracy in the presence of spacecraft platform jitter and attitude-control deadband, both of which can be several orders of magnitude larger than the required pointing accuracy. Inaccurate

beam pointing can result in large signal fades at the receiving site and a severely degraded system performance.

2.3.1 Downlink Beam Pointing

In order to achieve sub-microradian level pointing accuracy in the presence of spacecraft platform jitter and attitude-control deadband, a dedicated pointing control subsystem needs to be an integral part of any flight lasercom system design. Furthermore, design of the pointing control subsystem can impose stringent requirements across the optics, control, and mechanical design of the lasercom terminal. In contrast, the beam-pointing requirement for a RF communication system is much less stringent: a 1-m antenna operating at X-band requires a pointing control accuracy of 0.1–0.5 deg, and the same antenna operating at Ka-band requires a pointing accuracy of a few milliradians, both of which are well within the capability of current spacecraft attitude-control subsystems.

The problem of pointing the narrow deep-space return beam can be divided into a combination of jitter isolation/rejection, and precision beam-pointing functions. The former is the problem of isolating and rejecting the spacecraft jitter and attitude deadband in order to provide a stable transmit line of sight (LOS) in inertial space. The latter is the problem of pointing the stabilized line of sight in the direction of the Earth receiver.

2.3.1.1 Jitter Isolation and Rejection. To achieve a stable line of sight, the lasercom terminal must properly isolate and reject the spacecraft platform jitter and spacecraft attitude control errors. This is accomplished using a combination of vibration isolators and a pointing stabilization control loop as shown in Fig. 2-10.

Vibration isolation is an effective method of limiting the amount of high-frequency jitter. Figure 2-10 shows a set of vibration isolators that provides the principal mechanical linkage between the flight lasercom terminal and the host spacecraft. The platform jitter is low-pass filtered by the isolators, and the high frequency jitter components are severely attenuated. This effectively reduces the required tracking loop bandwidth, which in turn reduces the requirements on the tracking sensors and line-of-sight stabilization elements.

After vibration isolation, the residual jitter present at the lasercom terminal can be controlled with a pointing stabilization control loop, which must provide sufficient control bandwidth and dynamic range to compensate for the residual jitter. This is accomplished by measuring the jitter at the appropriate update rate and accuracy.

The update rate required for the jitter measurement is in general an order of magnitude higher than the required closed-loop bandwidth of the jitter-control loop. The latter will depend on the effectiveness of the vibration isolation. The

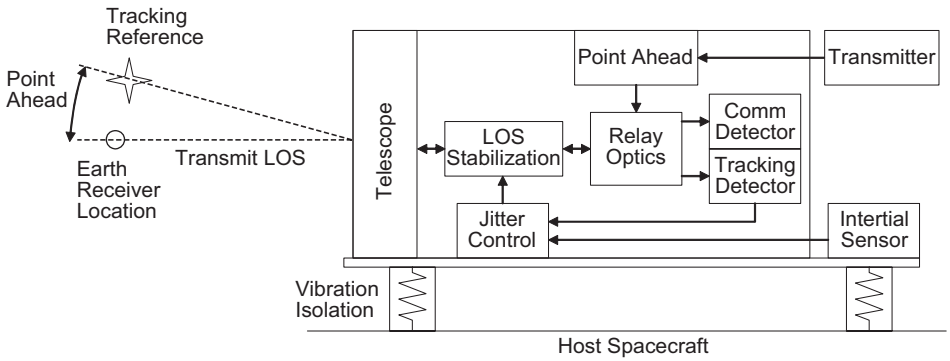


Fig. 2-10. Block diagram showing jitter isolation and rejection for a lasercom terminal.

implementation of the jitter sensor can be accomplished using a variety of means. For a near-Earth lasercom system, this is generally accomplishing by measuring the line-of-sight jitter using a beacon laser signal from the remote (ground) terminal. The narrow spectral line width of the beacon laser allows efficient background noise rejection and, because of the short range involved, the beacon signal is generally able to provide a sufficient signal-to-noise ratio to operate at the desired update rate and noise equivalent angle (NEA). For a deep-space system, on the other hand, the large link distance implies a reduced amount of beacon power available at the flight terminal. Furthermore, for a ground-based beacon, atmospheric turbulence experienced by the uplink can lead to deep and frequent signal fades that are difficult to compensate when the only reference for jitter measurement is the optical beacon.

Instead of relying only on the beacon signal from the ground terminal, the required sensing update rate and measurement accuracy can be accomplished using a hybrid pointing architecture, which utilizes a combination of inertial sensors, celestial references, and an uplink beacon. As shown in Fig. 2-10, measurements from the inertial sensors are blended with optical line-of-sight measurements derived from celestial references and/or uplink beacons to provide the jitter measurements. The inertial measurements are generally accurate at higher frequencies, but they have a lower frequency cutoff, whereas the celestial sensor and/or beacon measurements are limited by the available power to low-frequency measurements. The blending of the inertial sensor with celestial/beacon signals allows adequate jitter sensing over the frequency range of interest.

2.3.1.2 Precision Beam Pointing and Point Ahead. The net effect of the vibration-isolation and jitter-compensation control loop is to provide a stabilized line of sight referenced to the beacon (or celestial sensor) direction. The pointing and tracking subsystem must then point this LOS-stabilized

downlink signal accurately at the Earth receiver. This is accomplished by accurately referencing the position of the celestial and/or beacon signals to the Earth receiver location and then applying an appropriate amount of open-loop correction (point ahead) to account for the relative velocity between the flight lasercom terminal and the Earth receiving terminal.

Pointing architecture that relies on a ground-based beacon has the advantages that the beacon is well referenced to the receiver location and is generally located within the field of view of the optical system. However, within the United States, transmission of a laser beacon signal through the atmosphere is subjected to safety coordination with the United States Federal Aviation Authority (FAA) and with the United States Air Force Laser Clearing House (LCH), and the uplink session may be punctuated to prevent illumination of aircraft and spacecraft. Since the flight terminal relies on the ground-based beacon to provide the pointing reference, the potentially nondeterministic beacon outages can lead to occasionally large pointing error and can interrupt the downlink communication session; which must be addressed via proper operational workarounds (e.g., with retransmission protocols to ensure reliable downlink data delivery).

An alternative to the ground-based beacon is to use the Earth image or celestial references to provide the desired pointing reference. This architecture has the advantage that it allows the flight terminal to point the downlink at the receiving terminal without requiring an uplink beacon. This, in turn can greatly simplify mission operations. However, practical implementation of a beaconless pointing concept is very difficult. Earth image tracking is susceptible to albedo variation, which can cause a random and time-varying shift of the Earth image centroid from its geometric center site reference. Furthermore, since the Earth images fall within the same spectral band as the solar radiation, proper filtering of the solar background can be very difficult to accomplish, especially at low Sun-spacecraft-Earth angles. Beaconless pointing using celestial reference is equally difficult to implement as it will require at least a reference source within the optical field of view. This, in turn, drives the optical design. For outer planetary missions (i.e., Jupiter and beyond), Earth will only be a few degrees from the Sun, and solar stray light can lead to an elevated background level and a higher noise equivalent angle. Although separate tracking sensors with boresight pointed away from the downlink is a possible option, practical implementation of this concept will require maintaining the precision alignment between the boresights of the lasercom terminal and the celestial tracking sensor, and can greatly complicate the mechanical design of the optical system.

In general, the location of the pointing reference is different than that of the receiving station, and the flight lasercom terminal must off-point from the pointing reference in order to position the downlink over the receiving terminal. Even if a co-located beacon is used with the receiving terminal, the relative

motion between transmit and receive terminals will require that the transmit signal be off-point from the apparent beacon location so that the return signal will arrive at the receiving terminal at the proper spatial-temporal location. This pointing offset is known as the point-ahead angle. Because it is generally impractical to offset the beacon and the receiver location at precisely the point ahead angle, point-ahead function is usually accomplished open-loop. The point-ahead angle required for deep-space missions is typically on the order of several hundreds of microradians, compared to the tens of microradians for near-Earth lasercom systems. This large point ahead drives both the field of view of the optics as well as the design tolerances of the optical system as it needs to maintain its performance over the relatively large angular separation between the pointing reference (beacon) and the desired downlink direction.

2.3.2 Uplink Beam Pointing

Uplink transmission from the ground to the flight terminal is needed to provide an optical command path and to provide a pointing reference if the flight terminal relies on a beacon signal to point the downlink. In both cases the ground terminal must deliver the required irradiance at the flight terminal while minimizing the magnitude and frequency of the signal fade. The latter is due to the time-varying higher order modes in the wavefront distortion introduced by the atmospheric turbulence which, when propagated to the far field, can result in strong fluctuations of the far field irradiance.

In order to accurately point the uplink at the spacecraft, one must provide the requisite pointing reference. This pointing reference can be a nearby star or planet. Alternatively, the optical downlink itself can be used as a pointing reference, although since the uplink is also used to provide the pointing reference, one must carefully address the pointing acquisition issue. Since the atmospheric turbulence effectively broadens the transmit signal (by breaking up the wavefront into small cells with coherence diameter of approximately r0, the Fried's parameter), the required uplink pointing accuracy is generally looser than the downlink (on the order of a few microradians). Such a pointing accuracy is within the capability of a well-instrumented telescope; provided that a proper mount-calibration has been conducted using stellar references nearby to the spacecraft position.

Even though the required uplink pointing accuracy can be achieved, the presence of the uplink signal scintillation can affect both the communications and the beacon-tracking performance. For the communications link, the occasional signal fades translate to periods of high error rates, which can be controlled through coding and data interleaving. When the uplink is used to provide a pointing reference, the signal fades translate to periods of higher noise equivalent angle and can degrade the pointing control performance.

The period and duration of a signal fade is a function of the turbulence parameter r_0 the wind speed, and the uplink signal configuration (i.e., the beam divergence and number of transmitted beams). In general, fluctuations in the far-field irradiance profile introduced by turbulence can be reduced by increasing the beam divergence, and by transmitting multiple mutually incoherent beams. If these beams are spatially separated to the extent that they pass through portions of the atmosphere that are largely uncorrelated, the likelihood of all the beams simultaneously being directed off axis will be substantially reduced relative to the likelihood of the same result for a single beam. In general, increasing the number of beacon laser beams can lead to fewer scintillation fades, which in turn, can improve the spacecraft pointing control performance. A recent study by the Optical Science Company [8] indicated that 8–16 independent beams will generally lead to very infrequent fades. Figure 2-11 shows the result of a computer simulation which plots the cumulative probability versus the on-axis power at the far field. As can be seen from this study, the probability of experiencing a large signal fades decreases rapidly with increasing number of uplink beams.

In addition to the probability of fades, the frequency and duration of the fades are also significant performance drivers. The frequency and duration of fades are related to the temporal nature of the turbulence by which they are induced. When turbulence changes more rapidly, fade events occur more frequently, but with a correspondingly shorter duration. The time scale of the turbulence evolution can be characterized in terms of the Greenwood frequency, f_G,

$$f_G = 0.255\left[k^2 \sec\psi \int dh C_n^2(h) v(h)^{5/3}\right]^{3/5} \qquad (2.3\text{-}1)$$

where C_n^2 is the altitude-dependent turbulence profile and $v(h)$ is the wind profile. Turbulence-induced events will generally occur on a time scale of roughly $1/f_G$. The value of the Greenwood frequency, under typical conditions, is on the order of 30 Hz. That is, the turbulence-induced events will tend to occur on a time scale of tens of milliseconds. In order to employ a ground-based beacon as a pointing reference, the flight terminal pointing-control loop must be capable of tolerating the pointing-induced signal fade. That is, it must either have sufficient power margin or a sufficiently low pointing bandwidth such that the pointing-induced fades can be averaged over a sufficiently long period.

2.3.3 Pointing Acquisition

Prior to link establishment, the flight terminal and the beacon transmit terminal must establish the line-of-sight reference to each other. Since the

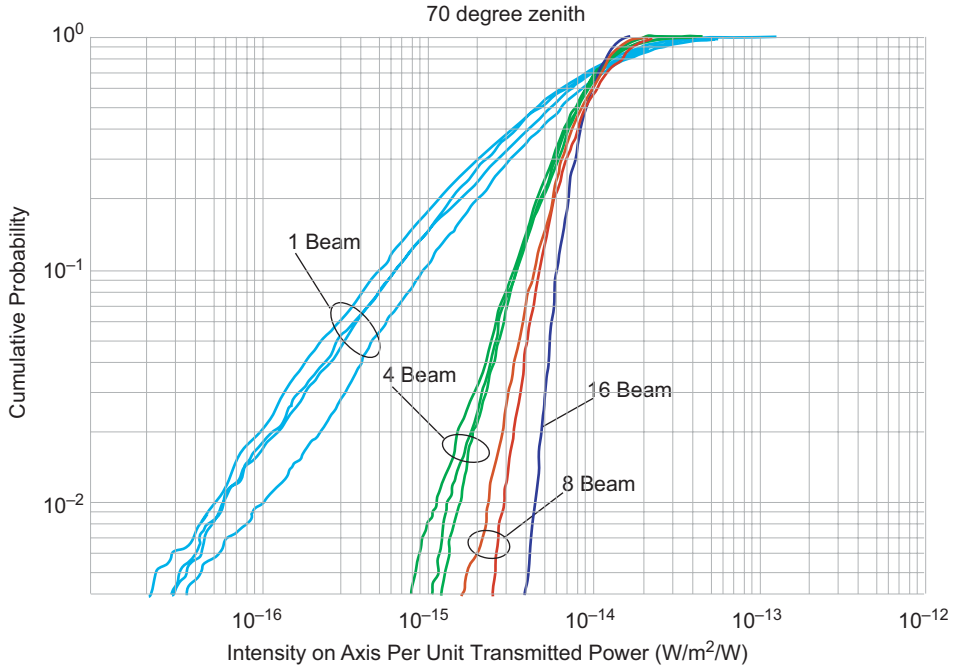


Fig. 2-11. Plot of cumulative probability versus on axis power from [7] illustrating the advantages of multiple beam uplinks.

initial pointing uncertainty can be much larger than the desired pointing accuracy, a separate pointing acquisition process is generally required to achieve this mutual line-of-sight reference. For a near-Earth lasercom system, this pointing acquisition process can be accomplished in several ways. In one concept, one terminal (the initiating terminal) slowly scans the initial uncertain region with its transmit signal. At the same time, the target terminal searches over its pointing field of view for the beacon signal. Once the beacon signal is detected, the target terminal then transmits a returns signal to the initiating terminal which, upon detecting the return link, stops its acquisition scan. Figure 2-12 illustrates this process.

The performance of this step-scan acquisition scheme depends strongly on the RTLT. At each scan step, the initiating terminal must wait at least one RTLT before proceeding. Furthermore, the drift in attitude for the transmit platform over the scan period must be smaller than or comparable to the scanning beamwidth in order to avoid missed acquisition. For deep-space missions, the long RTLT makes it impractical to employ such an acquisition scheme.

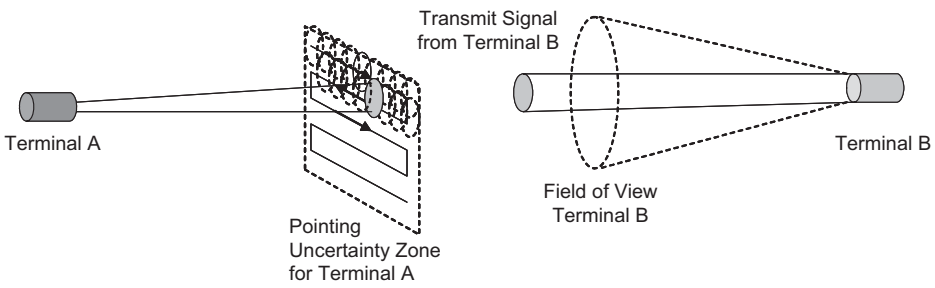


Fig. 2-12. Pointing acquisition concept in which one terminal (terminal A) slowly scans its transmit signal while at each step the other terminal (Terminal B) scans through its entire uncertain region.

Alternatively, one or both terminals can transmit a broadened beacon signal to illuminate its initial uncertainty region, while the other terminal searches for this broadened beacon signal. A variation of this scheme is for the terminals to rapidly scan the uncertainty zone with a narrow signal to provide an effectively "broadened" beacon. This "parallel" acquisition scheme is suitable if one terminal has a small initial pointing uncertainty to permit transmission of a broadened beacon while still providing adequate irradiance at the other terminal for pointing acquisition. For a deep-space lasercom system employing a ground transmit beacon, it is possible to transmit a high power signal with a suitably broadened beam to allow for adequate pointing acquisition at the flight terminal. Such a pointing scheme, however, may drive the beacon power as well as initial pointing uncertainty of the ground transmit terminal.

2.4 Other Design Drivers and Considerations

In addition to considerations on pointing and data links, a number of additional considerations may also affect the design of the flight lasercom terminal.

2.4.1 System Mass and Power

Because of the high launch cost associated with deep-space missions, flight system mass and power are generally considered to be premiums. As a result, deep-space telecom links are generally highly asymmetric: with a smaller aperture and limited transmit power on the flight terminal, and a larger aperture and higher transmit power on the ground. This approach minimizes the flight system mass and power consumption while maintaining the overall link performance. The asymmetric design is also more cost effective since it is easier to develop the large aperture and higher power transmitter on the ground. Furthermore, it is possible to amortize the cost of a larger aperture ground station over several missions. However, since the aperture and transmit laser power account for only a fraction of the overall flight terminal mass and power,

continuing reduction of the flight system aperture may not significantly reduce the overall mass and power consumption. Furthermore, even though the allowable pointing error increases with smaller aperture, a smaller aperture collects a reduced amount of received tracking signal, and hence, can have a higher sensor noise equivalent angle on the tracking detector. When designing a deep-space lasercom link, therefore, care must be taken to evaluate the potential design with respect to the impact on the overall system mass and power consumption.

2.4.2 Impact on Spacecraft Design

The tight pointing requirements may affect the design of the spacecraft bus and impose constraints on mission operations. In addition to the known mass and power constraints, special care will be required to design the spacecraft in order to accommodate the flight lasercom terminal. Some of the considerations include:

1) Platform Jitter Environment: The tight pointing requirement will lead to requirements on the spacecraft vibration environment, which in turn can impose constraints on the mass balancing and structural stiffness of the spacecraft.
2) Configuration: Providing a clear optical line of sight of the lasercom terminal may impose constraints on the spacecraft configuration. This is particularly true for a body-mounted lasercom terminal, which must be pointed toward Earth within the field of view of the optical system. If an RF link is also present, the line of sight of the optical system will also need to be co-aligned with the high gain antenna boresight in order to support simultaneous RF–optical downlinks. Additionally, temperature control requirements of the lasercom terminal may impose field-of-view requirement on the thermal radiator.
3) Attitude-Control Accuracy: The attitude-control performance of the spacecraft must be sufficiently tight such that the sum of attitude uncertainty, control deadband, and point-ahead angle, are smaller than the field of regard of the lasercom pointing-control subsystem. Furthermore, depending on the pointing-control loop bandwidth, there may be constraints placed on the maximum allowable attitude rate of the spacecraft over which the desired pointing accuracy can be achieved.
4) Data Storage and Management: For reliable operation over the optical link using an ARQ protocol, the amount of data storage onboard must be greater than the expected downlink data volume over the RTLT plus ground data processing time. For a flight lasercom terminal operating at tens of megabits per second, such a data storage requirement can be a significant design driver.

2.4.3 Laser Safety

Consideration of laser safety can also be a design driver for the optical ground station. In general, laser safety considerations are limited to those for the uplink since the downlink signal, after propagating through the deep-space distance, is generally much weaker.

Personnel safety is the first priority, and an operational facility needs to comply with known safety guidelines. In addition to safety protection for the operating personnel, the system design and operation must also address the issue of aircraft and spacecraft avoidance. Emission of all Class 4 lasers above the horizon requires coordination with the United States Air Force Space Command Laser Clearing House (LCH) and with the regional FAA office for laser radiation above the maximum permissible exposure (MPE) outside restricted airspace. The regional FAA office is responsible for evaluating and determining the effect of outdoor laser operations on users of the navigable airspace (NAS). Regional offices conduct an aeronautical review of all laser operations to be performed in the NAS to ensure that these types of operations will not have a detrimental effect on aircraft operations. Requests for laser operations are evaluated by the regional offices having jurisdiction over the airspace and coordinated, if necessary, with the affected facility.

The LCH acts as the focal point to authorize laser emissions into space which may result in interference or damage to United States or foreign satellite payloads. The LCH maintains the laser facility data base, receives laser facility emission requests, determines waiver status, sends approval/denial/restrictions to the laser facilities, and processes accidental illumination information. After receiving data on the conditions of uplink emission, LCH either grants a blanket waiver for the laser or coordinates to determine safe laser firing times. The Predictive Avoidance (PA) safe firing windows provide the laser facility with safe laser start/stop times ensuring no satellite payloads will be unintentionally illuminated. LCH monitors changes in space activity and may update issued PA windows.

Due to the long safety range for high-power laser operations, the design and operations of any ground-based laser beacon must carefully consider the issue of laser safety, which may impact site selection and operational strategy for the optical link (for example, can the system tolerate occasional safety shut down). Refer to Section 6.1 for more detailed discussions on laser safety.

2.5 Summary

Because of the long distances involved, deep-space lasercom link implementation is significantly more difficult than its near-Earth counterpart. In order to deliver performance comparable or better than current state-of-the-art RF systems, the deep-space optical link will need to achieve a performance (measured in data rate-distance squared product) that is 50 dB or better than the

performance achieved by the current state-of-the-art near-Earth system (i.e., 10 Gbps from GEO). This drives the transmit power and aperture sizing; the receiver's photon detection efficiency, modulation and coding; and the background rejection capability. Communications link performance considerations also lead to the use of data-interleaving and retransmission protocols to mitigate the effects of short-term outages introduced by scintillation and pointing-induced fades. Finally, considerations on the link availability also lead to ground system designs with multiple site diversity to mitigate weather-related outages.

The large link distance also drives the design of the beam pointing and acquisition. The pointing architecture used for near-Earth lasercom systems cannot be easily extended to deep-space distances due to the large propagation loss and long RTLT. Deep-space lasercom pointing will in general rely on a hybrid architecture involving the use of vibration isolators, inertial sensors, and pointing beacons.

The fact that the communications link performance and pointing acquisition and tracking considerations drive the overall lasercom system design is illustrated in Table 2-3, which shows the key performance parameters for the major flight and ground subsystems that are affected by these design drivers. Furthermore, because of the large number of common parameters, the design of the communications link is tightly coupled to that of the pointing architecture. As a result, a practical design of the lasercom system must consider performance of both the communications link and the pointing acquisition and tracking.

Table 2-3. Design considerations for the key subsystems and assemblies.

	Communications Link Performance	Pointing Acquisition & Tracking
Flight transmitter	Downlink wavelength Achievable peak and average power Modulation extinction ratio	
Flight optomechanics	Aperture size and obscuration Optics efficiency Receive optics bandwidth Transmit-receive isolation Stray light characteristics (surface quality + cleanliness) Optomechanical structural stability Transmit optics, Strehl Pointing bias and jitter	Aperture size and obscuration Optics efficiency Receive optics/solar rejection bandwidth Transmit/beacon receiver Isolation Stray light characteristics (surface quality + cleanliness) Optomechanical structural stability Vibration isolation bandwidth Optics field of view/field of regard, LOS stabilization mechanism (steering mirror) Precision point-ahead mechanism
Flight electronics	Modulator and encoder Data interleaver Downlink protocols Uplink data demodulator + decoder	Pointing control loop bandwidth and residual error Inertial sensor bandwidth and accuracy Celestial reference/beacon sensor bandwidth and noise equivalent angle
Flight receiver	Detector noise characteristic Receiver field of view (FOV)	
Spacecraft interface	Platform jitter and rate Data storage Mass and power allocations Spacecraft command and data interface Applications layer protocol stack	Platform jitter and rate Operational attitude constraints Pointing ephemeris
Ground receive optics	Downlink wavelength Aperture size and obscuration Optics efficiency Narrowband filter bandwidth Detector field of view Stray light control (surface quality and cleanliness) Receiver pointing bias and jitter	
Ground detector, receiver and decoder	Operating wavelength Detector noise characteristics Modulation format Coding Data de-interleaver Downlink protocols	
Ground network	Single vs. multiple aperture, Site diversity	Single vs. multiple aperture Site diversity
Ground beacon/uplink	Uplink wavelength Beacon pointing accuracy Beacon power Beam divergence Number of uplink beacons Laser safety Uplink data modulation and coding Uplink protocols	Uplink wavelength Beacon pointing accuracy Beacon power Beam divergence Number of uplink beacons Laser safety

References

[1] *National Aeronautics and Space Administration 2003 Strategic Plan*, 2003.

[2] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The Capacity of the Photon Counting channel," *IEEE Transactions on Information Theory*, vol. IT-27, no. 1, pp. 61–77, January 1981.

[3] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel – Part I, *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1449–1961, November 1988.

[4] J. Hamkins, S. Dolinar, and D. Divsalar, "Optical Channel Capacity Sensitivity," *The Telecommunications and Mission Operations Progress Report 42-143, July–September 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–16, November 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[5] B. Moision and J. Hamkins, "Constrained Coding for the Deep-Space Optical Channel," *The Interplanetary Network Progress Report 42-149, January–March 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–29, May 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[6] P. R. Spyak and W. L. Wolfe "Scatter from Particulate-Contaminated Mirrors, Part 4: Properties of Scatter from Dust for Visible to Far-Infrared Wavelengths," *Optical Engineering*, vol. 31, no. 8, pp. 1775–1784, August 1992.

[7] *Mars Polar Lander/Deep Space 2 Loss–JPL Special Review Board Report*, JPL D-18709 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 22, 2000.

[8] N. Steinhoff, *Irradiance and Fade Characteristics for the JPL MLCD Program*, the Optical Science Company, Report No. TR-1657, Anaheim, California, May 2004.

[9] B. Moision and J. Hamkins, "Deep-Space Optical Communications Downlink Budget: Modulation and Coding," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–28, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[10] A. Biswas and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, California, pp. 1–38, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[11] K. Wilson, M. Troy, M. Srinivasan, B. Platt, V. Vilnrotter, M. Wright, V. Garkanian, H. Hemmati, "Daytime Adaptive Optics for Deep Space Communications," Space Activities and Cooperation Contributing to All Pacific Basin Countries (10th ISCOPS), vol. 117, *Advances in Astronautical Sciences*, [Eds. P. M. Bainum, L. Furong, and T. Nakajima], American Astronautical Society, 04-401, p. 481, 2004.

[12] J. R. Lesh, J. Katz, H. H. Tan, and D. Zwillinger, "2.5-Bit/Detected Photon Demonstration Program: Description, Analysis and Phase I Results," *The Telecommunications and Data Acquisition Progress Report 42-66, September–October 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 115–132, December 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[13] J. Katz, "2.5 Bit/Detected Photon Demonstration Program: Phase II and III Experimental Results," *The Telecommunications and Data Acquisition Progress Report 42-70, May–June 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 95–104, August 15, 1982.

Chapter 3
The Atmospheric Channel

Abhijit Biswas and Sabino Piazzolla

An optical receiving network that can provide availability similar to the National Aeronautics and Space Administration's (NASA's) existing radio frequency (RF) Deep Space Network (DSN) is being considered for future planetary exploration, where large data volumes acquired using advanced sensors must be returned to scientists on Earth. Optical groundbased receiving networks are an option for realizing this objective. The interaction between laser beams and the atmosphere must be taken into account in order to fully exploit this option. Alternatively, atmospheric effects on the laser beam can be partially or completely eliminated by resorting to either airborne or spaceborne optical receivers. The reason for considering groundbased networks, however, lies in the perceived reductions in cost and risk. Deploying a receiving station in space with the flexibility of re-configuring receiver settings will be expensive and vulnerable to single-point failures. With the maturity in technologies for deploying relatively large apertures on airborne and orbital platforms the cost and feasibility of achieving spaceborne operations seems very likely. However, current estimates for deploying a single fully functional receiver in Earth orbit are comparable to the cost of building an entire ground network, a fact that motivates a close study of the prospects and consequences of implementing a ground network, at least until such time as the cost of deploying receivers in space becomes more competitive.

This chapter describes the effects of Earth's atmosphere upon laser communication beams. Clouds, fog, haze, mist, and other precipitation in the atmosphere cause strong attenuation of the laser signal and cause communication outages. Fortunately, weather diversity or the global cloud

distribution patterns can provide, with a reasonably high likelihood, a cloud free line of sight (CFLOS) for the optical link to at least one location on the ground. Note that, depending upon the design of the optical communication system, CFLOS must be maintained not only for reception of the laser beam from space, but also for transmission of a pointing reference beacon and/or an uplink command laser. This can be an important consideration when the receiving and transmitting terminals are not co-located on the ground. Analysis of local and global statistical distribution of clouds provides the quantitative estimates of CFLOS described in Section 3.1.

Even when CFLOS can be maintained, atmospheric effects continue to play a critical role in the link performance. Spectral attenuation due to absorption and scattering of the laser signal by atmospheric molecular and aerosol constituents determine the laser signal that can be received from or transmitted to space. Therefore, choosing laser wavelengths that do not coincide with strong absorption bands in the atmosphere is critical. Section 3.2 elaborates upon spectral attenuation predictions with a description of measurements made at Jet Propulsion Laboratory (JPL) to verify the predictions and their statistics.

Though optical reception of laser communications has similarities with low-light detection applications extant in astronomy and satellite laser ranging, a significant point of departure is the need to operate the links during day as well as night with a relatively high probability of correct detection. Therefore, sky radiance, or sunlight scattered and reflected by atmospheric constituents, plays a major role in determining link performance. Under the most stressing circumstances, the ground station design must support links when the spacecraft angular separation from the Sun, or Sun-Earth-probe (SEP) angle is approximately 2–3 deg. The sky radiance usually increases with angular proximity to the Sun and is influenced by the aerosol concentration and airmass through which communications must be maintained. Atmospheric modeling software can provide reasonable predictions of sky radiance under a varying set of assumptions. The background photon flux resulting from sky radiance, especially at low SEP angles can be considered to have a twofold contribution. First, the field of view (FOV) of sky imaged onto the detector by the light-collection system is directly proportional to the background. Second, light from outside the FOV (or “stray light”) is scattered and reflected onto the detector. The latter contribution is highly dependent upon the light-collection system used. The light-collection system must necessarily use properly designed narrow band-pass optical filters to minimize both the in-band light and leakage of out-of-band background light that is incident on the detector. Section 3.2 below also presents a discussion on models used to predict sky radiance and a comparison between the predictions and the field data.

NASA’s deep space programs so far have relied upon a global network of RF antennas to capture the signal return from the spacecraft. However, for the case of optical signal return from deep space, new architectural and system

problems are encountered because of atmospheric channel effects. Section 3.3 introduces the problem of deploying a ground optical deep-space network and its site selection.

The atmosphere is a dynamic medium with a randomly varying refractive index to a propagating laser beam. The resulting perturbations of the laser beam wavefront originating from space or being transmitted from the ground are broadly called atmospheric "seeing" effects. Thus, atmospheric "seeing" effects result in solid angles, or equivalently, the FOV of the ground light-collection system being many times the diffraction limit, assuming perfect optics. The penalty for the increased FOV is an increase in background light. Especially during daytime optical links, near the Sun the penalty can be severe. Other effects include the irradiance fluctuations of the laser beams received and transmitted, tilts and wander of the beam fronts, and spreading of the beam widths. These effects, and their impact on link performance, are discussed in Section 3.4

3.1 Cloud Coverage Statistics

Cloud opacity is an atmospheric physical phenomenon that jeopardizes optical links from deep space to any single ground station. Clearly, when clouds are in the line-of-sight, the link is blocked. Therefore, ground receiving telescopes need to be located in sites where cloud coverage is low and statistically predictable. Moreover, to guarantee continuity of data delivery from deep space to ground, while the Earth is rotating, a global network of telescopes is necessary. The selection of the sites for telescopes belonging to an optical deep space network (ODSN) is driven by considerations based, among other factors, on cloud-cover statistics.

To characterize the atmospheric channel and global cloud coverage, a number of resources are actually available to the scientific community. International agencies, institutions, and programs have made available sources of weather data and cloud coverage around the globe. For instance, the International Satellite Cloud Climatology Project (ISCCP) [1] is one source for weather data that can be utilized for selection of optimal telescope sites. ISCCP extracts and elaborates data from a multitude of weather satellites, e.g., Geostationary Operational Environmental Satellite (GOES), European weather observation satellite (METEOSAT), Geostationary Meteorological Satellite (GMS), Indian National Satellite (INSAT), and National Oceanic & Atmospheric Administration (NOAA) polar orbiting satellites. Another source of atmospheric data is the NOAA National Climatic Data Center (NCDC) that can provide surface observation data from observation sites distributed all around the globe [2].

In this chapter we present two different approaches to elaborate statistical characterization of the cloud coverage. First, we present a study of cloud

distributions in the United States Southwest using surface observation data from NCDC. Then, we discuss the results of ongoing research by TASC Inc. [3] under JPL contract and based on satellite observations.

3.1.1 National Climatic Data Center Data Set

The NCDC is the sole climatic record center for the Department of Commerce. The NCDC collects, prepares, and distributes climate data regarding the United States, and the NCDC is also responsible for the United States branch of the World Data Center (along with Russia, Japan, and China).

Among the different types of environmental and weather observations collected and maintained by the NCDC, the work presented here is specifically based on the elaboration and processing of surface observation data. Surface observations are meteorological data that describe the climate of an area (or a site) where an observation station is located. Surface observations for each site indicate temperature, humidity, precipitation, snowfall, wind speed and direction, atmospheric pressure, visibility, and other kinds of weather conditions, including obscurations. The observations are (in general) made hourly, recorded, and collected by a certified operator. The data that we analyzed are in a format DATSAV3 (NCDC designation).

Essentially, the DATSAV3 format consists of rows of data, where each row contains the weather observations made at a specific moment of the day. To indicate the cloud coverage of the celestial dome, a station operator uses standardized requirements specified by the Federal Meteorological Handbook [4]. According to these requirements, an operator during an observation specifies the cloud coverage in “eighths” or “oktas” that are assigned according to the following numeric code: “0” when no clouds are present in the celestial dome (clear sky); “2” when the celestial dome is less than half covered (0 <cloud coverage <4/8) (scattered sky); “7” when the celestial dome is less then 7/8 covered (4/8 <cloud coverage <7/8) (broken sky); “8” when clouds completely cover the celestial dome, except perhaps a small portion (overcast sky).

A study is presented here that considers ten different sites in the United Sates Southwest with locations ranging from California through Texas (Fig. 3-1), and it presents an analysis of cloud coverage records of these sites over a number of years [5]. The United States Southwest is home to a large number of telescopes due to its dry weather, which manifests itself in a limited number of cloudy days with respect to other areas of the United States (and North America in general). Such a large geographical area, however, does not present a uniform weather (and cloud coverage) pattern during the year. For instance, while California experiences dry summers and storms during the winter, Arizona and New Mexico are mainly affected by stormy summer seasons. Therefore, our intent in this study is also to understand the variation

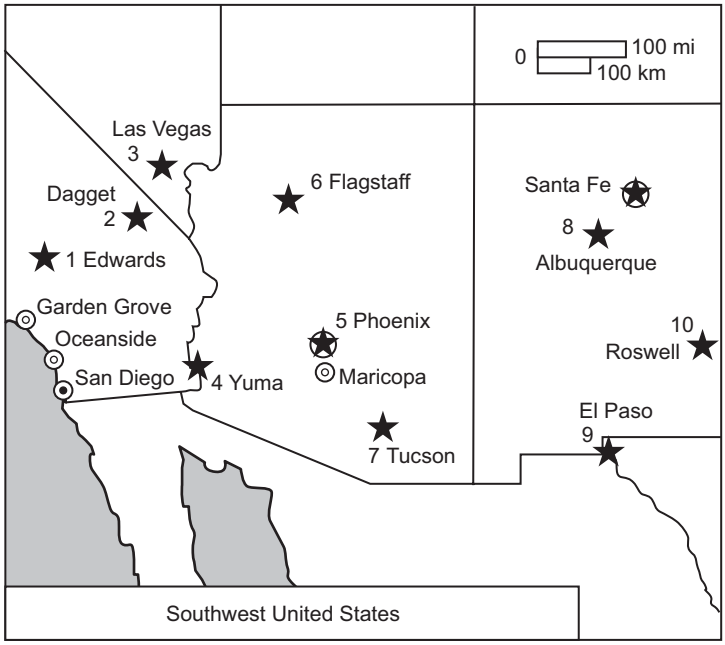


Fig. 3-1. Site Locations (denoted by stars) in the United States Southwest for which cloud coverage statistics have been characterized in this work.

and the correlation of cloud coverage in this area. To accomplish this goal, we have selected ten NCDC observation stations in the region in an area from California (Edwards Air Force Base, 101 km northeast of Los Angeles) to the border between New Mexico and Texas, as indicated in Fig. 3-1. The maximum distance between sites is 1241 km (from Edwards in California to Roswell in New Mexico). There are several reasons for the selection of these particular observation sites. The first one, as already noted, is the intention to cover the relatively dry Southwest. Another reason is to select sites that are near existing telescope facilities. For example, at Table Mountain (California), NASA/JPL deployed a telescope for optical communication; therefore, we selected the NCDC observation station of Edwards Air Force Base, California, which is in the vicinity of Table Mountain Observatory (65 km).

Finally, we selected observation sites near locations or peaks that can be considered for future installation of optical telescopes [5,6]. The selected locations are: (1) Edwards Air Force Base, California; (2) Daggett Airport, California; (3) Las Vegas McCarran International Airport, Nevada; (4) Yuma International Airport, Arizona; (5) Phoenix Sky Harbor International Airport, Arizona; (6) Flagstaff Pulliam Airport, Arizona; (7) Tucson International Airport, Arizona; (8) Albuquerque International Airport, New Mexico; (9) El

Paso International Airport, Texas; (10) Roswell Industrial Airpark, New Mexico.

Most of the NCDC observation centers are located at airports or other locations with relatively modest elevations. By contrast, telescopes are located at higher elevations (usually mountain peaks). Therefore, one may expect (and plan for) different sky visibility conditions between mountain peaks and lower elevation areas in their proximity. For instance, fog and smog do not usually appear at higher elevations because they remain constrained by the inversion layer. Moreover, at higher elevations, mountains may cut off lower clouds, and sometimes the orographic effects of the mountains trap and localize clouds.

3.1.2 Single-Site and Two-Site Diversity Statistics

Initially, we considered 27 years worth of data from NCDC for the ten observation stations (1973–1999). However, in order to have consistent statistics, we considered only years in which the percentage of missing data is at most of the order of one and a half months (13 percent of the year-time amount). Moreover, we considered only years with missing data distributed over the year (i.e., years with an entire month of missing data were disregarded). These principles led us to restrict our study to six years 1991–1993 and 1997–1999 only. Finally, owing to the proximity of Edwards to the Table Mountain Observatory where NASA/JPL has a telescope for optical communications, we emphasize data involving the observation station of Edwards itself [7]. Because of the interest in studying a possible telescope network for deep space optical communications, we present also site diversity data between two sites. The diversity data are obtained by comparing simultaneous cloud coverage between two sites and then by selecting the better sky condition (less cloud coverage).

In Table 3-1 we present and compare the improvement in the average yearly amount of clear sky of two-site diversity over single sites. Figure 3-2 compares the average cumulative distribution of the cloud coverage for single observation site and two-site diversity regarding Edwards and Roswell. Two-site diversity greatly increases the yearly amount of clear sky (66.57 percent), and for cloud coverage less than 4/8 (91.14 percent) with respect to the single site. Another interesting case to consider is two-site diversity between Edwards and Tucson as presented in Fig. 3-3, especially because they are close to existing observatories (i.e., the Goldstone facility (California) and the Mount Lemmon Observatory (Tucson, Arizona)). Even in this case, two-site diversity provides an overall improvement over Tucson and Edwards with clear sky amounts of 61.01 percent and 90.82 percent for cloud coverage less then four eighths.

Table 3-1. Clear sky percentage averaged over six years for single-site (diagonal bold) and two-site diversity.

	Edwards	Daggett	Las Vegas	Yuma	Phoenix	Flagstaff	Tucson	Albuquerque	El Paso	Roswell
Edwards	**26.8**	45.5	51.8	41.3	48.3	50.4	61.0	49.1	58.7	66.5
Daggett	45.5	**41.4**	56.6	50.0	55.5	58.9	68.7	58.4	66.6	74.9
Las Vegas	51.8	56.6	**46.5**	54.6	57.9	59.6	68.0	58.4	67.6	75.0
Yuma	41.3	50.0	54.6	**28.8**	44.5	52.8	57.6	48.2	57.1	66.3
Phoenix	48.3	55.5	57.9	44.5	**36.9**	55.6	57.1	50.1	58.6	68.2
Flagstaff	50.4	58.9	59.6	52.8	55.6	**40.6**	63.4	55.9	65.3	68.8
Tucson	61.0	68.7	68.0	57.6	57.1	63.4	**49.8**	58.1	62.2	68.4
Albuquerque	49.1	58.4	58.4	48.2	50.1	55.9	58.1	**33.3**	54.1	63.1
El Paso	58.7	66.6	67.6	57.1	58.6	65.3	62.2	54.1	**44.7**	65.7
Roswell	66.5	74.9	75.0	66.3	68.2	68.8	68.4	63.1	65.7	**55.0**

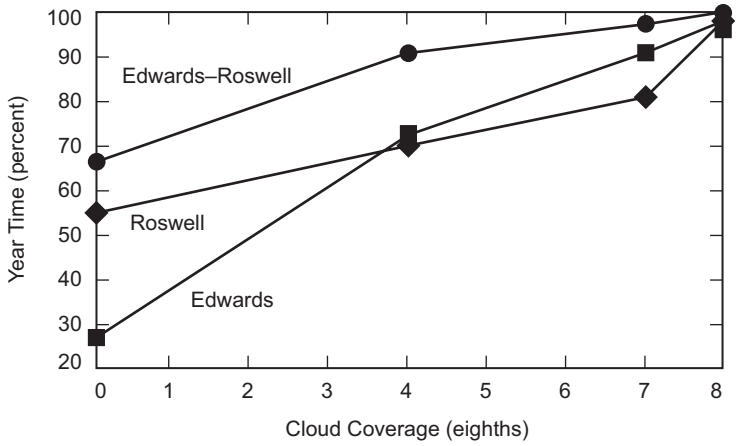


Fig. 3-2. Comparison of cumulative average cloud coverage among one-site Roswell, one-site Edwards, and two-site diversity for combined Roswell and Edwards for the years 1991–1993 and 1997–1999. Since NASA/JPL installed an optical telescope at Table Mountain in the proximity of Edwards, Roswell is a potential location in whose proximity a telescope will enhance the two-site diversity with Table Mountain.

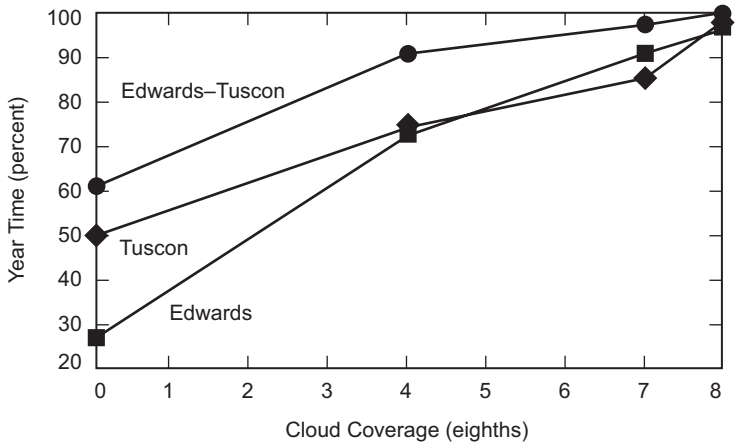


Fig. 3-3. Comparison of cumulative average cloud coverage among one-site Tucson, one-site Edwards, and two-site diversity for combined Tucson and Edwards for the years 1991–1993 and 1997–1999.

Because Edwards has the average lowest clear sky amount among the ten locations, one should expect that other locations could offer higher yields of clear sky amount for two-site diversity.

In fact, best results in two-site diversity involve Roswell and Daggett (see Fig. 3-4), or Roswell and Las Vegas (see Fig. 3-5). In these two cases, the clear sky amount is about 75 percent.

Cloud coverage data from each observation station present monthly variations that greatly differ during the year depending on the geographical

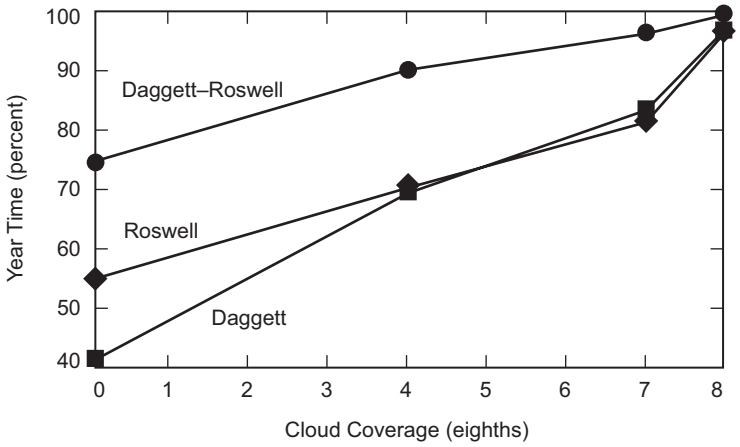


Fig. 3-4. Comparison of cumulative average cloud coverage among one-site Daggett, one-site Roswell, and two-site diversity for Daggett and Roswell for the years 1991–1993 and 1997–1999.

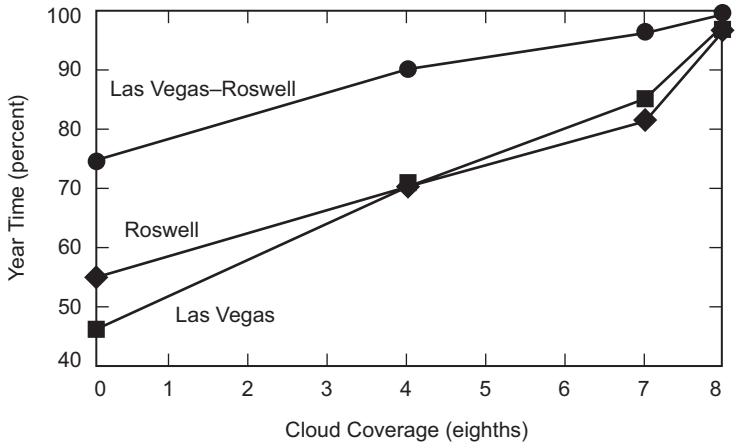


Fig. 3-5. Comparison of cumulative average cloud coverage among one-site Las Vegas, one-site Roswell, and two-site diversity for combined Las Vegas and Roswell for the years 1991–1993 and 1997–1999.

area. For instance, while the clear sky amount was smaller during the winter in Southern California, the same was not true in New Mexico and part of Arizona, where clear sky amount was reduced during the summer months of July and August. Therefore, upon selection of proper locations, one should expect that in two-site diversity statistics the monthly variation of clear sky (and other cloud coverage conditions) would be more uniform over the year. To better prove this last statement, in this segment we discuss monthly variations of two-site diversity statistics involving Edwards, Daggett, Las Vegas, Tucson, and Roswell.

Figure 3-6 presents average monthly variations of cloud coverage between Edwards and Tucson. The differences between Edwards and Tucson are evident in Figs. 3-6(a) and 3-6(c). Consequences of two-site diversity are shown in Fig. 3-6(e), where during the months of winter, spring, and autumn the dominant clear-sky contribution to the statistics is given by Tucson, while during the summer Edwards compensates for the lack of clear sky at Tucson. A monthly variation is still visible in the two-site statistics for the clear sky, but overall we can observe an average of 15 days of clear sky all year round. Two-site diversity between Edwards and Roswell (Fig. 3-7) presents more favorable conditions. In fact, the clear sky amount for the combination of these sites does not change much over the year. Except for January and February, the average clear sky amount for the two-site diversity is approximately 20 days. As previously noted, Daggett and Roswell (Fig. 3-8) along with Las Vegas and Roswell (Fig. 3-9) are the most advantageous choices for two-site diversity. Las Vegas–Roswell two-site diversity presents minimal variation of clear sky amount over the year, with an average of 2/3 clear sky for each month.

3.1.3 Three-Site Diversity

Three-site diversity may offer further improvement of clear sky (and clear plus scattered sky) over two-site diversity. Figure 3-10 presents some results of three-site diversity involving Edwards during the years 1991–1993 and 1997–1999. Among the curves in Fig. 3-10, there is the cloud-coverage cumulative distribution curve that describes a case study with Edwards–Daggett–Tucson. For this configuration, the average clear sky amount is 70.6 percent during the year. However, the fact that Daggett is in the proximity of Edwards and that they both belong to the same climate area does not constitute a good choice for three-site diversity. As a result, for this site selection example, the benefits of having three stations operating simultaneously are greatly reduced. In fact, one may notice that during the same period of time, the clear sky amount for two-site diversity of Tucson and Dagget is 68.79 percent, which suggests that the addition of Edwards to the other two stations provides only marginal improvement of the overall statistics. Adding Roswell to Edwards

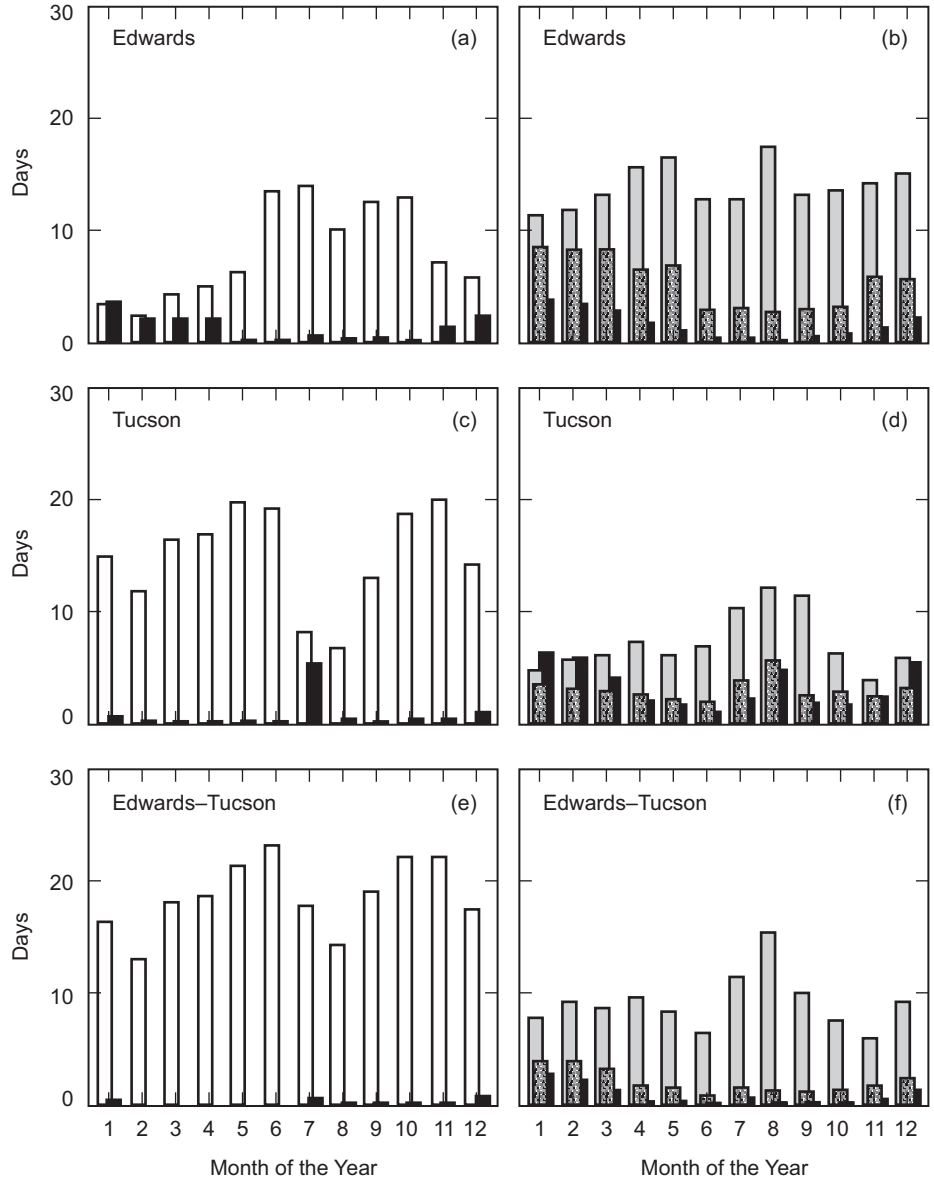


Fig. 3-6. Average monthly cloud coverage: comparison of cumulative average cloud coverage among one-site Edwards, one-site Tucson, and two-site diversity for combined Edwards and Tucson. February is averaged over 28 days: (a), (c), and (e) for each month indicate in sequence the equivalent days of clear sky (white bar) and missing data (black bar); (b), (d), and (f) for each month indicate in sequence the equivalent days of scattered (gray bar), broken (speckled bar), and overcast sky (black bar). The years studied are 1991–1993 and 1997–1999.

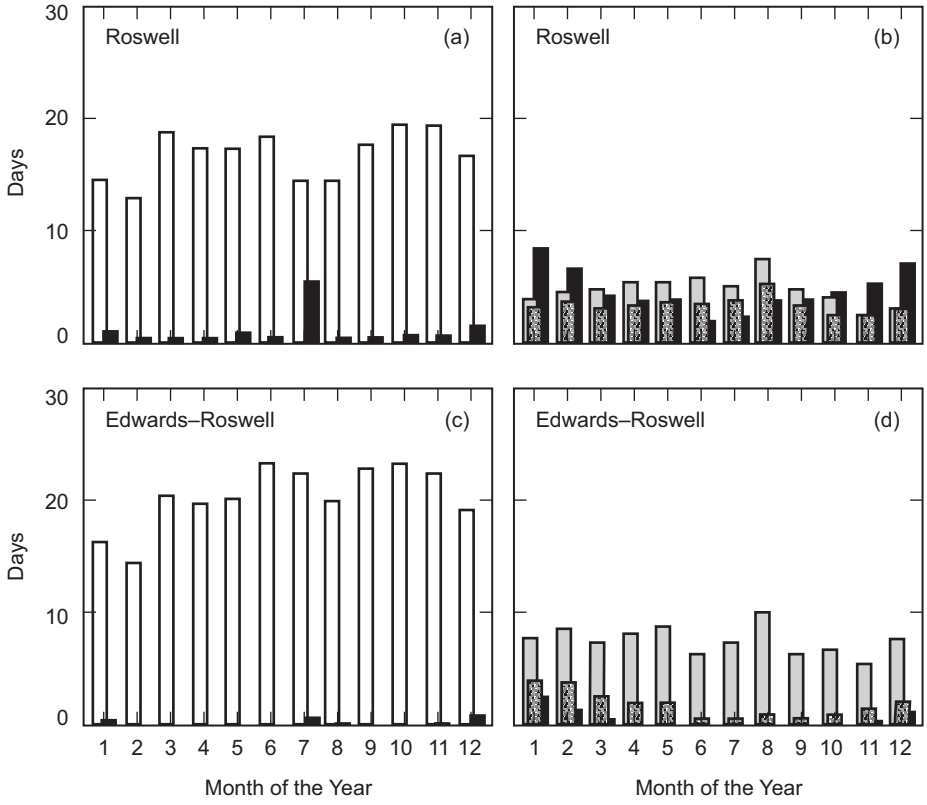


Fig. 3-7. Average monthly cloud coverage: comparison among one-site Edwards, one-site Roswell, and two-site diversity for combined Edwards and Roswell. February is averaged over 28 days: (a) and (c) for each month indicate in sequence the equivalent days of clear sky (white bar) and missing data (black bar); (b) and (d) for each month indicate in sequence the equivalent days of scattered (gray bar), broken (speckled bar), and overcast sky (black bar). The years studied are 1991–1993 and 1997–1999.

and Tucson improves the two-site diversity performances as seen in Fig. 3-3. For this last configuration, the average clear sky amount is 75.33 percent during a year with 19.52 percent scattered sky. However, if Edwards must be considered for three-site diversity, adding Roswell and Las Vegas gives the best contribution to enhance the clear sky statistics, with 77.54 percent.

Among the results analyzed here, the combination Daggett–Tucson–Roswell exhibited the best performance when considered for three-site diversity, producing a yearly clear sky amount of 81.24 percent with a scattered-sky amount of 13.34 percent.

Incidentally, one may notice in the Fig. 3-1 map that in this last configuration Tucson is symmetrically distant from the other two locations

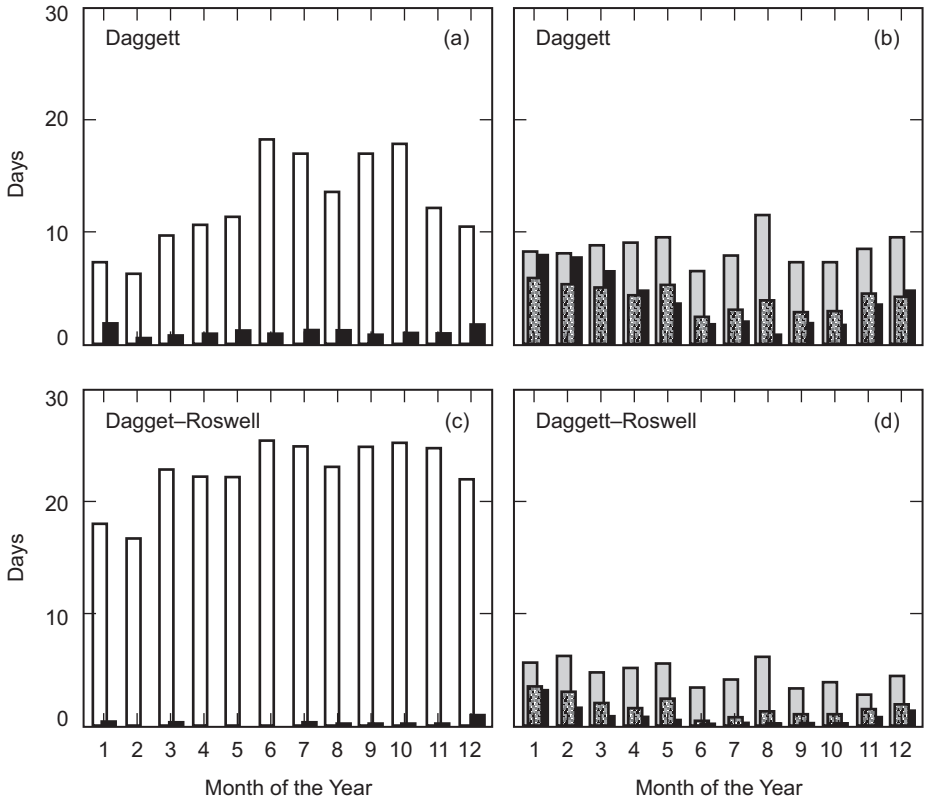


Fig. 3-8. Average monthly cloud coverage: comparison among one-site Daggett, one-site Roswell, and two-site diversity for combined Daggett and Roswell. February is averaged over 28 days: (a) and (c) for each month indicate in sequence the equivalent days of clear sky (white bar) and missing data (black bar); (b) and (d) for each month indicate in sequence the equivalent days of scattered (gray bar), broken (speckled bar), and overcast sky (black bar). The years studied are 1991–1993 and 1997–1999.

(each branch is of the order of 600 km). Moreover, all three locations were representative of three distinct climatic zones.

In this section we present results of monthly variations of cloud coverage for the same examples of three-site diversity analyzed in Fig. 3-10. One should hope, after selecting the proper locations for three-site diversity, to achieve extended durations of clear sky, with minimal monthly variation. For Edwards–Daggett–Tucson, a yearly variation of monthly amount for clear sky is still detectable, with relatively minimal amounts in January, February, and August, as seen in Fig. 3-11(a). During these months, the clear sky amount was in the range of 15–20 days. During the other months of the year it exceeded 20 days. For Edwards–Tucson–Roswell, the clear sky amount exceeded the 20 days with

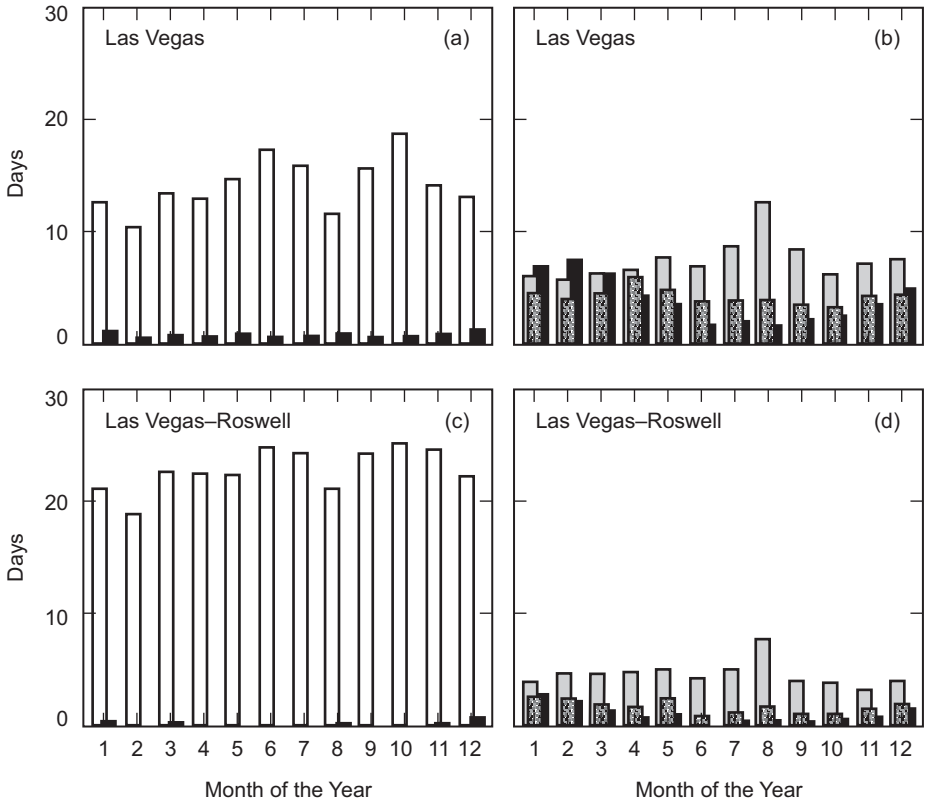


Fig. 3-9. Average monthly cloud coverage: comparison among one-site Las Vegas, one-site Roswell, and two-site diversity for combined Las Vegas and Roswell. February is averaged over 28 days: (a) and (c) for each month indicate in sequence the equivalent days of clear sky (white bar) and missing data (black bar); (b) and (d) for each month indicate in sequence the equivalent days of scattered (gray bar), broken (speckled bar), and overcast sky (black bar). The years studied are 1991–1993 and 1997–1999.

the exception of February (17 days). A reduction of the clear sky amount is detectable during August and the winter months, Fig. 3-11(b).

Monthly variation of the clear sky amount was less accentuated for Edwards–Las Vegas–Roswell, as seen in Fig. 3-11(c). In this configuration, February had only 19.11 days of clear sky amount (which, however, represented over 68 percent or the time during the month of 28 days), with the other months well beyond 20 days presenting a peak of 25.57 days in July.

The combination Daggett–Tucson–Roswell shows a similar trend of less variation, with an evident incremental increase of clear sky during the months of March, April, and May, Fig. 3-11(d).

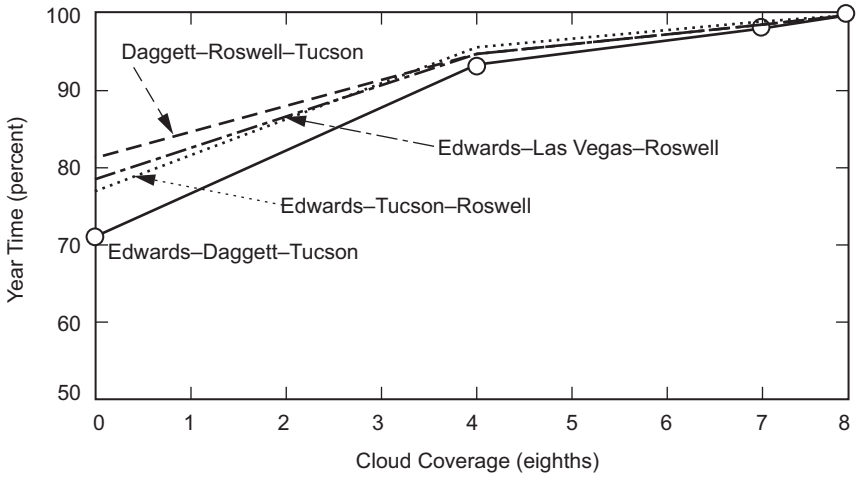


Fig. 3-10. Three-site diversity average cloud coverage statistics for Edwards–Daggett–Tucson (solid curve), Edwards–Tucson–Roswell (dotted curve), Edwards–Las Vegas–Roswell (dash-dotted curve), and Daggett–Roswell–Tucson (dashed curve). The years studied are 1991–1993 and 1997–1999. The piece-wise cumulative distribution curves are shown.

3.1.4 NCDC Analysis Conclusion

Using the surface observations, we were able to calculate a statistical representation of cloud coverage for single site, two-site diversity, and three-site diversity for the Southwest sites studied. Two-site diversity statistics clearly showed improvement over the single observation site statistics. By selecting a proper pair of sites among the 45 available combinations, we also demonstrated that two-site diversity statistics presented favorable periods of clear sky that were more uniform over the months when compared to single-site statistics. For instance, Las Vegas–Roswell, during the years 1991–1993 and 1997–1999, presented clear sky amount of about 75 percent as compared to the single site amount of 46.51 percent for Las Vegas and 55.03 percent for Roswell, Table 3-1.

Besides the overall improvement of the sky visibility, further analysis of data, also has shown that site diversity can be a robust solution against anomalies in the climate patterns that may affect the performance of a single telescope. To better explain this last concept, one may consider the hypothetical case of two telescopes: one located in the proximity of Edwards (e.g., Table Mountain) and the other in the proximity of Tucson (e.g., Mount Lemmon) during the year 1997. Table 3-1 indicates that on average one should expect a yearly amount 28.83 percent of clear sky at Edwards and 49.85 percent in Tucson. However, under the influence of “El Niño,” the cloud coverage in both locations greatly differed from the average that year. In fact, during 1997 at

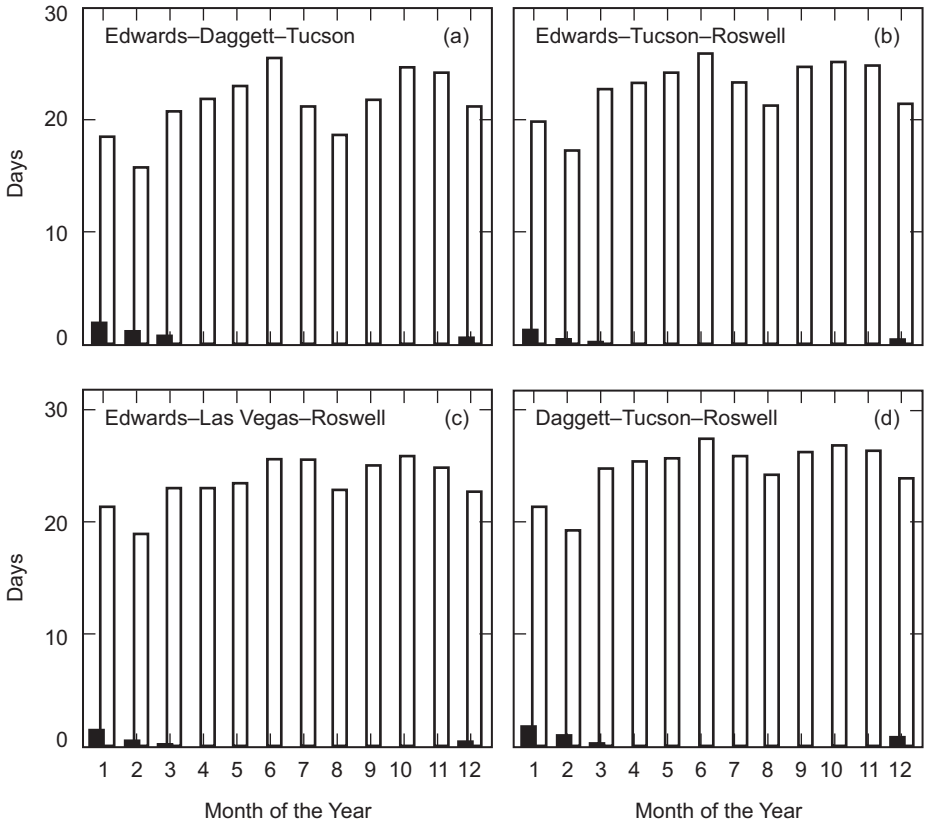


Fig. 3-11. Three-site diversity: average monthly clear sky for (a) Edwards–Daggett–Tucson, (b) Edwards–Tucson–Roswell, (c) Edwards–Las Vegas–Roswell, and (d) Daggett–Tucson–Roswell. February is averaged over 28 days. Bars indicate for each month in sequence the equivalent days of clear sky (white bar) and missing data (black bar) [5,6].

Edwards, only 19.39 percent of the time was the sky clear, while at Tucson the clear-sky amount was 61.36 percent, and the overall diversity clear sky was 66.9 percent. Therefore, the unusual climate pattern caused by El Niño affected the two locations in opposite ways; while the visibility for a telescope in the proximity of Edwards was greatly reduced, in Tucson the visibility was enhanced.

A further improvement with respect to two-site diversity is given by three-site diversity. For instance, statistics show that by adding Tucson (with 61.36 percent of yearly clear sky) to Las Vegas and Roswell, the clear sky amount is almost 80 percent compared to 75 percent for two-site diversity. The best overall results were observed for the triplet Daggett–Tucson–Roswell with 81.28 percent clear sky.

As a result of this investigation, one may notice that the average yearly clear sky amount improves on the order of tens of percent for two-site diversity over a single site for two properly selected locations. However, the additional improvement for three-site diversity compared with two-site diversity may not be as dramatic. System engineers should carefully evaluate the importance of a few percentage numbers in considering whether they can justify the expenses that the use of a third telescope would entail in order to reach a very high availability of the atmospheric channel at optical wavelengths.

Finally, among the sites characterized here, it was not possible to observe an amount of clear sky close to 100 percent in any (single, two-site diversity, or event three-site diversity) configuration. This result suggests that the study presented here based upon NCDC surface observation data be further expanded to other areas of the United States to find if this 100 percent limit of yearly clear sky is achievable by site diversity.

The approach in studying Earth cloud coverage using surface observation data, as proposed in the previous section, has a number of inherent advantages. The NCDC data bank has an extensive historical archive (decades). The data are readily available from the NCDC, and the observation stations are distributed all over the world (with a great emphasis on the continental United States). However a number of drawbacks are evident too. It is not possible to pinpoint a specific site on Earth, except by serendipitous coincidence with one of the observation stations. Observation stations are usually located at airports and not on mountain peaks. Moreover, the surface-observed data are prone to subjectivity of the observer and are at best qualitative. Finally, the number of stations is greatly reduced in the Southern Hemisphere.

3.1.5 Cloud Coverage Statistics by Satellite Data Observation

Satellite imaging measurements offer a more quantitative complement to surface observations. The satellite images acquired at different wavelength bands (both in the visible and in the infrared) are processed to extract spatial distribution of water vapor concentration in the atmosphere and therefore on the presence of clouds. The measurements are broadcast to Earth and distributed. The spatial distribution of clouds (water vapor) is determined using algorithms that perform a series of threshold tests. Over time, it is possible to provide cloud coverage statistics over the region that is imaged by the satellite. The spatio-temporal resolution of this approach is determined by the pixel angular FOV and the frequency at which images are acquired.

The remainder of this section, summarizes results of a study performed by TASC Inc. [3,8,9] on cloud-coverage statistics based on images collected by GOES9, METEOSAT-5, and METEOSAT-7 satellites. The spatial resolution is 5 km, and the temporal resolution is one hour. The averages span from

71.3 percent for Goldstone (California) to 55.29 percent at Starfire Optical Range (New Mexico).

The sites targeted in this study have astronomical observatories or NASA facilities that may be dedicated to space-to-ground communication. The sites chosen were concentrated in the United States Southwest, namely: Goldstone, Mount Wilson, Mount Palomar, and Table Mountain in California; Kitt Peak, in Arizona; Starfire Optical Range, in New Mexico (SOR); McDonald Observatory in Texas; and Mount Haleakala and Mauna Kea in Hawaii. The site altitudes vary from 1 to 4 km with the lowest being at Goldstone and the highest at Mauna Kea. Figure 3-12 shows the average yearly percentage of clear sky for the study sites.

The satellite study emphasized site-diversity statistics to determine link availability for a network of four sites among those considered. In other words, given the four sites distributed among the selected locations, what would the link availability be to any one of these four sites. Particularly, the case of a network of four telescopes supporting a deep space mission from Mars was considered. The determination of the cloud-free-line-of sight (CFLOS) took into account the angular variations on the plane of the ecliptic of Mars and Earth during the mission. It was found that a network of telescopes located in Goldstone, Kitt Peak, McDonald Observatory, and Mauna Kea yielded the best results in terms of availability. This is depicted graphically, based on data acquired from 1997–2002, in Fig. 3-13(a). Apparent in these data is a general evidence of seasonal variability of CFLOS between a spacecraft in Mars orbit and the selected ground stations. For all years with the exception of 1998,

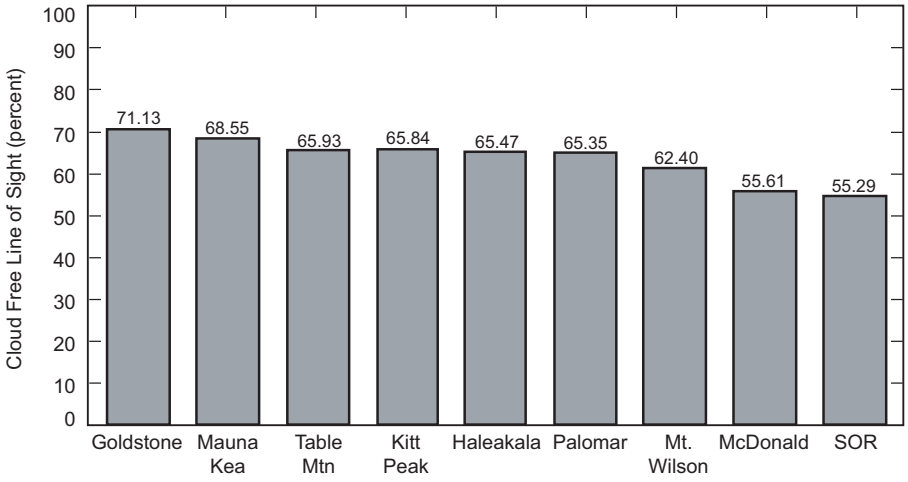


Fig. 3-12. Average yearly percentage of clear sky for the sites considered by the study. (SOR is an acronym for Starfire Optical Range facility.) The data are representative of six years of observations (1997–2002).

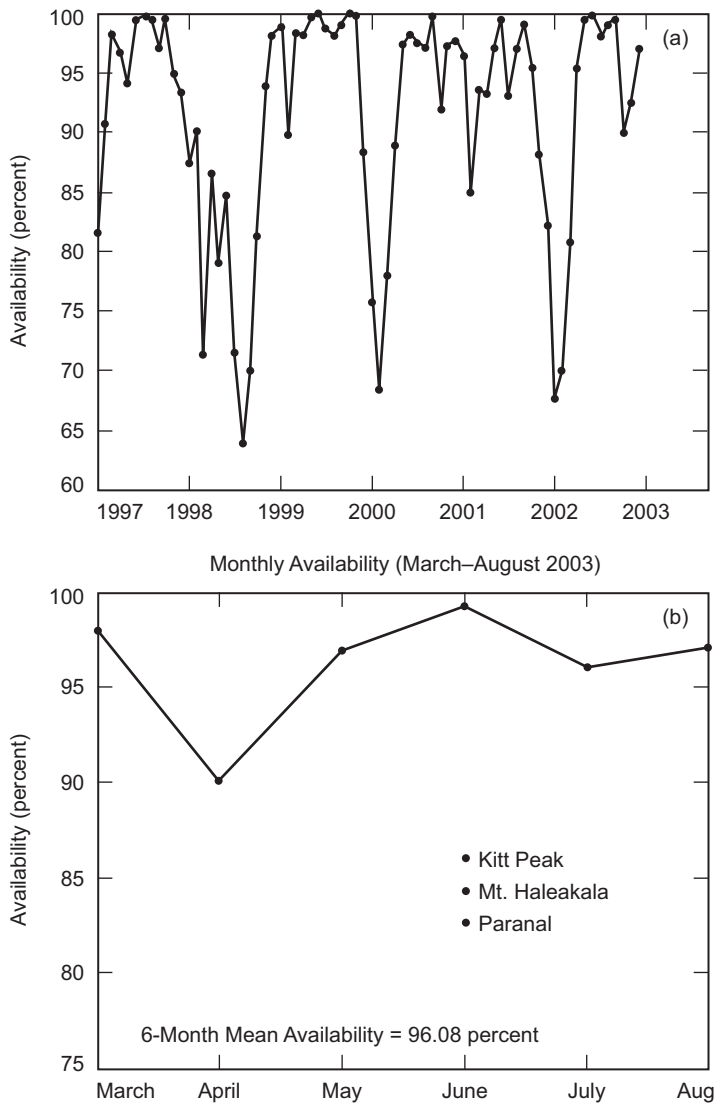


Fig. 3-13. (a) Average yearly percentage of clear sky for a network of four sites in spatial diversity. This combination of sites, Goldstone, Kitt Peak, McDonald Observatory, and Mauna Kea, gives the best results of CLFOS. The data are representative of six years of satellite observations (1997–2002). (b) 6-month statistics including a Southern-Hemisphere site (Paranal, Chile) shows the possibility of enhancing overall availability in a ground-based optical network.

winter and early spring availabilities are 70–80 percent increasing to 95–100 percent during the remainder of the year. The exception is 1998 because an El Niño condition occurred that year, which does not conform to this trend, showing nearly year-round relatively poor availability of 65–90 percent. Note that only Northern Hemisphere sites were considered in the data set presented. The overall network availability can be greatly enhanced by including Southern Hemisphere sites. The benefits would be twofold, namely, summer skies with better availability for the Southern Hemisphere sites as well as possible longer passes of the spacecraft including higher elevation angles. The long term statistical data to prove the better availability had not been analyzed at the time of writing this text. However, a proof of concept analysis based on a six-month period of cloud cover analysis using three sites (namely, Kitt Peak, Arizona; Mt. Haleakala, Hawaii; and Paranal in northern Chile) yield an overall 96 percent availability, as shown in Fig. 3-13b.

3.2 Atmospheric Transmittance and Sky Radiance

3.2.1 Atmospheric Transmittance

To quantify laser beam attenuation experienced by an optical communications link, one can introduce the concept of optical depth τ. The power reaching the receiver P_r is related to the transmitted power P_t [10] as

$$P_r = P_t e^{-\tau} \tag{3.2-1}$$

The fraction of the power transmitted in the optical link, T, is called transmittance and is given by

$$T = \frac{P_r}{P_t} = e^{-\tau} \tag{3.2-2}$$

The atmospheric transmittance and the optical depth are related to the atmospheric attenuation coefficient γ and the transmission path length r by

$$T = e^{-\int_0^r \gamma(\rho) d\rho} \tag{3.2-3}$$

and

$$\tau = \int_0^r \gamma(\rho) d\rho \tag{3.2-4}$$

where the atmospheric attenuation coefficient is laser-wavelength specific and depends on the path-integrated distribution of atmospheric constituents along

the line-of-sight. From the above equation, one can define the loss L in decibels that the beam experiences as

$$L = -10\log_{10} T = 4.34\tau \qquad (3.2\text{-}5).$$

An optical depth of 0.7, therefore, corresponds to a loss of 3 dB, or 50 percent of the signal power.

Generally speaking, the atmospheric attenuation coefficient can be expressed as the combination of absorption and scattering of the light beam due to gas molecules and aerosols present in the atmosphere:

$$\gamma = \alpha_m + \beta_m + \alpha_a + \beta_a \qquad (3.2\text{-}6)$$

where α_m and α_a are respectively the absorption coefficient for the molecular gas and aerosol, and β_m and β_a are the scattering coefficient for the molecular gas and aerosol. One should notice that although both absorption and scattering contribute to the attenuation coefficient, their attenuation mechanism is quite different. When a light beam propagates through the Earth atmosphere, it may interact with the gases and be absorbed. A light photon is absorbed when the quantum state of a molecule is excited to a higher state of energy. Specific states of energy absorb light at specific wavelengths with narrow line widths. The absorbed energy may be released at different wavelengths or as heat. During elastic scattering, there is no loss of energy at that determined wavelength. Instead the scattered light is redirected (at the same wavelength) into the total solid angle, with an effective loss along the propagation direction of the light beam. Scattering not only degrades the propagation of a signal beam in the atmosphere, but it is also the origin of the background sky radiance that introduces noise in a day-time communications downlink operation (sky radiance is discussed later in this chapter). Generally, scattering can be classified according to the size of the scattering particle and the wavelength: if the scattering particle is smaller than the wavelength the process is termed Rayleigh scattering [11]; if the size of the scatterer is comparable to the wavelength, it is termed Mie scattering [12]. Commonly, one may observe that molecular scattering is due to Rayleigh scattering while aerosol scattering is better described by Mie theory. When the size of the scatterers is much larger than the wavelength in consideration, diffraction theory can describe the scattering of light in a more proper fashion.

3.2.2 Molecular Absorption and Scattering

The Earth atmosphere is a combination of different gases [13]. The main constituents of the atmosphere are nitrogen (N_2) with 78.09 percent and oxygen (O_2) with 20.95 percent. Tables 3-2 and 3-3 summarize some data regarding gaseous composition of Earth troposphere.

Table 3-2. Main gases composing the Earth atmosphere.

Constituent	Volume Ratio (%)	Parts Per Million (ppm)
Nitrogen (N_2)	78.09	
Oxygen (O_2)	20.95	
Argon (Ar)	0.93	
Carbon dioxide (CO_2)	0.03	
Water vapor (H_2O)		40–40,000
Neon (Ne)		20
Helium (He)		5.2
Methane (CH_4)		1.5
Krypton (Kr)		1.1
Hydrogen (H_2)		1
Nitrous oxide (N_2O)		0.6
Carbon monoxide (CO)		0.2
Ozone (O_3)		0.05
Xenon (Xe)		0.09

Table 3-3. Earth atmospheric composition at sea level (10^{13} mb pressure and 300 K temperature).

Components	Mixing Ratio	Fraction in Mass	Mass Density (g/m^3)	Number Density (cm^{-3})
N_2	78.09%	76.50%	986.9	2.1×10^{19}
O_2	20.955	21.97%	283.7	5.7×10^{18}
Ar	0.93%	1.30%	16.8	2.6×10^{17}
H_2O	~1%	~0.63%	~8.1	$\sim 2.7 \times 10^{17}$
CO_2	400 ppm	615 ppm	0.8	1.1×10^{16}
CO	0.2 ppm	0.2 ppm	2.6×10^{-4}	5.6×10^{12}

Of course, the composition of the Earth atmosphere can vary with location on Earth and altitude over sea level. For instance water vapor density is larger in tropical areas and quite sparse in desert regions. In proximity to industrial areas and urban centers, the concentration of densities of carbon dioxide, carbon monoxide, ozone, and pollutants are relatively high. Finally, the density of Earth's atmosphere decreases with increasing altitude, as represented in

Fig. 3-14, with a resulting less severe interaction between a propagating optical beam and the atmospheric channel.

As stated earlier, photons (energy) absorbed by atmospheric molecules change the rotational, vibrational, electronic, or all of these energy states for the molecules. These molecular energy states are quantized and selected (bands of) wavelengths that are absorbed. Resulting from this quantized molecular absorption is a peculiar transmittance spectrum of the atmosphere with "forbidden bands" where the absorption loss is close to 100 percent. Figure 3-15 shows the atmospheric transmittance spectrum from which one may choose wavelengths of interest for optical space communication in the range 0.5–2 μm. Figure 3-15 describes the transmittance of the atmosphere for an observer (or a link) viewing local zenith (900 of elevation). The graph was obtained using MODTRAN 4.0 [15]. MODTRAN simulation results are used in this chapter both to obtain a measure of the atmospheric transmittance and background sky radiance under different condition of ground station altitude, aerosols distribution, meteorological visibility, etc. In particular, Fig. 3-15 simulates the transmittance spectrum experienced by a space-to-ground link with a telescope/transmitter at sea level looking at the zenith, at a mid-latitude location on Earth, with no aerosol concentration. Therefore, the spectrum described is generated only by molecular absorption and scattering (the action of aerosols is described in the next subsection).

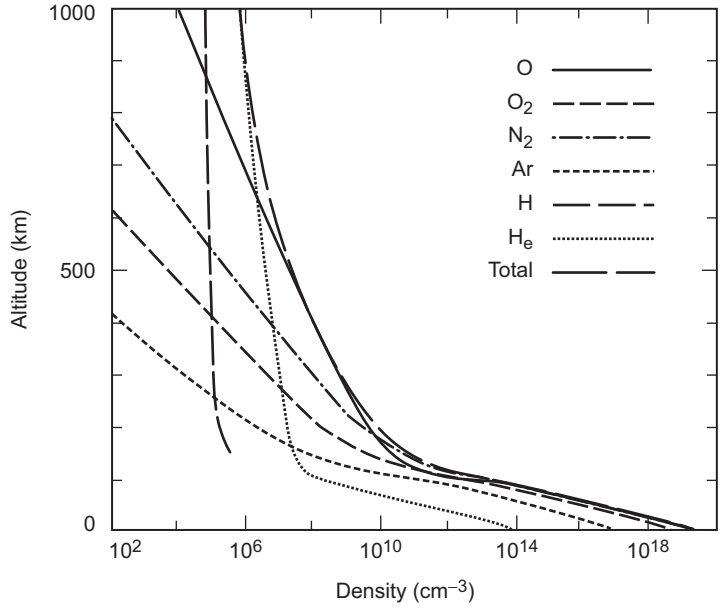


Fig. 3-14. Earth atmospheric number density profiles for individual species.

One may notice that Fig. 3-15 shows that a number of forbidden bands for laser propagation, mainly these around 0.7, 0.8, 0.96, 1.1, 1.38, and 1.9 μm. These forbidden bands are associated mainly with water vapor interactions [15] while oxygen and carbon dioxide have relatively weaker absorption lines in this range. By contrast, attenuation in the visible region of the spectrum is due to Rayleigh scattering [16]. The absorption coefficient of the Rayleigh scattering shows a functional dependence with the wavelength λ as λ^{-4}, clearly indicating a larger attenuation in the blue range (~0.425 μm) of the visible spectrum, a much smaller attenuation in red (~ 0.600 μm), and practically negligible attenuation for wavelengths in the infrared. Incidentally, this functional dependence of Rayleigh scattering wavelength dependence manifests itself with the scattering of the sunlight and the corresponding blue color of the sky.

Finally, the zenith transmittance spectrum of Fig 3-15 can be conveniently scaled at different observation angle. If we define as T_0 the transmittance at the zenith, one can easily calculate the transmittance T_θ at observation zenith angle θ as

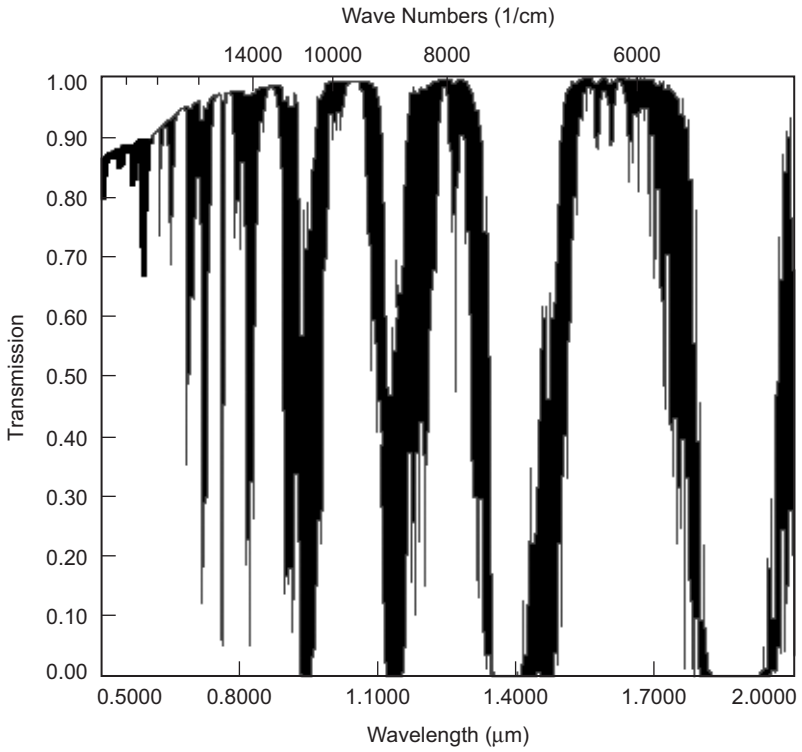


Fig. 3-15. Transmittance spectrum at sea level with zenith angle of zero. The plot is representative of a mid-latitude site on Earth with a hypothetical absence of aerosols.

$$T_\theta = T_0^{\sec(\theta)} \tag{3.2-7}$$

with the above relationship valid up to $\theta = 70°$ without loss of accuracy. The term "sec(θ)" is also referred to as airmass.

3.2.3 Aerosol Absorption and Scattering

Aerosols are atmospheric particles spanning a wide variety of constituents including dust, organic material, pollutants, ice, water droplets etc. Aerosol sizes can vary from sub-micrometer to a few tens of micrometers, and so its shapes (from spherical to irregular). Aerosols differ in distribution, components, and profile concentration in the atmosphere [17]; consequently, they influence the interactions with propagating laser beams in different ways (i.e., absorption and scattering). The largest concentration and variability of aerosols can be located in a boundary layer up to 1–2 km immediately above the Earth surface. Aerosols in the boundary layer are generally classified as maritime, rural, urban, and desert model [10]. Maritime aerosols are in the proximity or over the sea and ocean surfaces, and they typically consist of salt particles in aggregation with water droplets. Over land (distant from industrial settings), the rural model usually describes aerosol composition. Generally, aerosol composition of the rural model consists of dust particles and other substances (sulfates, organic materials originating from local flora) mixed with water droplets. Again, composition, density, and particle size distribution of aerosols belonging to the rural model vary with vegetation, land composition, weather dynamics, and seasonal climate variations. Manmade aerosols, those produced by industrial sources and other typical byproducts of urbanization (combustion, pollution, etc.), contribute to the urban model of the aerosol profile in the boundary layer. In the desert model, aerosols are mainly airborne dust particles, and their concentration mainly depends on wind speed. Above the boundary layer, aerosol concentration decreases in an exponential fashion until air convection and other atmospheric mixing mechanisms cause a globally uniform distribution of aerosols (independent of the sources in the boundary layer). With further increase in altitude, a region in the stratosphere between 10 and 20 km is reached, where particulate matter is essentially composed of sulfates and other products related to photochemical reactions with particles injected into the troposphere during volcanic eruptions [18,19]. This volcanic dust concentration is generally constant in time; however, the composition can change during episodes of volcanic eruption with further injection of volcanic dust that is usually dispersed in a few months. Above 30 km altitudes, aerosols are composed by meteoric and cometary dust.

The aerosol refraction index is described by a real and imaginary part (related to the material conductivity) as:

$$n_a = n_r - jn_i \tag{3.2-8}$$

aerosol absorption coefficient is related to the imaginary part of the refraction index as

$$\alpha_a = \frac{2\pi}{\lambda} n_i \tag{3.2-9}$$

A detailed list of the complex refraction index for a number of aerosol species can be found in [10].

Mie scattering theory describes scattering by aerosols [12]. Generally, according to the Mie theory, the scattering coefficient of aerosols depends on the particle concentration, size distribution, cross-section, and radiation wavelength. Detailed discussions of Mie scattering theory are beyond the scope of this work. However, to simply describe the action of aerosol Mie in the atmosphere, it is convenient to use the following practical relationships commonly used to describe the scattering coefficient, β_a, in horizontal path with constant aerosol concentration:

$$\beta_a = C_1 \lambda^{-\delta} \tag{3.2-10}$$

where C_1 and δ are constants [10,19] determined by aerosol characteristics (density, particle size distribution) and λ is the wavelength of interest in micrometers. The constant δ is related to the atmospheric visibility and varies from 1.0–1.6 (from poor visibility to clear line-of-sight). The constant C_1 is related to the visual (or meteorological) range V (in kilometers) as

$$C_1 = \frac{3.91}{V}(0.55)^{\delta} \tag{3.2-11}$$

where the visual range is referred to 0.550 μm [20]. The scattering coefficient therefore becomes

$$\beta_a = \frac{3.91}{V}\left(\frac{\lambda}{0.55}\right)^{-\delta} \tag{3.2-12}$$

where the wavelength is indicated in micrometers. Typical values of visual range are 5 km for hazy sky (high concentration of aerosol) and 23 km for clear sky. It is interesting to note, that despite an order of magnitude lower aerosol concentration compared to molecular gas concentration, aerosol scattering dominates Rayleigh scattering in the region of interest for optical communications wavelengths (0.5–2 μm).

Practical examples of the atmospheric transmittance dependence on aerosol concentration for a ground-to-space zenith pointing optical path are shown in Figs. 3-16 to 3-21.

Several interesting observations can be derived from these figures. For simplicity, all of these MODTRAN-generated plots are restricted to the case of rural aerosol model and location at mid-latitudes in the Northern Hemisphere during summer.

Figures 3-16 to 3-18 apply to 23-km visual range at different altitudes starting with sea level in Fig. 3-16. It is already evident, by comparing Fig. 3-16 to Fig. 3-15 describing the case of "aerosol free" atmosphere, that the atmospheric transmittance is reduced even with relatively benign high-visibility aerosol distribution. Of course, to mitigate the effects of aerosol concentration, telescope sites are usually located at higher elevation, so that the impact of the aerosols in the boundary layer is greatly reduced and the channel transmittance increased. Considering, for instance, the wavelength of 1 μm, Figs. 3-17, and 3-18, show the improvement in zenith atmospheric transmittance of 0.85 at sea-level to 0.93 at 2 km and 0.96 at 3 km. Note that at a 3-km altitude the

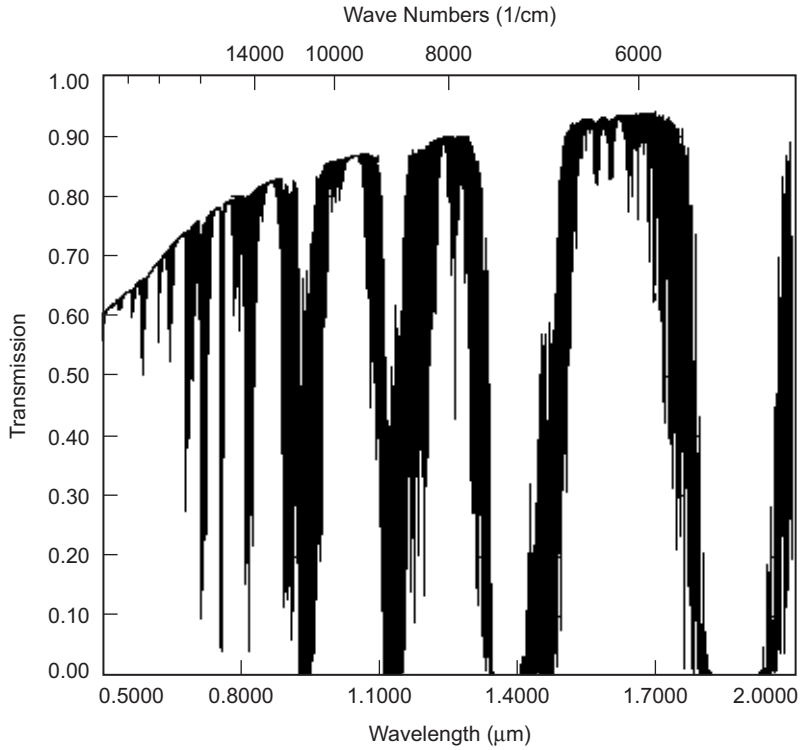


Fig. 3-16. Transmittance spectrum of the atmosphere for a space-to-ground link at sea level with zenith angle 0 deg, mid-latitude, 23 km of visual range.

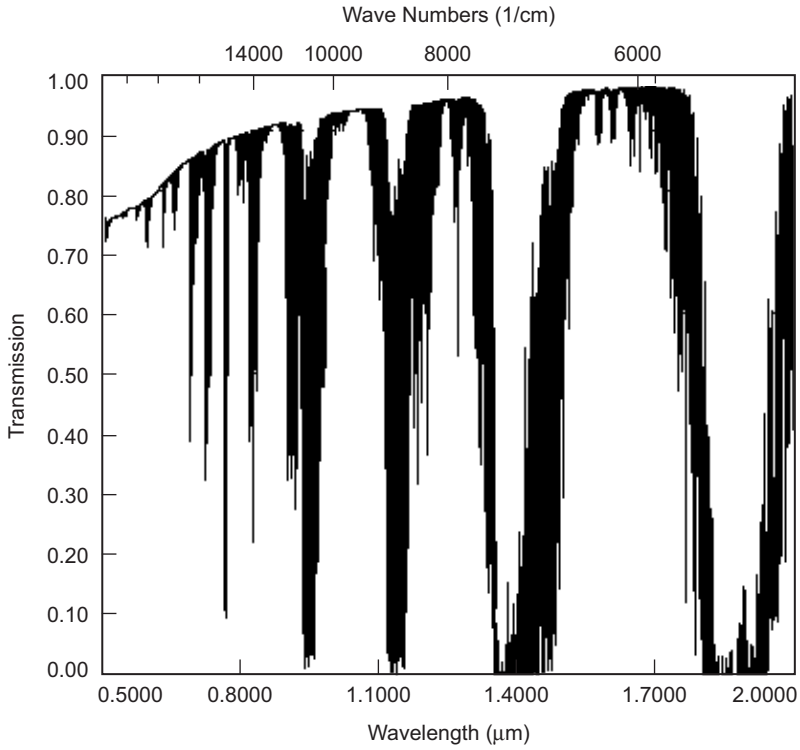


Fig. 3-17. Transmittance spectrum of the atmosphere for a space-to-ground link at 2 km of altitude with zenith angle 0 deg, mid-latitude. The boundary layer starts at 1 km with visual range of 23 km.

atmospheric transmittance starts becoming comparable to the no-aerosol case depicted in Fig. 3-15.

Figures 3-19, 3-20, and 3-21 describe hazy conditions at beginning of the boundary layer with visibility range at 5 km, again comparing sea level with 2- and 3-km altitude. For sea-level zenith transmittance, a wavelength of 1 μm (Figs. 3-19 and 3-16) is reduced to 0.60 from 0.85 or 1.7 dB and when scaled to 70 deg from zenith, this is nearly 4 dB worse (Eq. (3.2-7)). Figures 3-20 and 3-21 show a remarkable improvement in atmospheric attenuation with altitude assuming the beginning of the boundary layer is 1 km high. In fact, at 3 km altitude, the effect of aerosols has negligible impact on atmospheric attenuation at the wavelength of 1 μm (Figs. 3-21 and 3-18).

3.2.3.1 Atmospheric Attenuation Statistics. As shown above, MODTRAN can be effectively used to predict attenuation under a range of conditions that will be encountered during optical communication links through Earth's

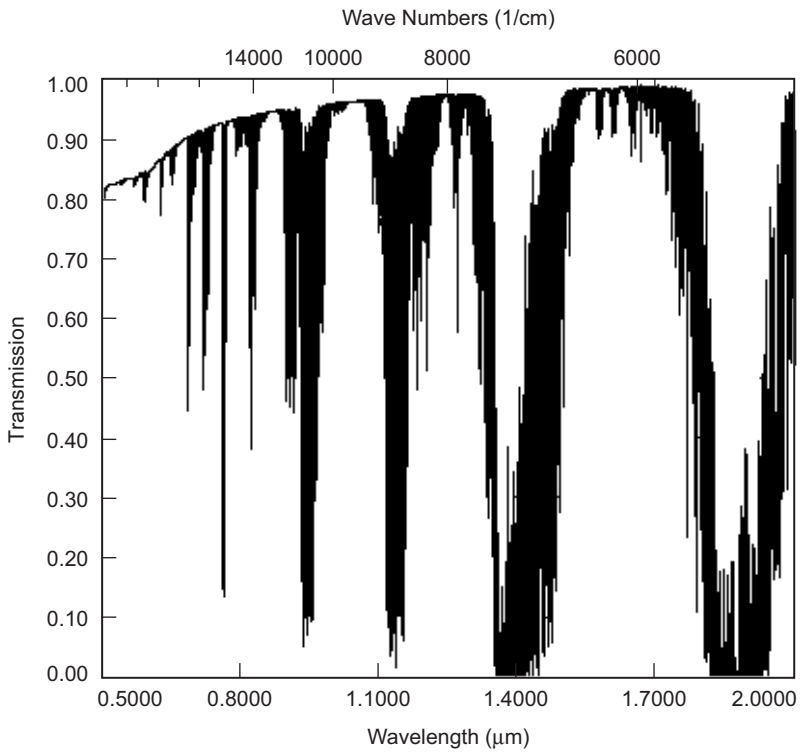


Fig. 3-18. Transmittance spectrum of the atmosphere for a space-to-ground link at 3 km of altitude with zenith angle 0 deg, mid-latitude. The boundary layer starts at 1 km with visual range of 23 km.

atmosphere. From an optical link design standpoint, what is not apparent from the model predictions is the statistical nature of the attenuation variations. How often actual atmospheric attenuation measurements will conform to one or the other kind of model assumptions is critical for designing and making long-term performance predictions for space-to-ground optical links. With this in mind, JPL initiated an Atmospheric Visibility Monitoring (AVM) program. The objective was to monitor the spectral attenuation through the atmosphere using stars as light sources observed through narrow bandpass optical filters. Currently the program is limited to use of silicon sensors and, therefore, measurements up to 1064 nm. At the time of writing this chapter, the state of knowledge is that statistics of atmospheric attenuation are available but purely in an empirical form. In general the AVM experience has been to obtain reliable day and nighttime data at the 860-nm band (10-nm wide), whereas with the 25-nm full-width half-maximum band centered at 1064 nm, mainly nighttime observations can be made since the silicon sensors decreasing response at 1064 nm yielded poor signal-to-noise for daytime measurements,

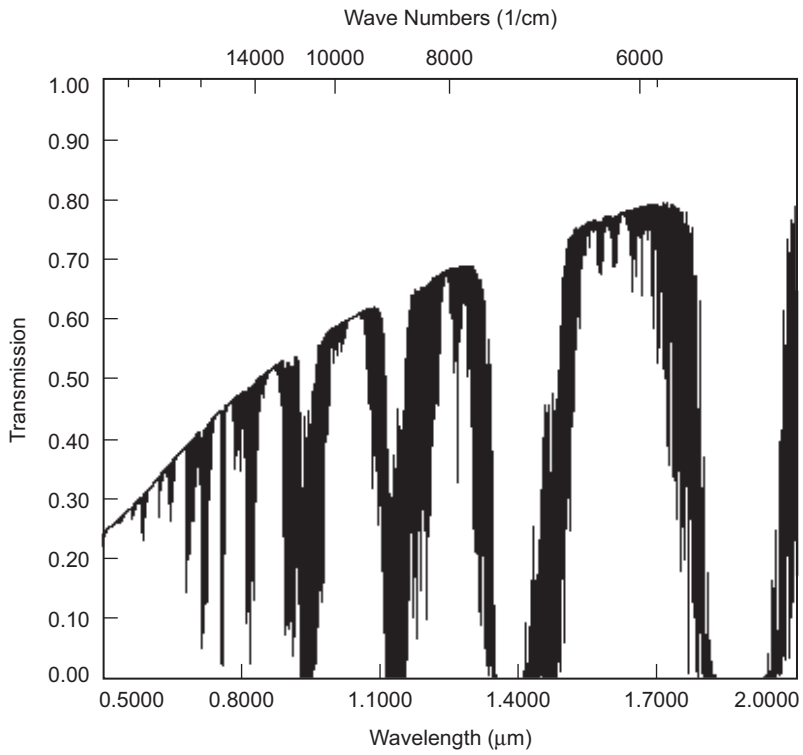


Fig. 3-19. Transmittance spectrum of the atmosphere for a space-to-ground link mid-latitude site at sea level with zenith angle 0 deg, and 5 km of visual range.

because of the increased background. Two tables of data (Tables 3-4 and 3-5) are reported below for 1064-nm and 860-nm reception, respectively. As stated, the 1064-nm results relate to nighttime only data, whereas the 860-nm data are relevant to daytime and nighttime data. The data presented in Tables 3-4 and 3-5 are almost all measured at the Table Mountain Facility (TMF), California, with just a single set of data shown for the Mount Lemmon (ML) observing station near Flagstaff, Arizona.

The site altitudes at TMF and ML are 2200 and 2800 m, respectively. Moreover, the cumulative probabilities apply to an airmass of 1. Table 3-5 shows that availability suffers during the first quarter months of January, February, and March due to increased cloud cover and precipitation associated with winter when it fell to as low as 40 percent availability. The remaining time availability was high, in the 60–80 percent range. Table 3-5 shows that wintertime availability at ML is better than at TMF. In general the observation can be made that the model predictions of benign (0.7-dB attenuation for the 23 km) visibility case at 2–3-km altitudes is borne out by the observations a very small fraction of time ranging typically from 15–30 percent of the time.

Table 3-4. Cumulative probabilities of atmospheric attenuation at 1064 nm (25-nm bandpass) for predominantly nighttime observations made with AVM at TMF.

Quarter	Station Uptime Fraction	<2 dB	<1.5 dB	<1 dB	<0.5 dB
Q3, 2002	0.33	0.8	0.79	0.73	0.15
Q4, 2002	0.62	0.62	0.6	0.56	0.45
Q1, 2003	0.85	NA	NA	NA	NA
Q1, 2003	0.4	0.62	0.58	0.5	0.2
Q2, 2003	0.85	0.73	0.73	0.7	0.4
Q3, 2003	0.7	0.68	0.65	0.53	0.4

Table 3-5. Cumulative probabilities of atmospheric attenuation at 860 nm (10-nm bandpass) for day and nighttime observations made with AVM at TMF.

Quarter	Station Uptime Fraction	<2 dB	<1.5 dB	<1 dB	<0.5 dB
Q3, 2002	0.33	0.87	0.83	0.78	0.35
Q4, 2002	0.62	0.65	0.62	0.58	0.35
Q1, 2003	0.85	0.4	0.35	0.25	0.05
Q1, 2003	0.4	0.5	0.46	0.45	0.27
Q2, 2003	0.85	0.83	0.8	0.75	0.4
Q3, 2003	0.7	0.81	0.72	0.53	0.2

From an optical-communications design standpoint, the statistics suggest designing links for 1-dB attenuation or equivalently 2.8 dB at zenith angles of 70 deg will cover 50–80 percent of the time across all seasons at any given station. The presumption here is that when conditions get worse there will be nearby sites in the network where the link will be switched to in order to achieve overall availability in the high 90s as discussed in the previous sections.

3.2.4 Sky Radiance

Just as atmospheric scattering deflects photons propagating from an optical communications transmitter to a ground receiver resulting in net signal loss, solar photons can be scattered in a manner that causes them to propagate along the transmit–receive line-of-sight path giving rise to unwanted background. The latter degrades the signal-to-noise ratio of the link. The extent of these daytime phenomena is dictated by the source geometry of the observer, the Sun, and the transmitting source. The exo-atmospheric solar spectral irradiance is shown in Fig. 3-22. Source irradiance describes the power emitted by a point source

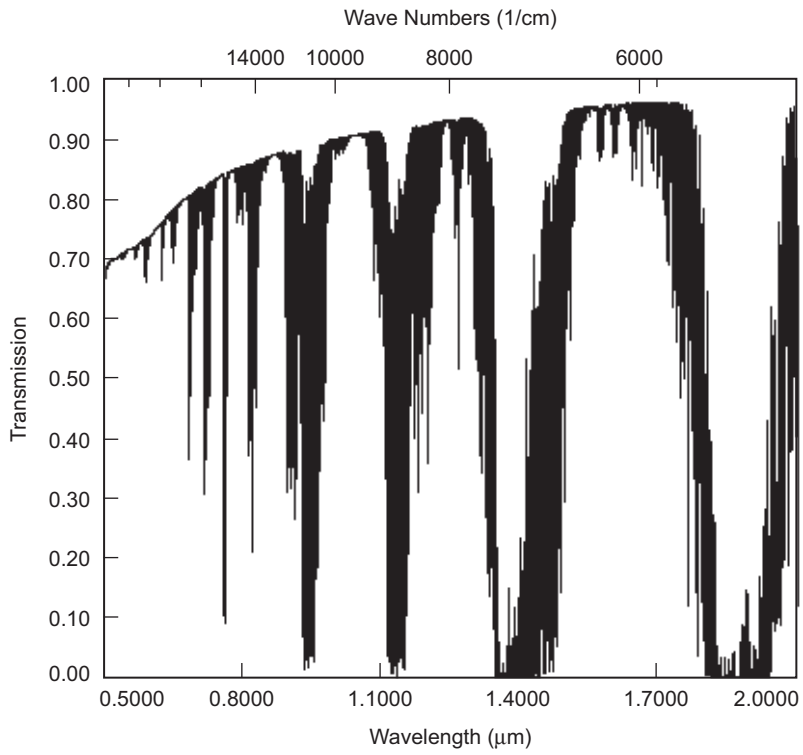


Fig. 3-20. Transmittance spectrum of the atmosphere for a space-to-ground link mid-latitude site at 2 km with zenith angle 0 deg. The boundary layer starts at 1 km with visual range of 5 km.

captured by unit area (receiver) over a spectral bandwidth, which in Fig. 3-22 is dimensionally W/(cm^2 μm).

Radiative transfer due to molecular and aerosol scattering in the atmosphere, and therefore the determination of the sky radiance, is not an easy problem to solve due to both the complexity of scattering theory and the fact that atmosphere is not a homogeneous medium, but instead it greatly varies with altitude. To simplify the problem, the atmosphere is modeled as a layered medium, with each layer consisting of a homogenous mixture of gas and aerosols. This concept is indicated in Fig. 3-23 where solar (S) radiation impinges on scatterer (P) and is redirected to an observer (O). In this figure the atmosphere is divided into a number of homogeneous layers (H_1, H_2, H_3…and so on), with the scattering angle γ between the forward direction of the Sun radiation and the observation point direction [21,22,23].

The basic premise of this model is to consider each single scatterer in a generic atmospheric layer, H_i, as a new scattering source of irradiance $J(\lambda)_i$

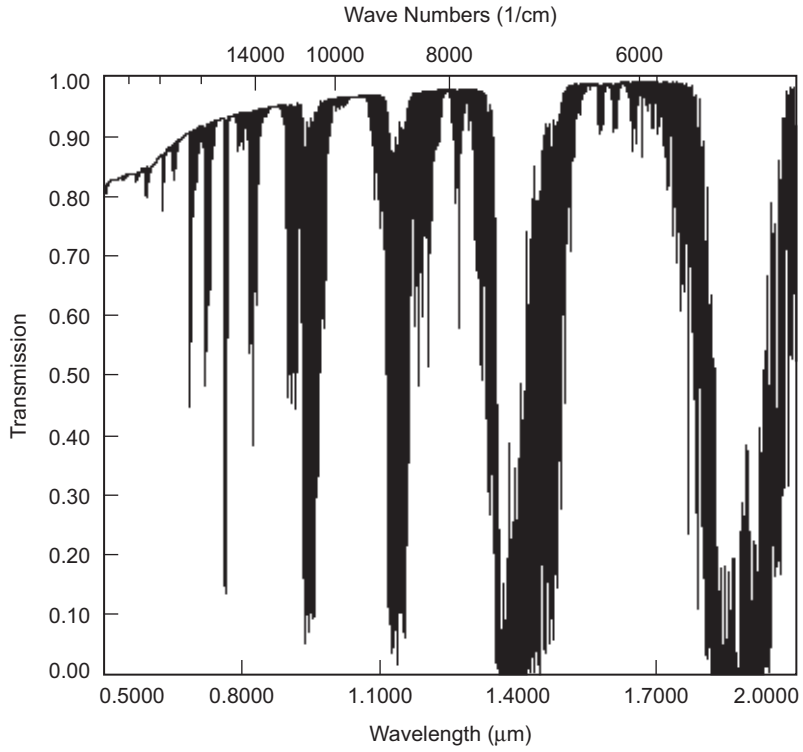


Fig. 3-21. Transmittance spectrum of the atmosphere for a space-to-ground link mid-latitude site at 3 km of altitude with zenith angle 0 deg. The boundary layer starts at 1 km with visual range of 5 km.

$$J(\lambda)_i = H_\lambda T_{sp}\left[p^a(\gamma)B_a + p^m(\gamma)B_m\right] \tag{3.2-13}$$

where H_λ is the exo-atmospheric Sun irradiance at the determined wavelength, T_{sp} is the atmospheric transmittance between Sun and the scatterer at the point P, while $p^a(\gamma)$ and $p^m(\gamma)$ are scattering phase functions for aerosol and molecular scattering describing the amount of energy scattered at the observer angle γ.

The total contribution from all the scattering source functions in the same atmospheric layer is therefore

$$L_i(\lambda,\theta,\varphi) = \int J(\lambda)_i T_{op} ds_{op} \tag{3.2-14}$$

where T_{op} is the atmospheric transmittance at the observation point represented by θ and φ, the observer zenith angle and angular distance between observer

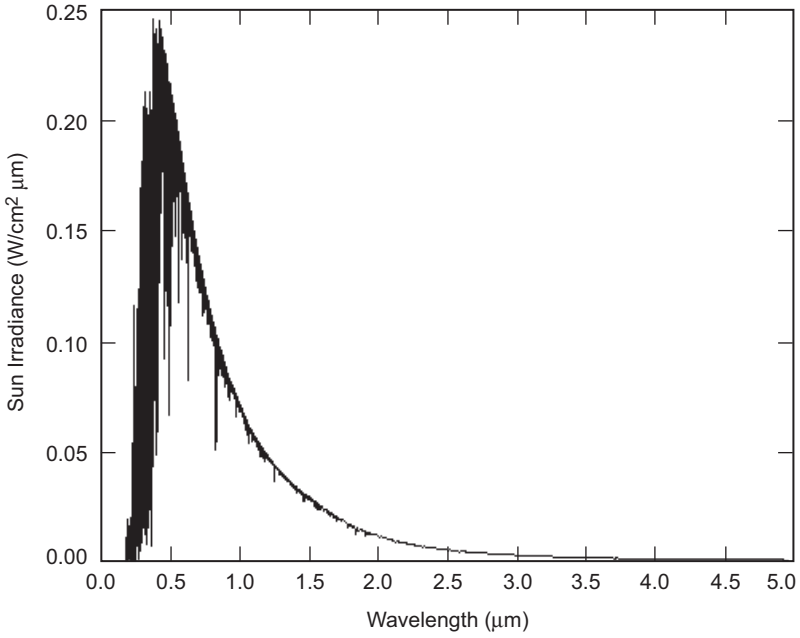


Fig. 3-22. Exo-atmospheric solar irradiance at 1 astronomical unit (AU).

and Sun zenith angles. The integral of Eq. (3.2-14) is defined over all the possible paths s_{op} between scatterer(s) and observation point. Finally, summing the contribution by all N atmospheric layers, one can get the total sky radiance [22,23]

$$L(\lambda,\theta,\varphi) = \sum_{j=1}^{N} L_j(\lambda,\theta,\varphi) \tag{3.2-15}$$

which dimensionally describes the power emitted by an extended source captured by unit area (receiver) over a spectral bandwidth at a given field of view, usually dimensionally is described as W/(cm^2 sr µm). Therefore, given a receiver aperture of diameter D cm, field of view Ω steradians, and $\Delta\lambda$ micron of bandpass filter, the total power P_{pbg} originated by sky radiance collected by the telescope is

$$P_{bg} = L(\lambda,\theta,\varphi)\frac{\pi D^2 \Omega \Delta\lambda}{4} \tag{3.2-16}$$

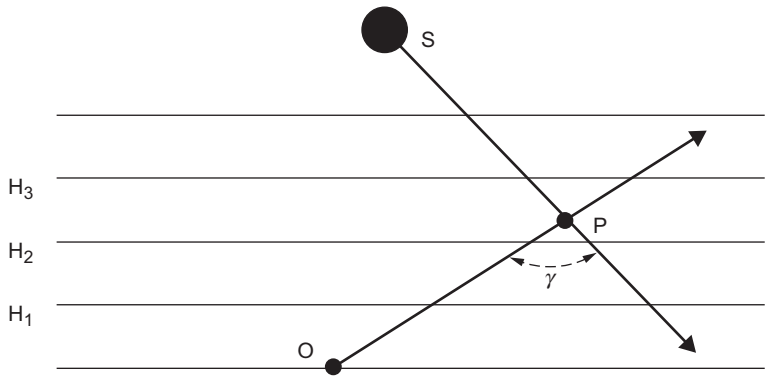


Fig. 3-23. Depiction of (single) scattering mechanism for a layered model of the atmosphere. S stands for the Sun (or any light source); P is the location of the aerosol; O is the observation point; γ is the scattering angle.

Scattering coefficients, phase functions, and transmittance vary greatly in the different atmospheric layers due to the diverse scatterer concentration.

To briefly summarize the consequences of Eq. (3.2-13) to Eq. (3.2-16), one may consider that:

- The higher the concentration of scatterers, the higher (generally, unless atmospheric transmittance is too low) is the sky radiance.
- The higher the altitude of the observer (telescope), the lower is the sky radiance because of lower concentration of scatterers.
- As the angular distance between observation direction and Sun decreases, (so does the scattering angle) the sky radiance increases.
- Sky radiance at small angular distance between observation direction and Sun φ is dominated by single scattering; however, with increased angular distance, the contribution of multiple scattering to sky radiance starts to dominate.
- Within 30 deg from the Sun, sky radiance is greatly dominated by aerosol contribution, as the angular distance from the Sun increases, molecular (Rayleigh) scattering, becomes more significant.

Figures 3-24 through 3-29 show sky radiance for different cases of telescope altitude and sky visibility over the spectrum of 500–2000 nm. For simplicity, we restricted the case of Sun zenith angle of 45 deg, the observer (telescope) zenith angle is instead at 10, 40, and 70 deg. The azimuth between observer and Sun is zero, and in all the cases, a rural aerosol model for a mid-latitude location during summer with observer location at sea level is assumed.

Figure 3-24 depicts the case of an observer at sea level. As clearly shown, the sky radiance is larger when the angle between the observation direction and the Sun is the smallest (5 deg). However, sky radiance is also large even when

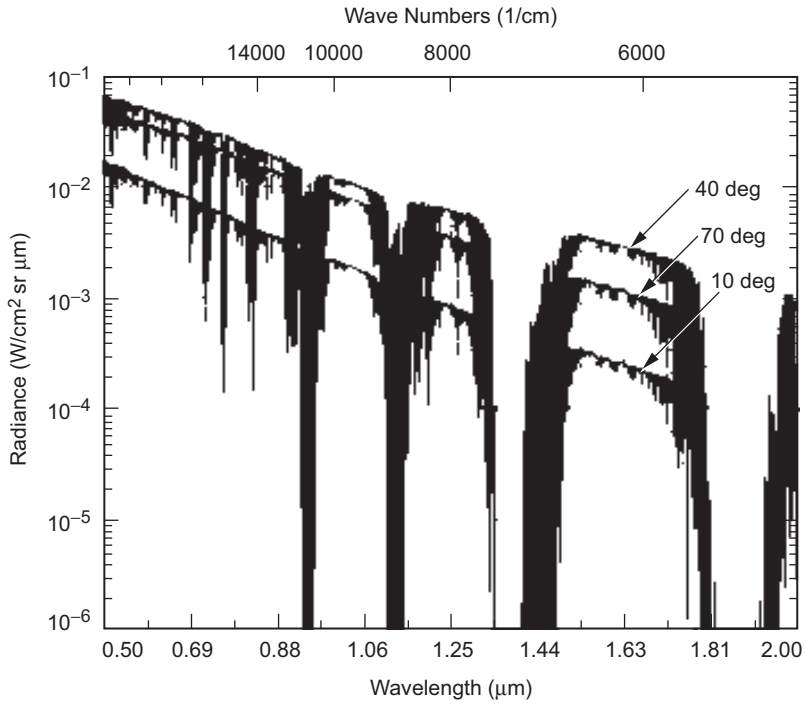


Fig. 3-24. Sky radiance spectrum experienced at an observation point at sea level for 23 km of visibility and Sun zenith angle of 45 deg while observer zenith angle varies as 10, 40, and 70 deg.

the observer angle is 70 deg. This is due to the fact that at a large zenith angle, the effective number of scatterers seen along the observation direction is larger. Figure 3-25 shows the sky radiance decrease upon raising observer location 2 km in altitude. Here the aerosol boundary layer is assumed to start at 1 km.

The reduction of sky radiance is due to the shorter path traversed by the light through the atmosphere, as well as, a smaller concentration of aerosols that scatter sunlight. Of course, such a reduction of the sky radiance is further accentuated at 3-km altitude, as seen in Fig. 3-26. Finally, for sake of completion, the same examples are repeated considering a visual range of 5 km (instead of 23 km) in Figs. 3-27 and 3-29. As expected, a larger aerosol concentration leads to larger sky radiance in Fig. 3-29, which corresponds to the case of 3-km altitude, showing that when the observer zenith angle is at 70 deg (30 deg from the Sun) the amount of sky radiance is larger than that one at 40 deg (5 deg from the Sun). This result, is related to the large difference in path integrated scatterer concentration along the two paths.

3.2.4.1 Sky Radiance Statistics. As discussed for atmospheric attenuation, the models presented in the preceding section are not indicative of sky radiance

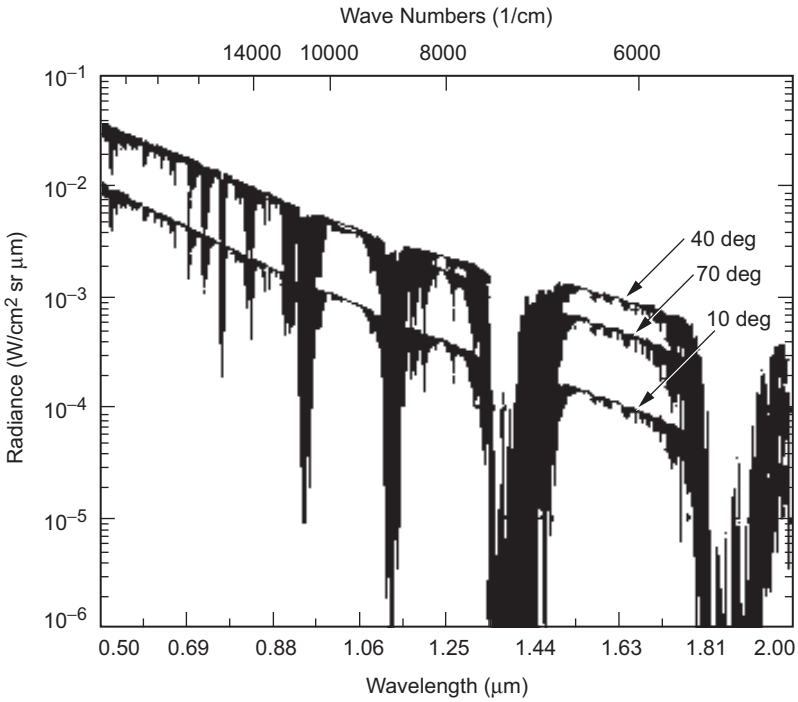


Fig. 3-25. Sky radiance spectrum experienced at an observation point at 2 km of altitude. The boundary layer starts at 1 km with a visual range of 23 km. Sun zenith angle is 45 deg while observer zenith angle varies as 10, 40, and 70 deg.

statistics as would be preferred by an optical communications systems designer. NASA has a global network of deployed Sun photometers that are used to monitor daytime sky radiance. This is known as the Aerosol Robotic Network (AERONET). The Sun photometers acquire sky radiance data in the course of extracting the aerosol optical thickness profiles. The sky radiance is reported in terms of principle plane and almucantar scans that provide a rich data set of sky radiance as a function of the position of the Sun in the sky. A few sets of data from this database were reduced in order to address sky radiance statistics at a few different sites.

Figure 3-30 shows a general comparison between the predicted and measured sky radiance at Table Mountain Facility (TMF), California. The sky radiance is presented as a function of Sun–Earth–Probe (SEP) angle. Noteworthy features displayed by Fig. 3-30 are the spread in measured sky radiance values at any give SEP angle. The measurements are over the period of a few months and are made at a wavelength of 1.026 μm for a zenith angle range of 55–60 deg. The spread is interpreted to be associated with a range of atmospheric conditions. The average of the measurements is shown by the

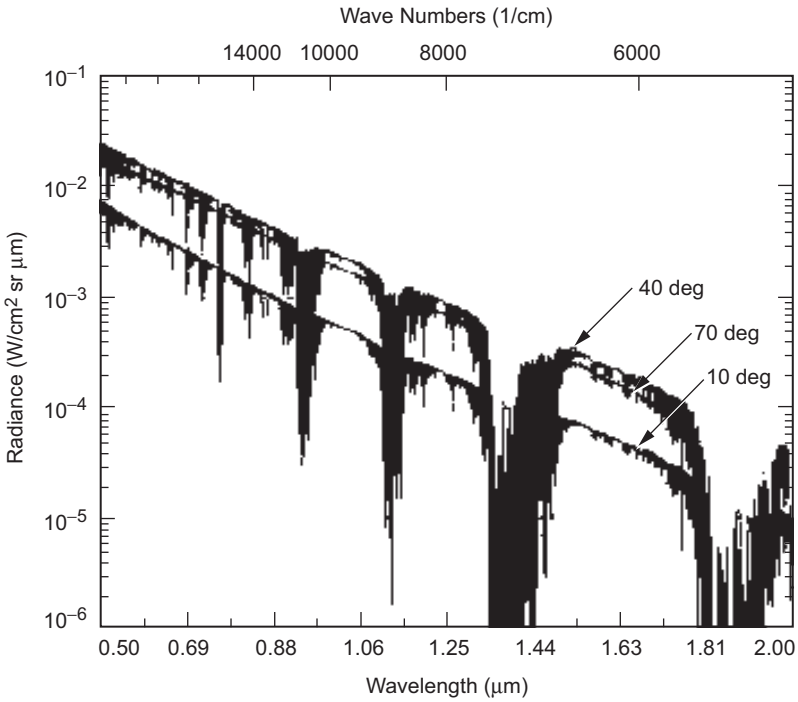


Fig. 3-26. Sky radiance spectrum experienced at an observation point at 3 km of altitude. The boundary layer starts at 1 km with a visual range of 23 km. Sun zenith angle is 45 deg while observer zenith angle varies as 10, 40, and 70 deg.

dotted line. A few MODTRAN sky radiance predictions for different aerosol and high cirrus cloud models at 1.064 μm are overlaid in Fig. 3-30. The comparison is reasonable.

While some evidence of the statistics is evident in Fig. 3-30, Fig. 3-31 shows cumulative distribution functions of sky radiance measured at TMF. The data are for SEP angles of 3 deg but for a few different solar zenith angles. Two separate data campaigns are represented, namely, data acquired in January–February of 2000 and then data acquired from June 2003–January 2004. The same calibrated instrument was used on the two separate occasions; however, the setup was dismantled and re-installed between the two data sets. Repeated in the data sets is the fact that for achieving any cumulative probability, larger zenith angles yield slightly lower sky radiance values, contrary to what the sky radiance models predict. For example, 50-percent cumulative probability at 65–70-deg zenith angle in the 2003 data set is 7×10^{-3} W/(cm^2 sr μm), whereas at the smaller zenith angle of 55–60 deg it is 1.5×10^{-2} W/(cm^2 sr μm). The same observation is generally true for the data gathered in 2000. The reason for departure from model behavior is not understood.

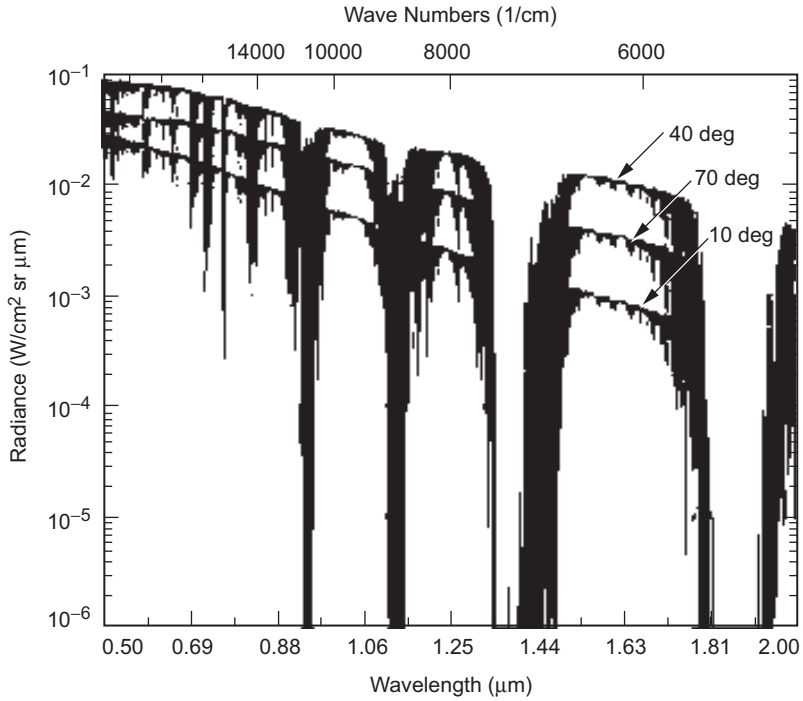


Fig. 3-27. Sky radiance spectrum experienced at an observation point at sea level with a visibility of 5 km. Sun zenith angle is 45 deg while observer zenith angle varies as 10, 40, and 70 deg.

Figure 3-32, on the other hand, shows data gathered at Roger's Dry Lake in California at an altitude of 680 m. In general, the sky radiances for this location appear to be higher than for TMF, consistent with the fact that it is located at a lower elevation. However, note that at Roger's Dry Lake the model behavior, namely larger sky radiance with zenith angle, is borne out.

3.2.5 Point Sources of Background Radiation

Sun-related sky radiance is the largest source of background noise that a telescope on Earth can collect pointing at a spacecraft. However, a telescope can also collect unwanted background light when (illuminated) planets or stars are in its FOV. These different sources of background radiations clearly need to be assessed in order to characterize the performances of a receiver.

In this subsection, therefore, we illustrate how to determine the background irradiance of a star of a given visual magnitude and temperature, and, later, the irradiance of a planet. To simplify, without lack of generalization, our discussion is limited to the case of absence of atmospheric interaction (e.g., a

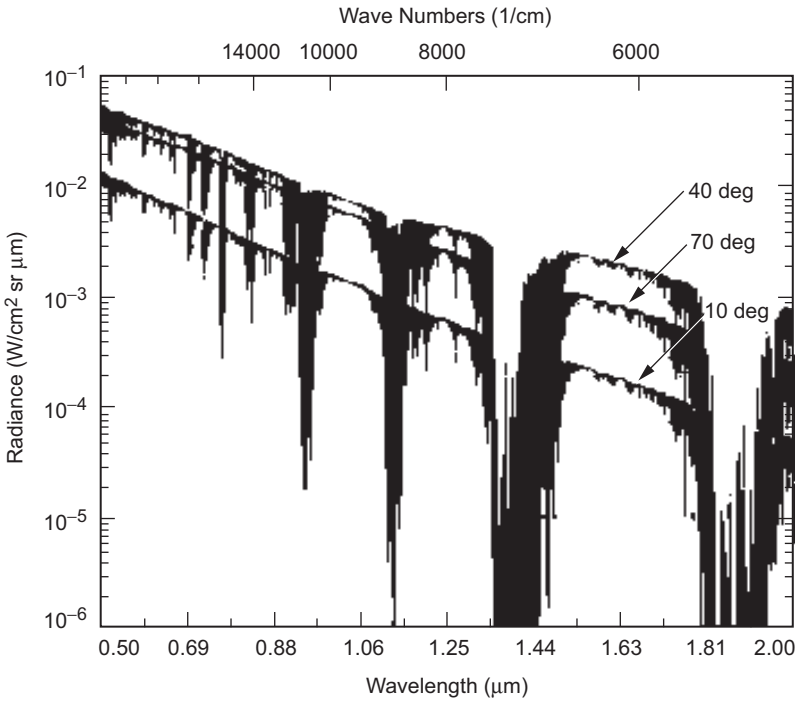


Fig. 3-28. Sky radiance spectrum experienced at an observation point at 2 km of altitude. The boundary layer starts at 1 km with a visual range of 5 km. Sun zenith angle is 45 deg, while observer zenith angle varies as 10, 40, and 70 deg.

receiver located aboard of spacecraft), while a comprehensive example (Mars in the FOV of a telescope on Earth) is shown at the end of this discussion.

Due to their small angular extension, stars (with the exception of the Sun) can be considered point sources because they are encompassed within a receiver FOV. To quantify the background light from a star, it is necessary to know the star's spectral irradiance S_λ. The star spectral irradiance defines the power collected by a receiver of a given collection area over a given spectral band (dimensionally W/m^3). Values of spectral irradiances for a number of stars can be found in Ref [24]. Otherwise, if the spectral irradiance of a star is not defined it can be calculated by knowledge of the radiation temperature T_s and its visual magnitude M_v. In fact, the spectrum of the star irradiance resembles (in shape) that of a black-body at given source temperature T_s such as

$$W(\lambda, T_s) = \frac{2\pi c^2 h}{\lambda^5} \frac{1}{\exp\left(hc / \lambda K T_s\right) - 1} \tag{3.2-17}$$

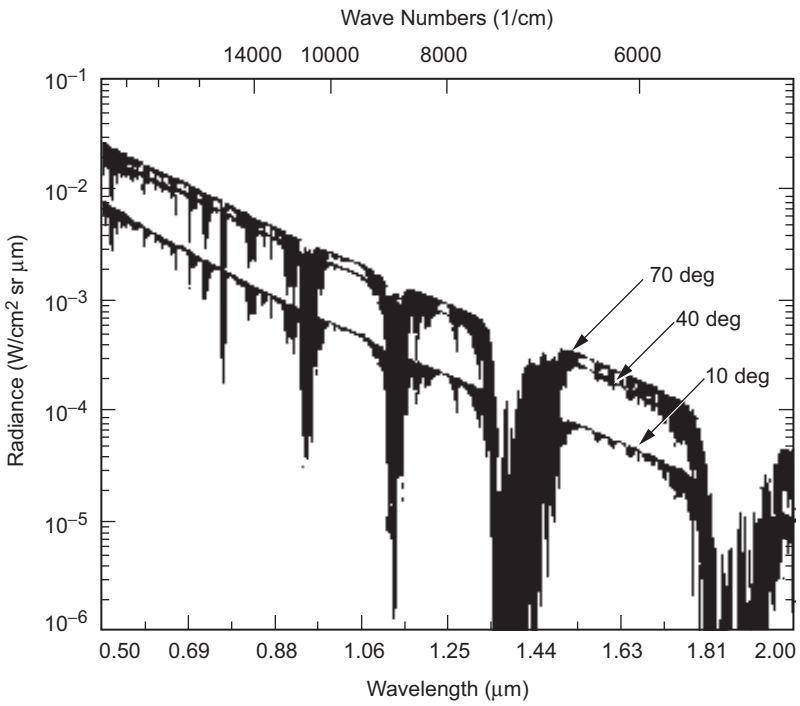


Fig. 3-29. Sky radiance spectrum experienced at an observation point at 3 km of altitude. The boundary layer starts at 1 km with a visual range of 5 km. Sun zenith angle is 45 deg, while observer zenith angle varies as 10, 40, and 70 deg.

where c is the speed of light in vacuo, h is Plank's constant, K is Boltzmann constant, and the blackbody emission is given in watts per meter. The spectral radiance emittance $W(\lambda, T_s)$ defines the power in Watts emitted by a square meter in the wavelength region $\lambda + d\lambda$. The wavelength in Eq. (3.2-17) is in meters. At its maximum value, $W(\lambda, T_s)$ and its corresponding peak wavelength λ_M are related to the source temperature as

$$\lambda_M = \frac{0.00289}{T_s} \tag{3.2-18}$$

where λ_M is in meters and T_s in kelvins. In essence, Eq. (3.2-17) teaches that given a star whose emission peaks at λ_M its spectral irradiance will be proportional to that one of a black body whose temperature T_s can be derived by Eq. (3.2-18). The proportionality constant that help us to calculate the star irradiance from the black body spectrum can be derived by the star visual magnitude.

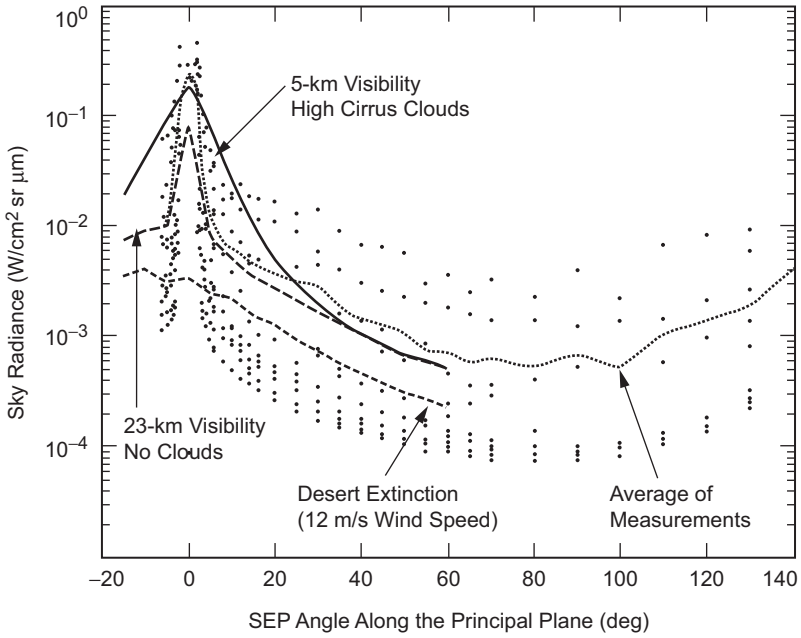


Fig. 3-30. A comparison between measured and predicted sky radiance as a function of SEP angle. The measurements (•) are at a wavelength of 1.02 μm and a zenith angle range of 55–60 deg. The solid lines are MODTRAN predictions for the indicated aerosol models at 1.064 μm.

The star visual magnitude M_v is a function of the star irradiance in the visible spectrum I_v defined as

$$M_v = -2.5\log_{10}\left(\frac{I_v}{3.1\times10^{-17}}\right) \tag{3.2-19}$$

where here I_v is in W/m^2. Using Eq. 3.2-17 to Eq. 3.2-19, it can be shown [24] that the star spectral irradiance of the can be written as

$$S_\lambda = 3.1\times10^{-\left(17+\frac{M_v}{2.5}\right)}\frac{W(\lambda,T_s)}{\int_0^\infty W(\lambda,T_s)e(\lambda)d\lambda} \tag{3.2-20}$$

where $e(\lambda)$ is the eye spectral response. The eye spectral response can be approximated by a triangular function that peaks e = 1 at 0.55 μm and is zero at 0.4 and 0.7 μm. Finally, one should notice that S_λ here is given in watts per square meter (of the receiver area) over the spectral bandwidth of interest.

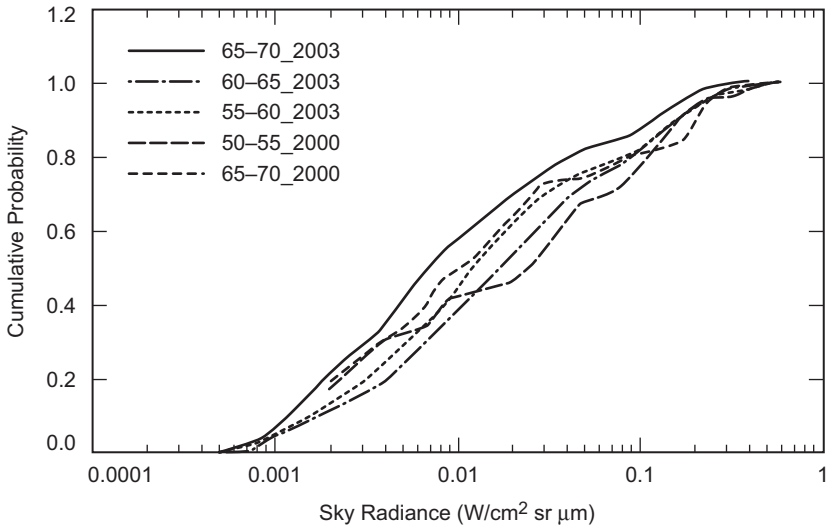


Fig. 3-31. Cumulative probability distribution of measured sky radiances at Table Mountain, California. The data sets marked 2003 were measured during June 2003–January 2004, while the data sets labeled 2000 were measured in January and February of 2000.

In the visible and near-infrared spectrum (our spectrum of interest so far), the planet irradiance is directly related to the reflection of sunlight (in the wavelength range between the mid infrared and far infrared range thermal emission of the planet must be taken into account [24]: this last case will not be discussed here).

Essentially, the planet irradiance can be described as the amount of the sunlight reflected by the planet surface (which is considered as a lambertian disk) and redirected towards the receiver. If a the receiver is located in space (outside the Earth atmosphere or in deep space where there is no interaction with gases and aerosols, implications for a receiver on Earth are introduced later in the section), one can write the planet irradiance E_λ at the receiver as [25]

$$E_\lambda = \frac{H_\lambda}{R_{AU}^2}\left(\frac{R_p}{Z_{PR}}\right)^2 a(\lambda) \tag{3.2-21}$$

where H_λ is the Sun spectral irradiance at 1 AU (Fig. 3-22) at the wavelength of interest, R_{AU} is the planet–Sun distance in AU, R_p is the radius of the planet, Z_{PR} is the planet–receiver distance, and $a(\lambda)$ is the planet spectral albedo. One can clearly notice, that the planet's irradiance is dimensionally related to the Sun spectral irradiance, which is usually indicated in the

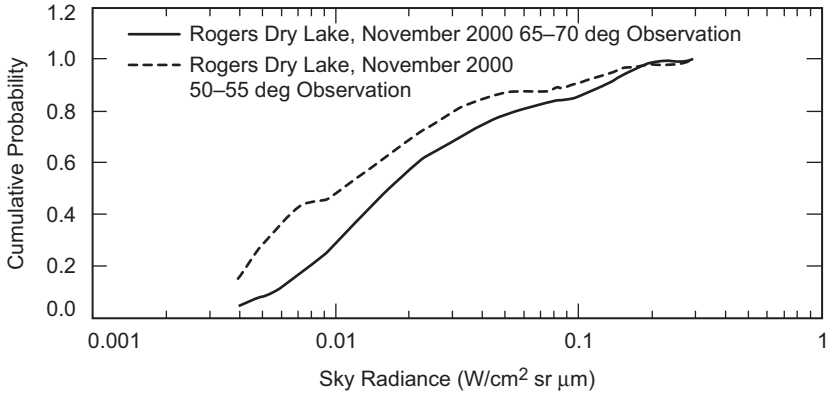


Fig. 3-32. Cumulative probability of sky radiance at a different location, Roger's Dry Lake, California.

Literature as W/(cm^2 μm). Therefore, if the receiver has an aperture of D centimeters in diameter, and the optical bandpass filter is $\Delta\lambda$ micrometers, the total planet background power, P_{pbg}, collected by the receiver is

$$P_{pbg} = E_\lambda \frac{\pi D^2 \Delta\lambda}{4} \tag{3.2-22}$$

Of course, Eq. (3.2-22) is valid if the angular extent of the planet is contained in the receiver FOV. If the angular extent of the planet exceeds the receiver FOV, only the fraction of the planet corresponding to the surface of the planet in the receiver FOV contributes to the background power. In this latter case, greater care must be taken when considering the planet geometric albedo. Some areas of the planet may not reflect uniformly because their albedo greatly depends on the composition (atmospheric and geological) of those specific areas. Therefore, the value of spectral albedo in Eq. (3.2-21) must correspond to these regions. For example, Fig. 3-33 depicts the variation of spectral albedo for different areas of Mars [25, 26].

As shown in Fig. 3-33, the Martian albedo can vary by a factor of four if the reflected light is coming from the dark mare areas or from bright desert area. Concerning the Sun–planet distance R_{AU}, since the orbit of the planets is not circular, there will be some variation due to the orbit eccentricity, and this variation must be taken into account. Table 3-6 summarizes data of orbital constants of planet, radius, and variation of solar irradiance at planetary distances. Notice that the ratio between the maximum and minimum irradiances for a fixed planet–receiver distance r_{Irr} is directly related to the eccentricity ε of the planet orbit as

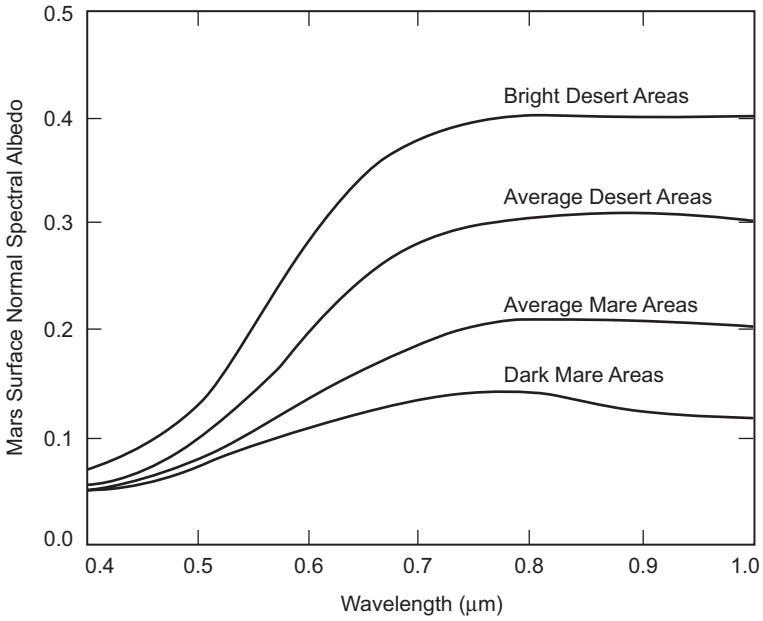


Fig. 3-33. Estimates of Mars spectral albedo for different surface areas.

$$r_{Irr} = \frac{(1+\varepsilon)^2}{(1-\varepsilon)^2} \tag{3.2-23}$$

To have a more precise evaluation of the planet irradiance as in Eq. (3.2-21), one must also consider the possible dependency of the planet irradiance on the sunlit sector of the planet seen from the receiver. The sunlit sector of the planet seen by the receiver depends on the phase angle. The phase angle, here indicated as ϕ_a in Fig. 3-34, is defined by Sun–planet–receiver angle. In fact, depending on the phase angle, one can notice that not all the disk corresponding to the surface illuminated by the Sun contributes to the planetary irradiance, but just a fraction of it. Considering the planet–Sun–receiver angle γ_s, elementary geometry shows that

$$\pi - \phi_a = \gamma_s + \theta_r \tag{3.2-24}$$

where θ_r is the Sun–receiver-planet angle, which is given by

$$\theta_r = \frac{1}{\sin(\gamma_s)}\left[\frac{R_{SR}}{R_{AU}} - \cos(\gamma_s)\right] \tag{3.2-25}$$

Table 3-6. Orbital constants and radii of the planets.

Planet	Semi-Major Axis of Orbit (AU)	Planet Radius (km)	Sidereal Period (days)	Eccentricity of the Orbit (ε)	Ratio Max/Min Irradiance
Mercury	0.387	2439	87.96	0.205	2.303
Venus	0.723	6051	224.7	0.006	1.028
Earth	1	6371	365.257	0.016	1.069
Mars	1.523	3390	686.98	0.093	1.454
Jupiter	5.203	69882	4332.58	0.048	1.212
Saturn	9.55	58234	10759.2	0.052	1.236
Uranus	19.18	25362	30685	0.049	1.218
Neptune	30.07	24622	60188	0.004	1.018
Pluto	39.44	1151	90700	0.252	2.806

Trigonometric calculations show then that the fraction of lit planetary disc is

$$\sin^2\left(\frac{\pi-\phi_a}{2}\right)=\sin^2\left(\frac{\gamma_s+\theta_r}{2}\right) \tag{3.2-26}$$

Of course, as long as the planet appears as an extended background source, its noise contribution is not affected by the above considerations. However, when the planet appears as a point source, its irradiance must be corrected by the lit fraction, as in Eq. (3.2-26). One should notice that this dependency is stronger for inner planets with respect to the receiver (e.g., receiver on Earth and with Mercury in the FOV). For outer planets (e.g., receiver on Earth and with Mars in the FOV), the fraction of the area lit from the Sun is closer to unity. For example, the Mars lit fraction is at the minimum of 87 percent, while for planets from Jupiter and beyond it is 99 percent.

Besides the lit fraction of a planet, there is also a dependence of the geometric albedo on the phase angle because the planet does not act as perfect lambertian reflector. This dependence is expressed by the phase function $f(\phi_a)$, which is shown for Mars [27] in Fig. 3-35. Typical features of the phase function are a linear part for phase angle exceeding approximately 10 deg, and higher order components for phase angles smaller approximately 5 deg. This enhanced reflectivity at small phase angles is called the “opposition effect.” Generally, albedos of planets with atmospheres have a smaller dependence on the phase angle than planets without atmospheres. Moreover, one must consider that the planet’s geometric albedo depends strongly on the wavelength of operation. A number of physical reasons contribute to this dependence, such as

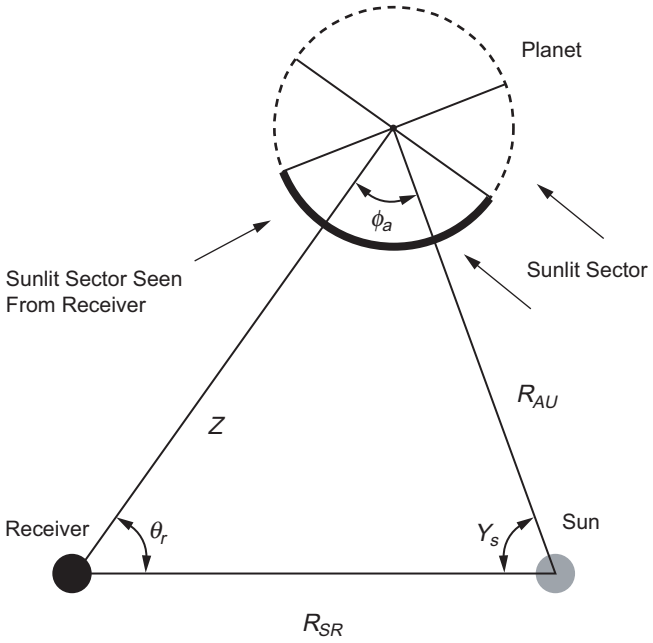


Fig. 3-34. Sun–planet–receiver system (for an inner planet).

planet atmosphere and Raman scattering. Examples of spectral variation of geometric albedo have already been shown in Fig. 3-33 for different area of Mars, while the average spectral variation of the geometric albedo for several other planets is indicated Fig. 3-36. The spectral albedos of Fig. 3-36 are at low resolution and therefore do not take into account of the absorption lines deriving by the reflecting planet atmospheric interactions.

Finally, one must remember that Eq. (3.2-20) and Eq. (3.2-21) refer to star and planet irradiance outside the atmosphere. For a receiver located on Earth, the irradiance is filtered by Earth's atmospheric transmittance $T(\lambda)$, and therefore, for example, a planet irradiance on Earth must be written as

$$E_{\lambda-Earth} = \frac{H_\lambda}{R_{AU}^2}\left(\frac{R_p}{Z_{PR}}\right)^2 a(\lambda)T(\lambda) \tag{3.2-27}$$

As an example of a planet irradiance seen on Earth, one can consider the case of a Mars–Earth downlink. Mars is in the FOV of a telescope located at 2 km above sea level with atmospheric visibility as in Fig. 3-17. The zenith angle of the telescope on Earth is 70 deg. Mars is at 2.4 AU, and the Sun–Mars distance is 1.4 AU. Considering the Sun irradiance as in Fig. 3-22, the Mars average albedo as in Fig. 3-31, for a phase angle close to zero, the Mars

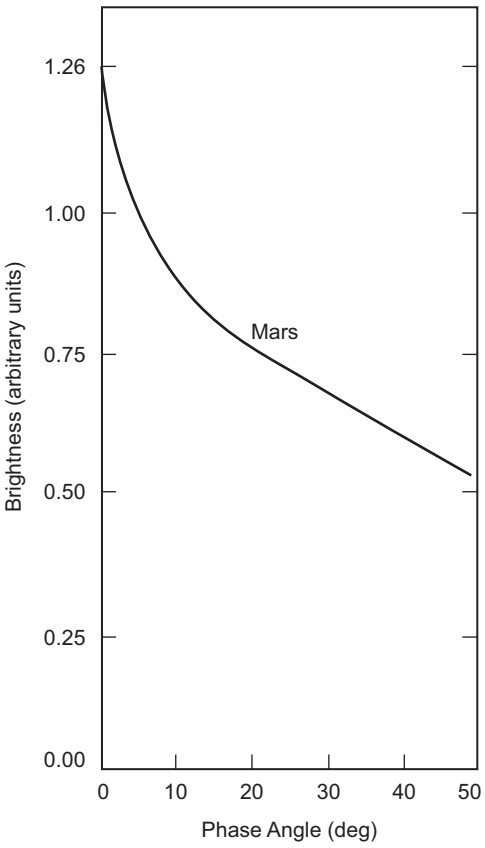


Fig. 3-35. Phase function for Mars after [25].

irradiance seen at the Earth receiver is indicated in Fig. 3-37. From Fig. 3-37, one may notice how the initial Sun irradiance is filtered by the Mars spectral albedo, which reduces spectral components in the visible, and also by the Earth's atmospheric transmittance, which blocks the forbidden bands in its spectrum.

Finally, to understand the impact of planet irradiance in a groundbased downlink scenario, one should compare the possible contribution of planet irradiance with sky background during daytime operations. In fact, comparing values of sky radiance from the previous section and Mars irradiance from Fig. 3-37, it is easy to convince oneself that background noise collected by Sun sky radiance is a number of orders of magnitude greater than that due to planet irradiance. Therefore, during daytime, the greatest source of background noise is represented by sky radiance, which otherwise is absent during nighttime.

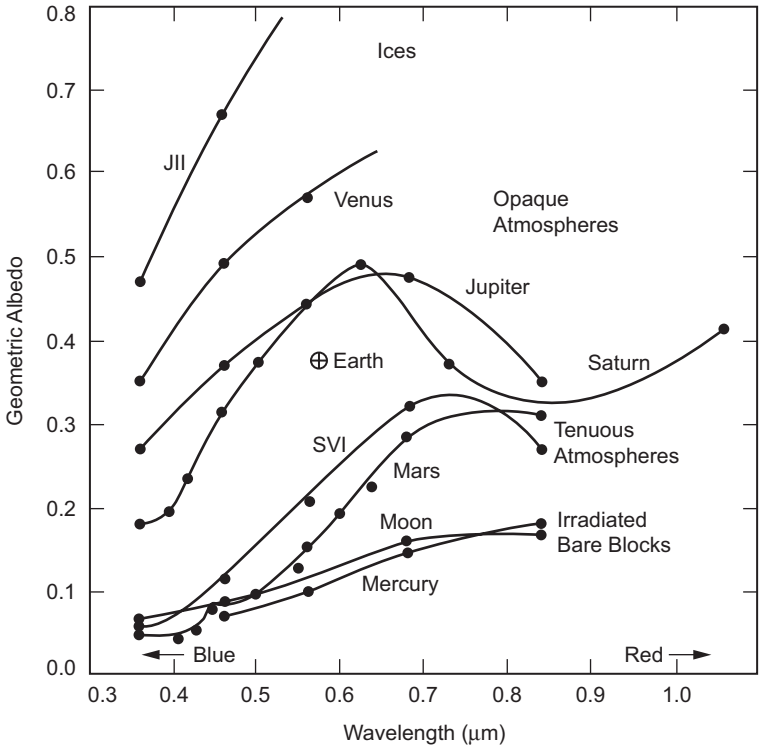


Fig. 3-36. Spectral variation of the geometric albedo for various planetary bodies (after Ref. 10 and Ref. 13).

3.3 Atmospheric Issues on Ground Telescope Site Selection for an Optical Deep Space Network

3.3.1 Optical Deep Space Network

To support deep space missions aimed to the exploration of the universe for the last four decades, NASA has designed and operated a global network of radio-frequency ground stations termed the Deep Space Network (DSN). Clearly, as the use of optical wavelengths has become a feasible technological option for deep space missions, future deployment of an optical deep space network (ODSN) might replicate the function of the DSN. However, the design of an ODSN poses new challenges in terms of mission requirements, mitigation of weather effects, life-cycle cost, and optimization of antenna (telescope) performances. Figure 3-38 summarizes some of the dynamic interactions among ODSN parameters and logistics, environmental, and technological variables.

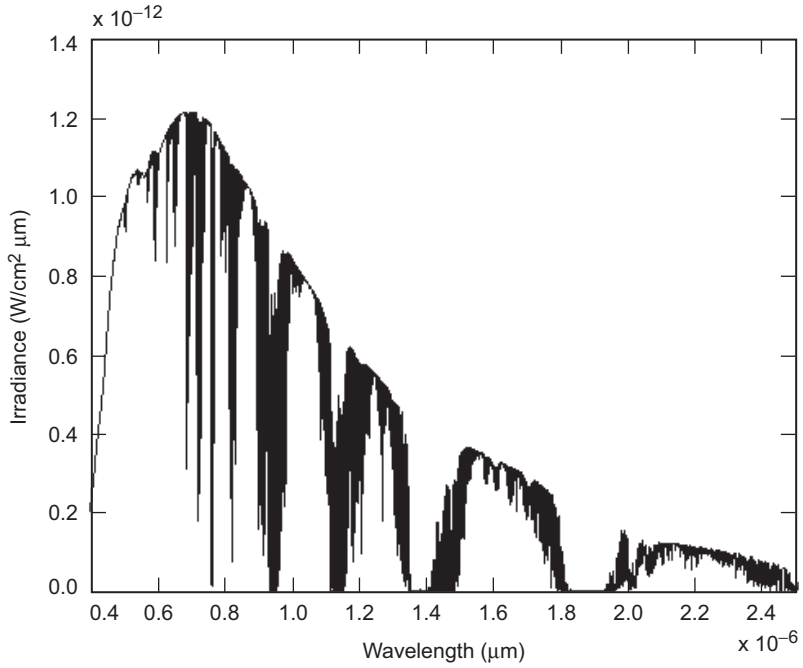


Fig. 3-37. Mars irradiance as seen when the planet is in the FOV of a telescope on Earth, located at 2 km above sea level, with an observation zenith angle of 70 deg, and with clear sky. The Earth–Mars distance is 2.4 AU, and the Mars–Sun distance is 1.4 AU. The phase angle is close to zero.

Note that in Fig. 3-38, three main parameters of the ODSN are listed as being the data delivery capacity and accuracy (data rate/BER), the continuous Earth coverage (network continuity), and the proper location of the nodes (i.e., optical communication telescopes) of the network itself (ground telescope sites). The data rate/BER is clearly influenced by the amount of signal photon flux collected by the aperture of the ground telescope site. This signal flux depends upon, among other factors, the angular spread of the laser beam and the atmospheric transmittance experienced at the receiver. Both atmospheric transmittance and laser beam width are related to the selected transmitter (spacecraft) wavelength, which also determines the sky background radiance, which during daytime operation contributes to increasing the noise level at the detector level.

The number of sky background photons collected by the detector is also determined by the receiver FOV, which is related to effects of atmospheric turbulence. Analysis of atmospheric turbulence effects is presented in a later section of this chapter. To ensure the continuous coverage of the Earth from deep space, despite its rotation, it is necessary to distribute a number of ground

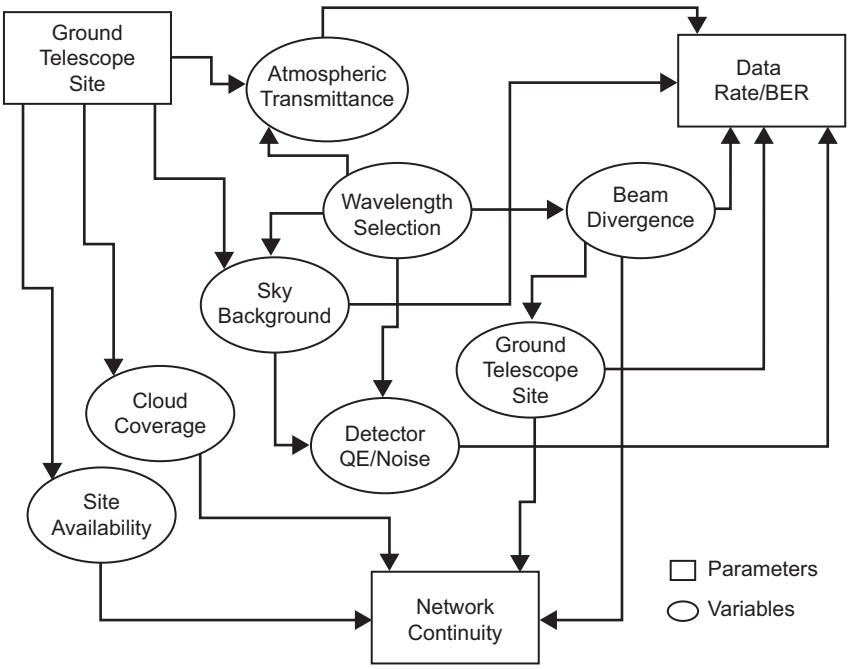


Fig. 3-38. Flowchart illustrating the dynamics among the main ODSN parameters and variables.

telescopes around the globe. Today NASA's DSN only requires three radio-telescope hubs (ground station complexes) to successfully operate the network. The DSN stations (located at approximately 120 deg of separation around the Earth: Goldstone, California; Madrid, Spain; and Canberra, Australia) allow continuous coverage of deep space from Earth. However, operation of the future ODSN will require a different geographical and logistical approach. Since the laser transmitter beam width from space can (usually) cover a limited area (footprint) on Earth it is necessary that the ODSN consists of a number of ground stations located around the Earth as a linear distributed optical subnet (LDOS) [28], Fig. 3-39. The idea behind LDOS is to have the spacecraft always pointing at a visible station belonging to the LDOS. When either the line of sight is too low on the horizon (20 deg of elevation) or is blocked by atmospheric conditions (e.g., clouds or low transmittance), the spacecraft beam is switched to a different station (or network node) by pointing to the adjacent optical ground station. Of course the adjacent station must be located in a geographical area where the atmospheric conditions are uncorrelated (or better, anti-correlated) with the previous station in order to optimize network continuity.

To simplify both the spacecraft re-pointing process and the network hand-off between stations, another network architecture has been proposed. The clustered optical subnet, or COS [28], consists of a number of optical hubs (three or more) distributed around the Earth, with the difference that each hub is composed of more than one ground station (e.g., two or three). Each ground station of a hub (circle in Fig. 3-39) is located in a geographical area having (dry) weather pattern that is uncorrelated (or better anticorrelated) to the other stations to optimize the overall hub availability having at least one station with clear line-of-sight with the spacecraft (Fig. 3-39).

Of course, the location of the ground telescope is critical to the assurance of network operation continuity and is directly linked to the data rate/BER performances. As stated earlier, cloud coverage at the ground station has to be as low as possible to minimize link blockage, and it must be somehow predictable for program station operation. Moreover, at the ground station, the link must experience the highest atmospheric transmittance and lowest sky background possible during daytime. All of the previous atmospheric conditions are optimized when the optical ground station is located at high altitude because the signal atmospheric path is reduced and so, consequently, is its interaction with the atmosphere. Moreover, local microclimatic conditions that usually generate low clouds are not influential at high altitudes (usually more than 2000 m), which reduce the overall cloud coverage at the station. At the same time, the ODSN network continuity requirements demand a regular distribution around the Earth of peaks that may accommodate potential ground stations. Unfortunately, the global scarcity of potential telescope sites around the Earth and their uneven distribution (along with ever-present geopolitical implications) makes their identification even more complex for the design of a global ODSN.

Therefore, we describe an analysis and methodology that can be used to identify possible peak candidates for a future ODSN. The approach is as follows. First, we define a baseline optical deep space mission. By determining characteristics of an optical communication payload on the spacecraft and using a link budget, we calculate the photon flux reaching the Earth. Then, modeling the atmospheric effects along the atmospheric profile, we determine the atmospheric losses, the background photon noise, and the receiver performance at different peak altitudes, which helps identify the optimal peak elevation for an individual ground station in an ODSN. Finally, we study the global distribution of the Earth’s peaks and landmass elevations at the required altitude, and we introduce determined conditions about the required low-cloud coverage. Results from this last step will help in selecting the telescope sites for the ODSN and in analyzing the advantages of LDOS versus COS (or vice versa).

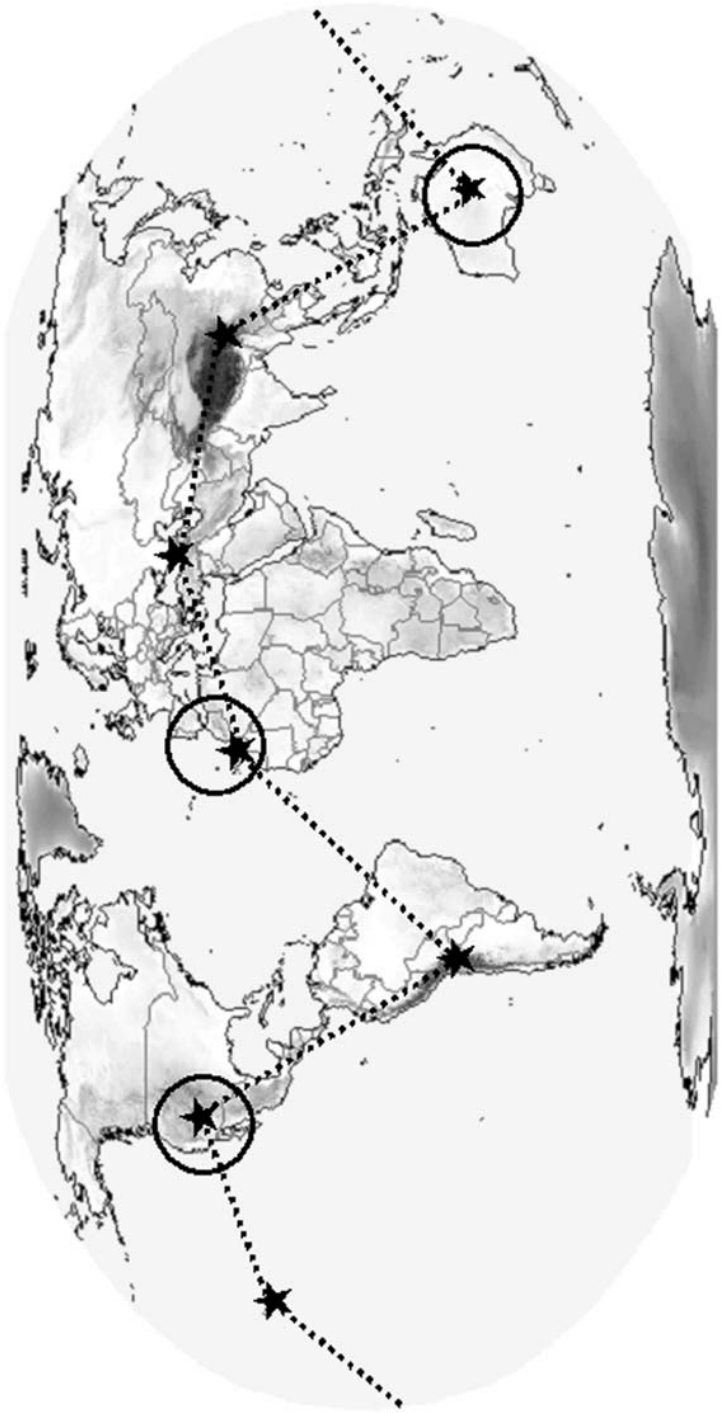


Fig. 3-39. Example of LDOS (star = telescope) and COS (circle = hub) architectures for an optical deep space network (ODSN).

3.3.2 Data Rate/BER of a Mission

The long-term objective of the ODSN is to provide ground support for Solar System exploration. In doing so, a practical and logical step is to base the ODSN analysis and site selection strategy around a specific mission and use it as a reference model to begin the point design.

A first logical choice for a reference mission for the ODSN is to understand how it may support a Mars mission. To design the ODSN as a support for a Mars mission, the next logical step is to derive an initial link budget based on the requirements, and then to analyze how the telescope aperture and the telescope location (via the atmospheric transmittance and daytime sky radiance noise) may affect the link budget itself. Specifically, the mission is required to provide a link at 1 million bits per second (Mbps), with an uncoded bit error rate (BER) of 0.001 at the largest distance of separation between Mars and Earth of 2.4 AU. The spacecraft laser has 5 W of average power, and the wavelength selected is $\lambda = 1064$ nm. (Another possible option is to consider 1550 nm for the laser wavelength.) The modulation used is M-ary Pulse Position Modulation (M-PPM) with M = 256, which corresponds to a 31-ns pulse. The spacecraft telescope has an aperture of 30 cm diameter with a linear obscuration of 10 percent. Transmitter loss was set to be 1.42 dB (72% of the laser power). Given these data, we ran a link with results as summarized in Table 3-7.

Because we did not restrict the ground telescope to a specific site (and therefore to a specific atmospheric condition), the link budget of Table 3-7 does not indicate any atmospheric loss. Concerning the receiver, we supposed an optical loss of 2.21 dB (60 percent transmission), and we normalized the receiver aperture to 1 m in diameter with linear obscuration of 20 percent to better describe the photon/flux per telescope aperture at the detector of a telescope on Earth. Losses of non-ideal synchronization and pulse amplitude were also added. Table 3-7 shows that in these conditions photon flux is 10.45 photons per pulse at the detector. To complete the information on link performance, a brief characterization of the receiver is necessary. Because our intent in this monograph is to consider a general detection case, we hypothesized a photodetector of quantum efficiency of 50 percent. Thermal noise is then not considered, which can be an appropriate hypothesis in the case of a cryogenic receiver with low noise amplification [29]. Noise from photodetector dark counts is also not considered (photodetector dark counts are greatly reduced when the photodetector is cooled to cryogenic temperature [30]).

3.3.3 Telescope Site Location

As already presented in a previous subsection, Earth's atmosphere affects the optical signal from deep space in two ways. First, when the optical signal

Table 3-7. Link Summary.
Bit Rate: 1.0 Mbps Modulation: PPM (M = 256)
Range: 3.59×10^8 km BER: 0.0010

Parameter	Description		Budget
Transmitter power	5.0-W average	31-ns slot time	61.08 dBm
Optical transmitter losses	72% transmitted		–1.42 dB
Transmitter gain	30.0-cm aperture	5.98 μr beam-width	117.67 dB
Pointing losses			–2 dB
Space loss	3.59×10^8 km	2.4 AU	–372.54 dB
Atmospheric transmission	100.0% transmitted	No atmosphere	0.0 dB
Receiver telescope gain	1.0-m aperture	20% obscured	129.40 dB
Optical receiver losses	60% transmitted		–2.21 dB
Non-ideal bit synch. adjustments			–1.0 dB
Pulse amplitude variation adjustments			–1.0 dB
Peak signal power at detector	10.45 photons/pulse	0.06262-nW peak	**–72.03 dBm**

goes through the atmosphere, it is partially (if the wavelength is not in the forbidden bands) attenuated. The longer the path through the atmosphere, the lower the atmospheric transmittance. Therefore, the higher the telescope's altitude, the higher the atmospheric transmittance. Moreover, the larger the observation zenith angle, the lower the atmospheric transmittance. Second, during daytime, the sunlight scattered by the atmosphere causes a number of unwanted photons to be collected by the telescope aperture, increasing the noise level at the receiver and badly affecting the receiver performance (BER) itself. Again, sky radiance is dependent on the sunlight's path through the atmosphere. Moreover, sky radiance depends on the concentration of aerosol suspended in the atmosphere, and finally it depends on the Sun–Earth–Probe (SEP) separation angle. To guarantee the largest continuity of the data delivery, it is recommended that the SEP angle be as low as possible. In this study, we assume a SEP of 5 deg. Also, in order to limit the number of stations deployed by the ODSN, an optical communication telescope must be able to observe the sky at a large zenith angle (low elevation angle). In our study, therefore, we set this limit at 70 deg of zenith angle.

A good baseline for the ODSN is to require that the ground stations work in the worst conditions for transmission and sky radiance (except the case of overcast sky where the link cannot be closed at all) that correspond, from our

assumptions above, to the case of 70 deg from zenith for observation angle and 5 deg of separation from the Sun during daytime. (One should notice that star and planet irradiances in the FOV of the telescope during daytime are much less than the sky radiance; and therefore, it is possible to ignore them without loss of accuracy.)

The MODTRAN simulation program [14] was used to describe values of sky radiance and atmospheric transmittance at different altitudes over the Earth. The Fig. 3-40 simulation considers altitudes between 0.5 and 3.5 km. The simulation refers to an atmospheric profile typical of a mid-latitude region, with the rural aerosol model, having its boundary layer starting at 0.5 km. Two cases of aerosol concentration are indicated, clear sky (visual range of 23 km at the bottom of the boundary layer) and hazy (visual range of 5 km at the bottom of the boundary layer). Keeping in mind that the aerosol concentration decreases exponentially starting at the beginning of the boundary layer, Fig. 3-40 shows that at 2 km of altitude, transmittance and radiance are independent of the aerosol concentration at the boundary layer. In Fig. 3-40, the dashed line describes the case of a rural aerosol model with a visual range of 5 km (hazy

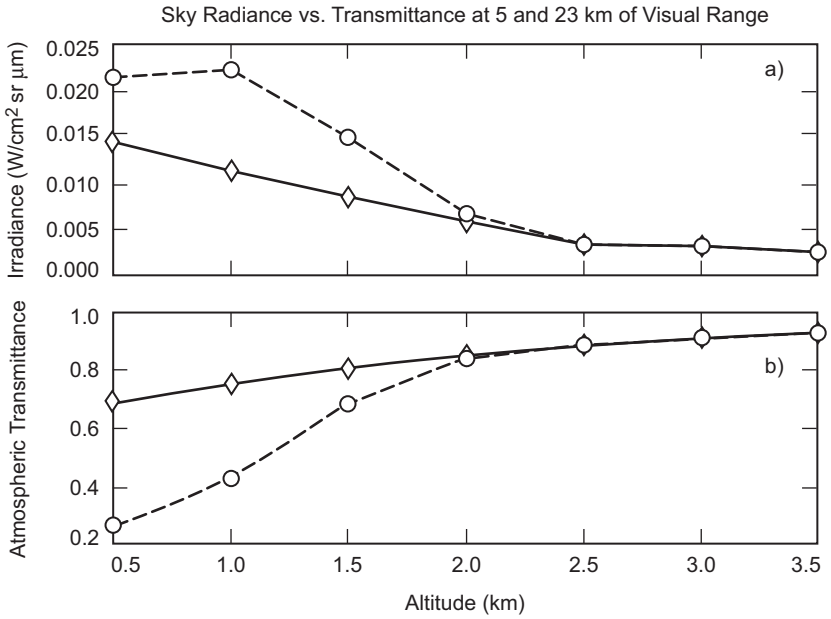


Fig. 3-40. (a) Sky radiance versus transmittance at 1064 nm for varying altitude of a telescope at 70-deg zenith angle with a 5-deg SEP angle. The dashed line describes the case of a rural aerosol model with visual range of 5 km (hazy sky) at the bottom of the boundary layer. The continuous line is for visual range of 23 km (clear sky) at the bottom of the boundary layer. (b) Atmospheric transmittance at 1064 nm for varying altitude for a telescope at 70-deg zenith angle. The dashed and continuous lines relate to atmospheric conditions as in (a).

From examination of Fig. 3-41, one can see that there are obvious compromises between aperture size and site altitude. Of course, the deployment of a smaller-aperture telescope has its own advantages, mainly related to the cost [31,32]. However, the scarcity of peaks available at higher altitude may also make it easier to find a lower altitude point that can house a large-aperture telescope (e.g., 10 m) for deep space optical communications. Moreover in the design of a global ODSN, the dichotomy of the problem "telescope aperture vs. site altitude" is even more critical. In fact, in a global ODSN, each single telescope must be located with precise coverage requirements that depend on the location of all the ground stations in the ODSN itself.

3.3.4 Network Continuity and Peaks

In a global ODSN, in principle the sites selected need to meet most, if not all, of the following conditions.

1) Latitude in proximity of the equator to better track spacecraft in the Solar System ecliptic. In this work we consider the latitude within range of ±40 deg.
2) Longitude according to the architecture requirements, in our case according to LDOS or COS requirements.
3) A minimum mutual view period of 4 hours with at least one other site, to allow smooth hand-off of the operations.
4) Absence of geopolitical issues for site locations outside the United States.
5) Close to pre-existing facilities for easy installation and operation.
6) Low time-duration (year-long) cloud coverage with fairly constant and predictable weather.
7) High altitude for high atmospheric transmittance and low sky radiance, as derived in Subsection 3.2.4.
8) Favorable atmospheric seeing (as explained in the next subsection).

Considering the results obtained in the previous subsection, we derived the baseline that when selecting a site for a 5-m aperture telescope, an optimal site altitude would be 1.9 km, while for a 10-m aperture the requirement can be relaxed to 1.2 km. Unfortunately, there is an overall scarcity of high-elevation land on Earth, as indicated in Fig. 3-42. Overall, only 7.5 percent of the Earth is above 1 km, 3.2 percent is above 2 km, and 1.38 percent is above 3 km. The latitude restriction of ±40 deg of the Earth surface dictated by the above selection criterion 2) further restricts the landmass availability to 3.5 percent, 1.2 percent, and 0.76 percent respectively for altitude above 1, 2, and 3 km. Furthermore, geopolitical restrictions imposed by selection criterion 4), and the fact that the peaks are not regularly distributed in the Earth's landmass, greatly limit the availability of candidate sites for a global ODSN.

As a first approach to analyze the global availability of global peaks, we elaborated a digital topographic map of Earth with resolution of 2 km × 2 km. The Earth surface to be analyzed was restricted in the latitude interval [–40, +40] and longitude interval –180, 180]. Moreover, to better view the potential ODSN site distribution, we divided the Earth altitude in three interval ranges as 0–1 km, 1–2 km, 2–3 km, 3–4 km, and higher than 4 km. Results of this altitude level division of the Earth surface are presented in Fig. 3-43. Again, one should notice that only 7.5 percent of the Earth is above 1 km, 3.2 percent is above 2 km, and 1.38 percent above 3 km when considering the entire globe (solid curve). Restricting the available landmass within latitude +40° (dashed curve), we have only 3.5 percent, 1.2 percent, and 0.76 percent of Earth above altitudes of 1, 2, and 3 km, respectively.

Analyzing the Earth elevation map in Fig. 3-43, provides a number of useful indications for the construction of an ODSN. For instance, if the LDOS design approach is going to be taken for the global ODSN architecture, there is large area of Earth, mainly defined by the Pacific Ocean that lacks available peaks. In that case, a sure stop for a station in the LDOS must to be Hawaii, where incidentally there are already a number of astronomical telescopes housed on high-altitude peaks (e.g., Mauna Kea and Mount Haleakala). At the same time, Australia (where incidentally there is already a DSN radio-antenna complex), is relatively poor in high-altitude areas. Mainly, these locations, all within in the first range of 1–2 km of elevation, are concentrated in the center of the continent (Alice Springs) or close to the east coast of the continent. This

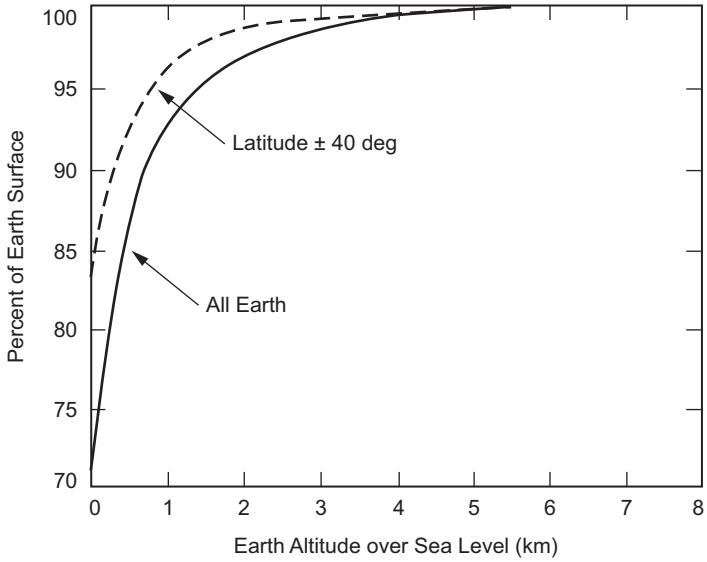


Fig. 3-42. Cumulative altitude distribution function of Earth's landmass.

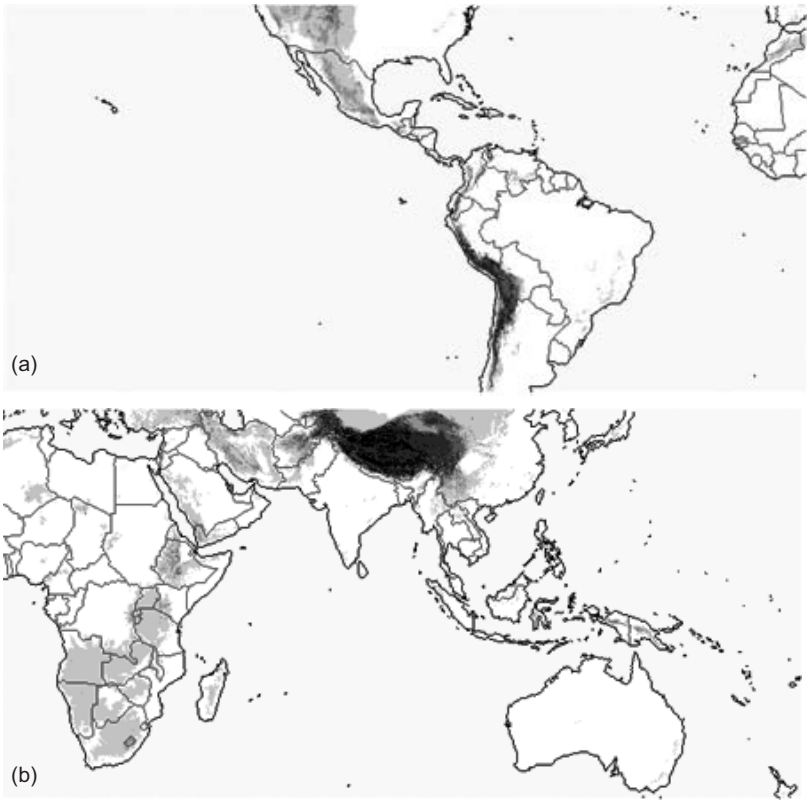


Fig. 3-43. Depiction of Earth landmass altitude in the latitude range of [–40 deg, 40 deg] at different ranges, as 0–1 km (white), 1–2 km (light gray), 2–3 km (gray), 3–4 km (dark gray), more than 4 km (black): (a) Earth map in the longitude range [–180 deg, 0 deg] and (b) Earth map in the longitude range of [0 deg, 180 deg].

scarcity of peaks in Australia can hamper the possible design of COS with a possible hub in this continent as previously suggested in the literature [28]. However, a more definitive answer to this last problem can come only after a careful evaluation and measurements of the sky background radiance and atmospheric transmission at candidate sites in Australia. However, as also stated by selection criteria 6), the altitude of the station of the ODSN, is not the only atmospheric/environmental requirement.

The location of a network station must be in an area where the cloud coverage has minimal impact on the operation of the network itself. Therefore, a more powerful indication of the site suitability for belonging to the ODSN can be made after simultaneously considering cloud coverage statistics of the area and peak availability. To better explain the study approach for joint correlation of low cloud coverage and higher altitude peaks, we first present in

Fig. 3-44 the cloud coverage in the section of Earth of interest for the ODSN using data from the International Satellite Cloud Climatology Project (ISCCP) [33].

The Earth map in Fig. 3-44 is within latitude range [–40 deg, 40 deg] and longitude range [–180 deg, 180 deg], and the map resolution is 2.5 deg × 2.5 deg. The figure indicates in grayscale-coded fashion the annual average of cloud coverage in percent in the region of Earth of interest for the ODSN. Clearly, for ODSN site selection, average cloud coverage duration must be as low as possible.

While the map in Fig. 3-44 may indicate regions of favorable cloud coverage for the installment of optical telescopes for deep space communication, it does not convey any information about the site/area altitude. However, one can further reduce the search for ODSN sites around the Earth, by introducing the simultaneous selection criteria of low cloud coverage (less than 50 percent) and altitude higher than 1 km (other more restrictive conditions about cloud coverage and altitude can also be used). Results from this last operation are shown in the maps of Fig. 3-45.

Figure 3-45 gives us more precise indications about the possible locations for ODSN and its possible architectural solutions. Starting from the Eastern Hemisphere as depicted in Fig. 3-45 a), beside Hawaii, other candidate areas are the United States Southwest and the Andes region (including northern Chile, southern Peru, and portions of Ecuador. Unfortunately, there is a lack of available sites east of these regions in both North and South America. Proceeding eastward, we can observe a number of candidate sites in the northern African continent and southern Spain. Southern Africa and eastern Africa (especially close to the horn of Africa) may also be regions of interest. A number of interesting regions are located in the Middle East, particularly in the Arabian Peninsula. Unfortunately, after the a region west of Pakistan and the Karakorum, moving eastward, (according to this first analysis) there is a great scarcity of peaks available in the map, except for the region around Alice Springs in the Australian Outback, and on the Australian east coast itself.

In conclusion, the approach just described is a practical methodology for the selection of potential sites for a global ODSN. In our approach, we first baselined a possible deep space mission and its requirements in terms of BER, link margin, data rate, and a few design figures (e.g., modulation, spacecraft, and ground telescope optical transmission). Then, to study link performance, it was supposed in our link scenario that we considered the worst case of optical signal interaction with Earth (i.e., 5 deg of SEP angle, 70 deg of observation zenith angle, and 2.4 AU range). It was demonstrated that at different altitudes on Earth, link performances differ greatly, and also that for different telescope apertures, there are different requirements for Earth altitude in order to successfully close the link. Next, we projected our study from single telescope

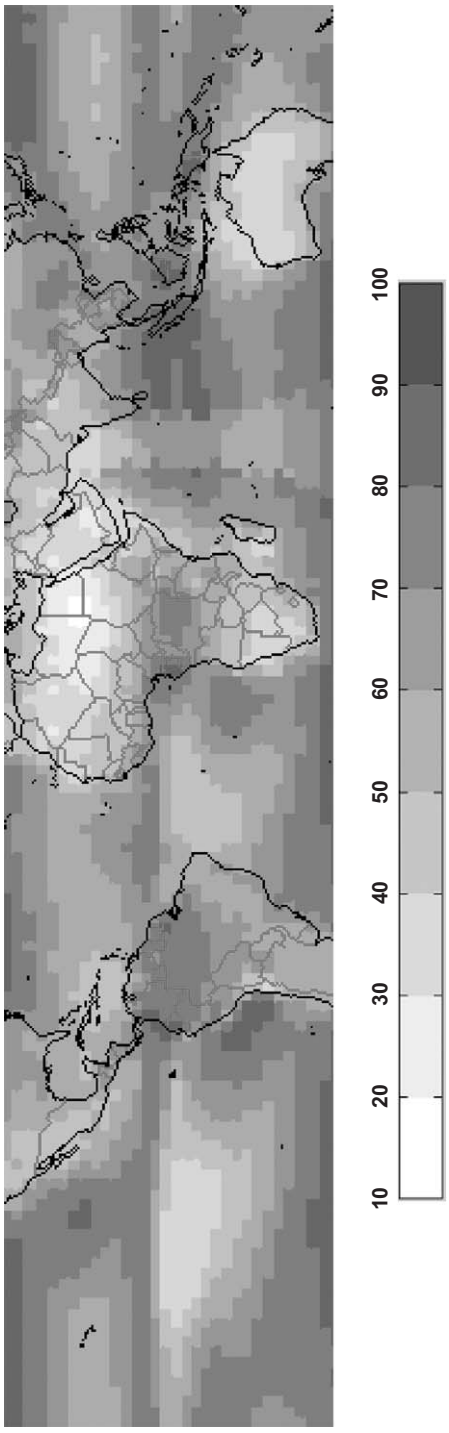


Fig. 3-44. Average annual percentage of cloud coverage for the Earth surface in the latitude range of [–40 deg, 40 deg]. The cloud coverage percent duration is indicated by the gray-scale per horizontal bar.

location to a global ODSN, and we demonstrated that high-altitude siting requirements, jointly with those of global cloud coverage, greatly restrict the landmass availability to house ODSN ground stations.

However, a number of issues must be further explored and amplified to provide a more precise answer to the problem of ODSN site selection. For instance, we limited the deep space mission requirements to a minimum 5 deg SEP angle separation. Conversely, to extend the duration of link coverage during a mission, the SEP angle requirement can be further reduced. Consequences of a smaller SEP angle separation will be a greater background sky radiance captured by the ground station and a greater noise in the receiver. A direct consequence of a noisier receiver is that our minimum altitude per station requirement will be raised, and fewer sites will be suitable for the ODSN use.

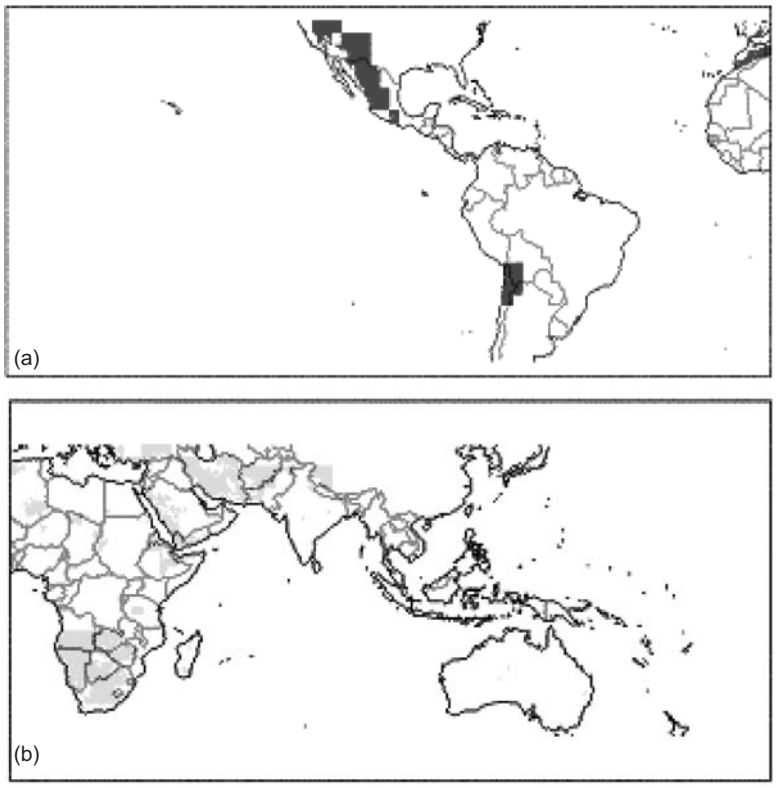


Fig. 3-45. Locations on Earth in the latitude range [–40 deg, 40 deg] that satisfy the simultaneous conditions of altitude higher than 1 km and average annual cloud coverage less than 50 percent: (a) Earth map in the longitude range [–180 deg, 0 deg] and (b) Earth map in the longitude range of [0 deg, 180 deg].

This work analysis can be further improved by providing a more precise model of the receiver channel. For instance, we did not consider effects of thermal noise in the receiver, or other noise factors derived by the detector dark counts. In other cases, a Poisson channel may be more representative of the receiver statistics. In any case, a precise modeling of the receiver channel can provide better information on the BER statistics and, therefore, the necessary signal photon flux that can satisfy the link requirements. Conversely (as previously demonstrated in this chapter), from these requirements we can derive the ground station diameter and/or the altitude of the ODSN stations.

Considering the meteorological activity of Earth, we introduced in our analysis a methodological approach to use global information on the cloud coverage. However, to further improve this analysis it is also necessary to consider diversity statistics [6]. In fact, to optimize Earth coverage, one of the operational principles of the ODSN is that at least two stations must be contemporarily seen by the spacecraft pointing towards Earth. In this case, the positioning of the stations with respect to each other cannot be done without considering weather diversity.

Finally, in this study we did not consider the action of atmospheric turbulences. It is known that Earth turbulence badly affects the signal both on the downlink and on the uplink in a number of ways. One of the most evident effects of turbulence is the spreading of the received signal focused on the photodetector with a consequence of net loss of power [34]. Again, to limit effects of atmospheric turbulences, the ground station should be located higher in altitude (i.e., less turbulent atmospheric path for the uplink and downlink signal), which again may further raise the threshold of minimum altitude for ODSN ground stations and the related available sites. Implications of atmospheric turbulence effects on the deep space downlink analysis and on the global ODSN need therefore to be included. A description of such effects is presented in the next subsection of this chapter.

3.4 Laser Propagation Through the Turbulent Atmosphere

3.4.1 Atmospheric Turbulence

As noted in the opening paragraphs of this chapter, laser beam propagation is perturbed by random refractive index fluctuations of the Earth’s atmosphere. Therefore, a plane wave arriving at the top of the atmosphere, from a deep-space-transmitted laser beam, undergoes phase distortions prior to incidence on a groundbased receiver. This subsection discusses the impact of these phase distortions upon laser communications. The discussion is initiated with a brief summary of relevant atmospheric properties. Following this description, the dominant effects on link performance (namely, turbulence-induced spatial and

temporal effects on the received irradiance) are discussed. Experimental results of laser beam propagation through horizontal atmospheric paths are also included, and even though these are not truly representative of deep space optical link configurations, they are indicative of the relevance of the theory. For gaining a deeper understanding of the problem, relating atmospheric turbulence to laser beam propagation the reader is urged to explore the vast amount of existing literature [35–40] on the subject. The following discussion is a very high-level summary extracted from the detailed discussions presented in these references.

The atmosphere can be thought of as a medium into which energy is injected in the form of thermally induced convection or wind shear. The upper scale at which this energy is dissipated is known as the outer atmospheric scale, commonly designated as, L_0. This outer scale is representative of the largest sized eddies involved in flows, that upon reaching a critical point break down to smaller scales by a cascading process. The critical point occurs due to the Reynolds number (Re) exceeding a value that demarcates laminar and turbulent flows. Re is defined as:

$$\mathrm{Re} = \frac{Vl}{n} \tag{3.4-1}$$

where V is the flow velocity, l is the scale size, and ν is the kinematic viscosity. For air, the kinematic viscosity is 1.5×10^{-5} m^2/s; therefore, if the product $V \times l$ exceeds 0.033, the critical Re (2200) that distinguishes between laminar and turbulent flows will be exceeded, hence the term atmospheric turbulence. This stepwise reduction in dissipation-scale or "eddy" size continues until an inner scale size, l_0 is reached. Below the inner scale size, energy dissipation occurs by viscous effects. The scale size bounded by L_0 and l_0 is referred to as the inertial scale, and it typically ranges from 10–100 m down to 0.1–1 cm.

The mathematical description of atmospheric turbulence, and in particular its effect on optical beam propagation, relies on idealized assumptions that treat the fluctuations of atmospheric parameters as stationary random processes that are homogeneous and isotropic. Within this mathematical framework, Kolomogorov [35,40] showed that the structure function follows a $r^{2/3}$ dependence, where r refers to the spatial scale defined as:

$$r = \left|\bar{r}_1 - \bar{r}_2\right| \tag{3.4-2}$$

with $\bar{r}_1$ and $\bar{r}_2$ referring to position vectors with $l_0 \le r \le L_0\, l_0 \le r \le L_0$. The structure function for a random variable $x(r)$ is defined as:

$$Dx(\bar{r}_1, \bar{r}_2) = \left\langle \left| x(\bar{r}_1) - x(\bar{r}_2) \right|^2 \right\rangle \tag{3.4-3a}$$

If the random variable $x(r)$ has a slowly varying mean with a superimposed fluctuation as represented by Eq. (3.4.3b)

$$x(\bar{r}_1) = \langle x(\bar{r}_1) \rangle + x'(\bar{r}_1) \tag{3.4-3b}$$

with the term in angle brackets denoting a slowly varying mean. With this definition of the random variable $x(r)$, the structure function can be rewritten as

$$Dx(\bar{r}_1, \bar{r}_2) = \left[\langle x(\bar{r}_1) \rangle - \langle x(\bar{r}_2) \rangle \right]^2 + \left\langle \left[\langle x'(\bar{r}_1) \rangle - \langle x'(\bar{r}_2) \rangle \right]^2 \right\rangle \tag{3.4-3c}$$

The first term in Eq. (3.4-3b) goes to zero for a stationary random process thereby emphasizing the merits of using the structure function in describing the fluctuations. The discussion above can be extended to any of the random variables associated with the atmosphere, such as velocity, temperature, and (of particular interest to optical propagation) refractive index $n(r, t)$

$$n(\bar{r}) = n_0 + n'(\bar{r}) \tag{3.4-4}$$

where $n_0 \cong 1$ represents the mean and $n'(\bar{r})$ represents random fluctuations. Furthermore, the long-term mean of the fluctuations is also equal to zero. The left-hand side of Eq. (3.4-4) can be related to visible and near-infrared optical wavelengths (λ), pressure $p(r)$, and temperature $T(r)$ by the relation:

$$n(\bar{r}) = 1 + 77.6 \times 10^{-6} \left(1 + 7.52 \times 10^{-3} \lambda^{-2} \right) \frac{p(\bar{r})}{T(\bar{r})} \tag{3.4-5}$$

Furthermore, the structure function for refractive index $D_n(r)$ can be expressed as:

$$D_n(r) = C_n^2 r^{2/3} \text{ for } l_0 \le r \le L_0 \tag{3.4-6a}$$

$$D_n(r) = C_n^2 r^{-4/3} r^2 \text{ for } r << l_0 \tag{3.4-6b}$$

The quantity C_n^2 is called the structure function constant of the refractive index, and mathematically it represents the slope of the structure function versus $r^{2/3}$ plot in the inertial range. The structure constant can be related to other structure function constants of temperature C_T^2 and velocity C_v^2, where

$$D_T(r) = C_T^2 r^{2/3} \text{ and } D_v(r) = C_v^2 r^{2/3} \text{ for } l_0 \le r \le L_0 \tag{3.4-6c}$$

The discussion so far has emphasized spatial fluctuations of atmospheric parameters; however, knowledge of temporal characteristics is also needed for

optical communications designers to devise strategies for mitigating temporal fluctuations. Atmospheric perturbations that contribute to temporal fluctuations owe their time scale to two types of effects, those associated with atmospheric eddies flowing past a fixed observer and those having to do with the dynamics within the eddy or atmospheric "cell." The former time scale is of the order of L_0 / V_T or D / V_T where V_T is the traverse wind speed and D is the diameter of the collecting aperture. For example, with the V_T = 12 m/s and D ranging from 5–10 m, the time scales of the order of 0.5–1 s are obtained. The dynamics within an atmospheric eddy or "cell" are thought to be of a longer duration, and the Taylor frozen-atmosphere hypothesis [40] is invoked to ignore this effect.

The preceding brief discussion of atmospheric turbulence introduced the refractive index structure function constant. Various moments of this function can be defined, and these in turn can be used to make first-order predictions of the effects expected on optical-link performance. The validity of these predictions will to a large extent depend upon the link configuration, as well as the location of the optical transmitter and receiver. Before initiating a discussion of these effects, however, a brief description of the methods followed to describe C_n^2 are in order.

Measurements of the refractive index structure function constant or C_n^2 can be classified into boundary-layer and free-atmosphere measurements. The boundary layer is the region close to the surface over which convective instabilities extend due to temperature differences. This region can extend from hundreds of meters to 2 km above the surface. Furthermore, it is dynamic, depending upon the time of day and the extent of solar heating. A good example of boundary layer C_n^2 measurements [41] shows diurnal variations with peaks at local noon, dips during the neutral durations close to sunrise and sunset, and a near-constant value at night. The variations have to do with the characteristics of heat transfer between the dry soil and surrounding atmosphere. One can well imagine the dependence on local terrain, vegetation, and prevalent wind speed that will influence this process. Thus, for sharp mountain peaks the boundary layer effects are not prevalent by virtue of the relief above the surrounding terrain. Thermosonde daytime measurements of the structure constant profiles as a function of altitude showed a –4/3 dependence of C_n^2 on altitude [42]. However, with winds that cause mixing in the boundary layer, deviations to the said dependence are observed. The daytime measurements of C_n^2 are particularly relevant to deep space optical communications links. The line-of-sight to deep space probes from Mars and all outer planets involves extended daytime durations. Thus, future operational optical communications systems must devise strategies to successfully mitigate daytime turbulence effects in order to be viable. In contrast with the boundary layer, the free-atmosphere layer involves the altitudes in the vicinity of the

tropopause (15–17 km) and higher altitudes. C_n^2 at these altitudes is less well known, though measurements have been reported [43].

Based upon the measurements, empirical and parametric models of C_n^2 have been derived, and a good description of these models is found in [39]. The Hufnagel-Valley (HV) parametric model is widely used. This model includes an upper atmospheric part of 3–24 km that applies to both daytime and nighttime and can be rewritten as:

$$C_n^2(h) = 8.2\times10^{-16} W^2 \left(\frac{h}{10}\right)^{10} \exp(-h) + 2.7\times10^{-16} \exp\left(\frac{-h}{1.5}\right) \tag{3.4-7}$$

Here, h is the height in kilometers, and W is the root mean square (rms) wind speed in meters per second in the range 5–20 km above the ground, specifically

$$W^2 = \left(\frac{1}{15}\right)\int_5^{20} V^2(h)\,dh \tag{3.4-8a}$$

where $V(h)$ is the wind speed in meters per second is also given by the Bufton wind model [40]:

$$V(h) = \omega_s h + v_g + 30\exp\left[-\left(\frac{h-9400}{4800}\right)^2\right] \tag{3.4-8b}$$

where ω_s is the residual slew rate between the spacecraft and ground that can be neglected for deep space probes where good tracking can be assumed. v_g refers to the ground wind speed. By adding a boundary layer term to Eq. (3.4-7) above the model can be extended to include the boundary layer effects. Therefore, the HV model takes the form:

$$\begin{aligned} C_n^2(h) = {} & 8.2\times10^{-16} W^2 \left(\frac{h}{10}\right)^{10} \exp(-h) + 2.7\times10^{-16} \exp\left(\frac{-h}{1.5}\right) \\ & + A\exp\left(\frac{-h}{0.1}\right) \end{aligned} \tag{3.4-9}$$

where A represents the ground level structure constant value or $C_n^2(0)$. The parameters A and W are adjustable in order to correspond to a desired value of "seeing" and isoplanatism, as explained further below. An alternate model for C_n^2 is the CLEAR 1 Night Model that can be extrapolated to the ground invoking the –4/3 height dependence to represent day time turbulence as

described earlier. Figure 3-46 below shows a comparison of the HV and 2 × CLEAR 1 models to represent daytime refractive index structure constants.

The moments of the refractive index structure constant allow estimates of parameters for predicting the influence of the atmospheric turbulence on optical link performance. These are described in the chapters that follow. The dependence of the moments derived from C_n^2 is a path integral, with the path extending from the height of the receiver above sea level to the top of the atmosphere. Therefore, the manner in which atmospheric turbulence influences space-to-ground propagation relative to ground-to-space propagation is different. Consider for a moment that a laser beam originating from deep space spreads to hundreds or thousands of kilometers and far exceeds atmospheric characteristic spatial scales, whereas for an upward propagating laser beam originating from an optical communications ground station and transmitted to space, the beam diameter is expected to be within the inertial range. In a qualitative sense, this causes the uplink beams to be steered by large angular displacements whereas the same is not true for downlink beams.

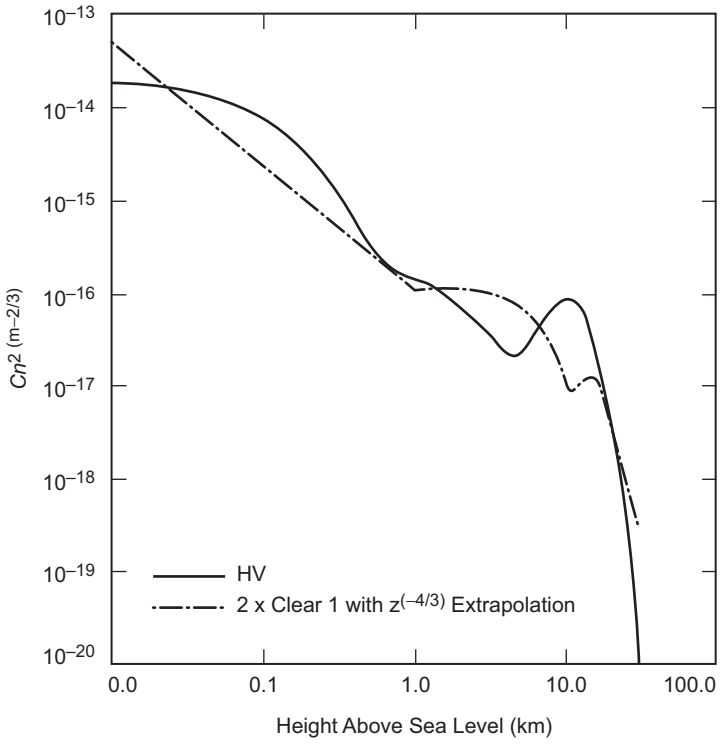


Fig. 3-46. Refractive index structure constant model derived using the Hufnagel-Valley and 2 x CLEAR 1 models. The CLEAR 1 model was extrapolated to the ground by a –4/3 dependence.

3.4.2 Atmospheric "Seeing" Effects

The atmospheric coherence diameter (also commonly referred to as the Fried parameter [44] or r_0 usually expressed in centimeters) is an important atmospheric parameter in determining optical link performance. Physically r_0 represents the spatial extent over which the phase of a propagating optical beam is preserved. For a plane wave, r_0 can be estimated from C_n^2.

$$r_0 = 1.67\left[\sec(\eta)k^2\int_{h_0}^{H} C_n^2(h)dh\right]^{-3/5} \tag{3.4-10}$$

where η represents the zenith angle of the observer, and k is the wave number defined as $2\pi/\lambda$ corresponding to the wavelength λ. This form of r_0 is applicable to the downlink beam. For the uplink beam, variations to this relation have been shown for the propagation of an assumed Gaussian beam intensity profile where r_0 increases with path length. As mentioned previously, the HV model parameters (such as A and W) can be adjusted in order to scale r_0 to desired values. Figure 3-47 shows the r_0 derived from the C_n^2 models discussed above, where the parameters A and W were adjusted to provide zenith r_0 values of 16 and 13.3 cm for the HV-models while the 2 × CLEAR 1 model was extrapolated to the ground by a –4/3 dependence on height yielded the smallest r_0 value of 9.5 cm at zenith.

These models provide a basis for predicting Fried parameters; however, as should be obvious from the discussions so far, the dependence on terrain, wind speed, topology, and solar illumination all together complicate the r_0 problem sufficiently to merit confirmation by measurement. Figure 3-48 shows a compilation of measured [39,45] r_0 values over mostly night, twilight, and dusk hours. These measurements were scaled to 1064 nm by the wavelength dependence indicated in Eq. (3.4-10). As shown in Figs. 3-48a and 3-48b, the r_0 measurements show very little dependence on altitude but show evidence of a relation to the relief.

An abundance of daytime measurements are found in the solar astronomy literature [46,47,48] where r_0 is extracted from a variety of solar irradiance and limb image motion measurements. Figure 3-49 shows a 2-year compilation of data measured at Sacramento Peak, New Mexico. The data are compared to the diurnal variation of r_0. A partial agreement is observed; however, note that the diurnal model is predicted for fully developed boundary layers.

One noteworthy difference between solar and stellar techniques for measuring r_0 is that the latter are usually not capable of making measurements in the vicinity of the Sun. Generally, stellar measurement systems are capable of observing daytime stars as close as 30 deg from the Sun in order to allow

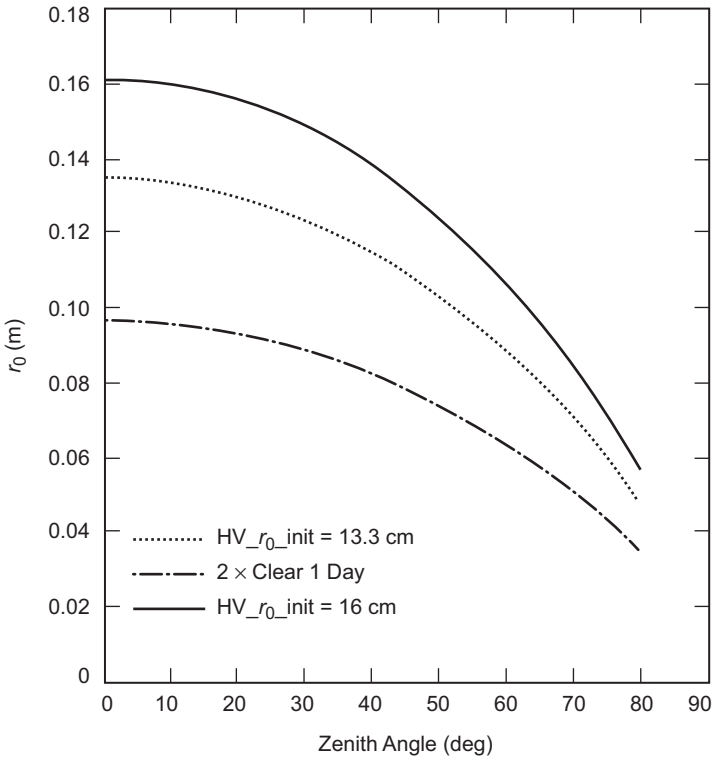


Fig. 3-47. Plots of r_0 versus zenith angle derived from the C_n^2 models discussed in 3.5.1. The Hufnagel-Valley models were scaled to provide zenith r_0 values of 13.3 and 16 cm, whereas the 2 × CLEAR 1 model extrapolated to the ground by the −4/3 dependence yields a zenith r_0 of 9.5 cm.

enough contrast in measured stellar images. Similarly solar observations are limited to direct observations of the Sun and cannot be used to look at zenith angles away from the Sun.

The beam perturbations associated with r_0 are also commonly referred to as atmospheric "seeing" effects with the "seeing being defined as $\sim \lambda / r_0$ expressed in angular units, or radians. Atmospheric "seeing" effects upon a deep space optical link are related to the fact that the diameter or Fried parameter presents a limiting aperture size to the laser beam wavefront. For a perfect optical collection system, the spot size in the focal plane to first order is determined by 2.44 $f\lambda/D$, with f and D representing the focal length and aperture diameter of the collection system. However, when the beam propagates through the atmosphere, D is replaced by r_0 so that first, the spot diameter at the focal plane is increased by the ratio D / r_0. This increase in focal

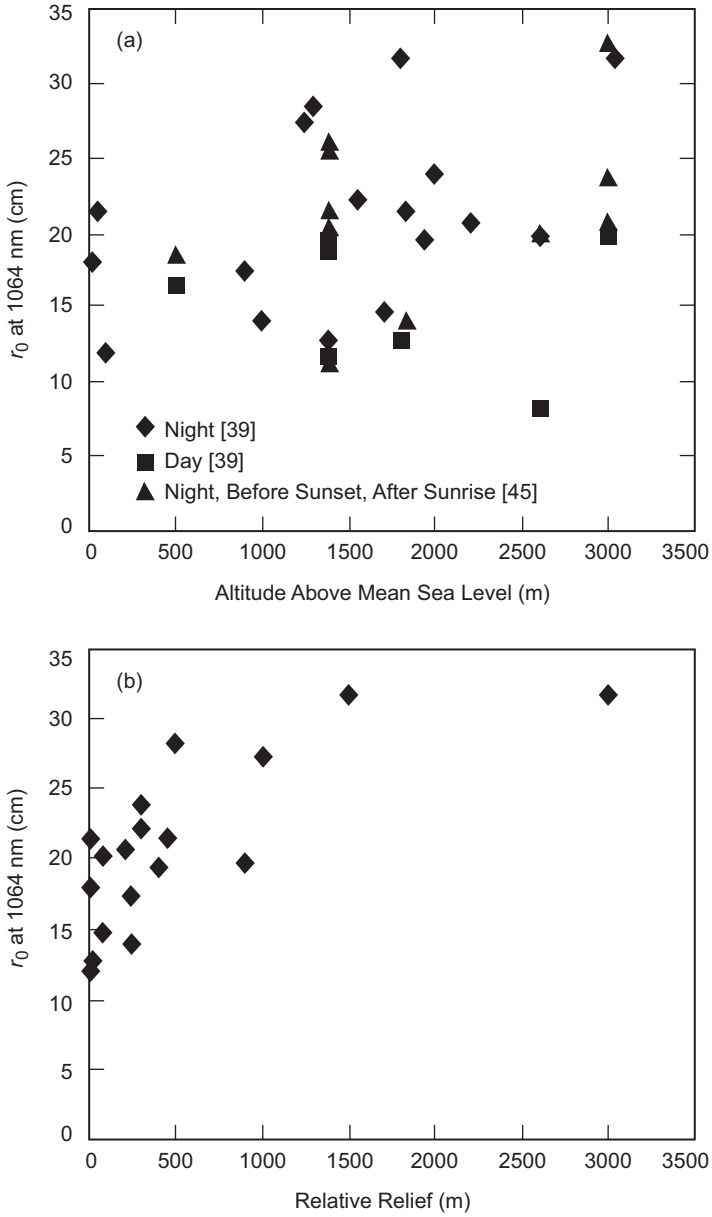


Fig. 3-48. (a) Distribution of r_0 values [39,45] as a function of height of location shows no correlation with altitude. (b) Measurements from [45] plotted as a function of the relief above the surrounding terrain where evidence of functional relationship seems apparent.

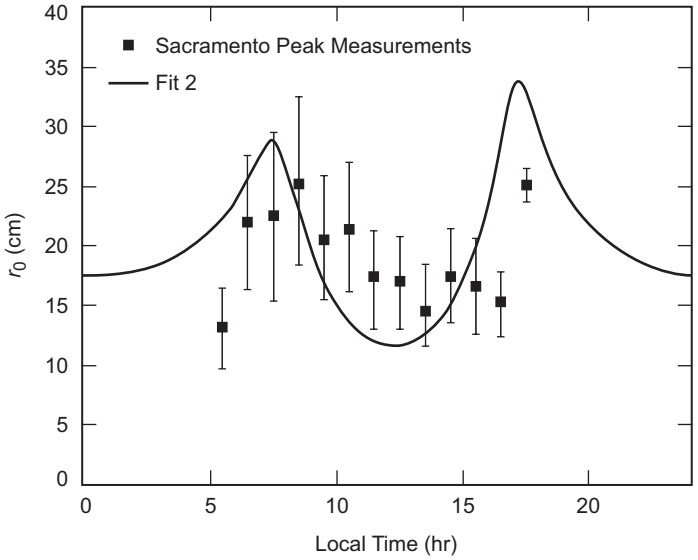


Fig. 3-49. Summary of r_0 obtained from solar observations [46], compared to the diurnal variation model. Note that the diurnal variation applies to measurements where the boundary layer is fully developed.

spot or atmospheric "blurring" increases the effective FOV required to collect all the laser signal photons. The two big disadvantages are (1) increased sky background noise contribution, especially for daytime links, and (2) increased detector area that could be bandwidth limiting. The latter problem is not insurmountable since detector sizes satisfying the FOV and bandwidths to support 10-Mbps class links are available. However, the former problem cannot be circumvented without the use of an adaptive optical system. Using adaptive optical systems in the daytime, very likely with an artificial guide star because the communications laser signal from the spacecraft will be too weak, poses a formidable challenge [49]. Conversely, being able to accommodate the turbulence-degraded FOV, since imaging is not required for communications, could greatly simplify matters. Alternatively, adaptive signal processing techniques using array detectors is a viable approach discussed later in Section 6.2.2. Detectors with adequate collection area do not suffer from turbulence in the nighttime when background light levels are sufficiently low. Furthermore, the background light penalty is really a product of the prevalent sky radiance and the "seeing" limited FOV or solid angle. Thus, if at large Sun-separation angles "seeing" is poor, the penalty on the optical communications system may not be very large; however, low Sun-separation angles combined with poor seeing present the worst case for the optical link performance.

Designs that address the worst-case condition will provide much better performance at other times. To cite an example, consider a spacecraft orbiting Mars. The farthest range between Earth and Mars occurs at solar conjunction when optical systems must communicate at Sun-separation angels as low as 3 deg in order to limit the outage to 30 days. Here of course, the spacecraft and the Sun rise at approximately the same time, and the sky radiance can be very high. If the ground-based receiver is operating under fully developed boundary layer conditions with worst seeing, performance is severely impeded. Furthermore, the time of day when the spacecraft rises has to be factored in while evaluating performance.

Next, some analysis is presented to further elaborate on the spot blurring due to atmospheric turbulence. Consider an ideal annular light-collection system with diameter D and obscuration ratio ε. The mean fraction $P(r)$ of the incident energy collected within the normalized detector radius $r \equiv a / \lambda F$, (a is the detector radius, and $F \equiv f / D$) is given, by the approximate relation [50,51]

$$P(r) = 2\pi r \int_0^1 \langle \tau(\rho,\varepsilon) \rangle J1(2\pi r\rho) d\rho \tag{3.4-11}$$

where $\langle \tau(\rho,\varepsilon) \rangle$ represents the time-averaged optical transfer function (OTF) of the annular optical system corresponding to a spatial frequency ρ with a linear obscuration ratio, ε and $J1$ is the first-order Bessel function of the first kind. The time-averaged OTF can further be expanded

$$\langle \tau(\rho,\varepsilon) \rangle = \tau_1(\rho,\varepsilon)\tau_a(\rho) \tag{3.4-11a}$$

Here $\tau_1(\rho,\varepsilon)$ is the OTF of the aberration free annular pupil, and $\tau_a(\rho)$ represents the effect of turbulence.

$$\tau_a(\rho) = \exp\left[-3.44\left(\frac{D\rho}{r_0}\right)^{5/3}\right] \tag{3.4-11b}$$

where a Kolmogorov model of atmospheric turbulence has been invoked.

$$\tau_1(\rho,\varepsilon) = \tau_1(\rho) + \varepsilon^2 \tau_1(\rho,\varepsilon) - \tau_{12}(\rho,\varepsilon) \tag{3.4-11c}$$

with

$$\tau_1(\rho) = \frac{2}{\pi\left(1-\varepsilon^2\right)}\left[\cos^{-1}\rho - \rho\left(1-\rho^2\right)^{1/2}\right], 0 \le \rho \le 1 \tag{3.4-11d}$$

$$\tau_{12}(\rho,\varepsilon) = \frac{2\varepsilon^2}{1-\varepsilon^2}, 0 \le \rho \le \frac{1-\varepsilon}{2}$$

$$\tau_{12}(\rho,\varepsilon) = \frac{2\varepsilon^2}{1-\varepsilon^2}\left[1 - \frac{1+\varepsilon^2}{2\pi\varepsilon^2\beta} - \frac{1}{\pi\varepsilon}\sin\beta + \frac{1-\varepsilon^2}{\pi\varepsilon^2}\tan^{-1}\left(\frac{1+\varepsilon}{1-\varepsilon}\tan\frac{\beta}{2}\right)\right],$$

$$\frac{1-\varepsilon}{2} \le \rho \le \frac{1+\varepsilon}{2} \qquad \text{(3.4-11e)}$$

$$\tau_{12}(\rho,\varepsilon) = 0,$$

otherwise

$$\beta = \cos^{-1}\left(\frac{1+\varepsilon^2-4\rho^2}{2\varepsilon}\right)$$

The ratio D/r_0 determines the quantity r required to collect a desired fraction of the signal energy where *r* is measured in units of λF. Adapting the daytime models represented in Fig. 3-47, where zenith-viewing r_0 values ranged from 9.5–16 cm and 5–8.5 cm for zenith angles of 70 deg. Table 3-8 below gives the approximate corresponding D/r_0 values that can be expected for a number of representative ground antennas.

Figure 3-50 shows a plot of the integral given by Eq. (3.4-11) for a number of the aperture diameters spanning the range of Table 3-8. Included in Fig. 3-50 is a plot for $D/r_0 \to 0$, or $r_0 >> D$, meaning no atmospheric turbulence with a diffraction-limited ideal optical system. So the curve with $D/r_0 \to 0$ and $\varepsilon = 0$ is simply the Airy pattern plot from diffraction theory that shows 84 percent of the energy corresponding to the dimensionless spot size of 1.22 in units of $1/\lambda F$, as expected. As the D/r_0 increases, the long term average spot size for encircling a given fraction of energy also increases as shown. Figure 3-50 shows that the dimensionless spot size required to encircle 84 percent of the energy can be approximated by the quantity D/r_0. This approximation gets

Table 3-8. The D/r_0 ratios for a number of large-aperture telescopes.

	D (m)	Obscuration (m)	D/r_0 Best Day	D/r_0 Worst Day
Hypothetical	10	1.4	62.5	200
Palomar	5	1	31.25	100
AEOS, Mt. Haleakala*	3.67	0.86	23	73.4
TMF	1	0.2	6.25	20

*AEOS = Advanced Electro-Optical System

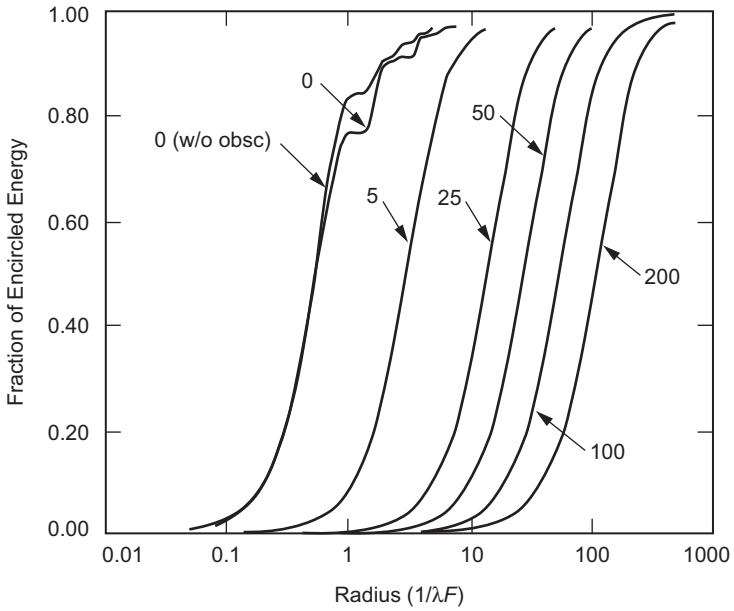


Fig. 3-50. Fraction of encircled energy versus radius at focal plane in units of 1/λ*F*, λ being wavelength and *F* being the focal ratio of the optical system with a circular aperture of diameter *D* with a central obscuration 20 percent of *D*. The labels indicate the D/r_0 ratio, with $D/r_0 = 0$ representing the diffraction-limited case. For comparison, a diffraction-limited unobscured circular aperture corresponding to an ideal Airy pattern is also included.

better as the D/r_0 increases For example, $D/r_0 = 200$ (the percent error in dimensionless spot size) is <3 percent, while for $D/r_0 = 5$ the percent error in dimensionless spot size is 15 percent. Thus, the FOV required for encircling 84 percent of the energy or the ratio of the detector diameter to the focal length is given by:

$$FOV_{84\%} \approx \frac{2\dfrac{D}{r_0}\lambda F}{f} \approx 2\frac{\lambda}{r_0} \tag{3.4-12}$$

where *f* is the focal length of the system, and the approximation confirms the simple assertion that the half-angle FOV required to collect on an average 84 percent of the signal energy is the atmospheric "seeing" λ/r_0. Thus, the more severe the turbulence and the smaller the r_0, the larger the FOV and corresponding solid angle required to gather 84 percent of the energy.

The discussion presented above for the mean encircled fraction of energy = 84 percent can be applied for any spot size along the abscissa in order to convert it to FOV, which in turn can be converted to solid angle in steradians using the relation:

$$\Omega = 2\pi\left(1 - \cos\frac{\Theta}{2}\right) \tag{3.4-13}$$

where, Θ is the FOV in radians. The conversion of the abscissa to FOV or solid angle will depend upon the focal ratio of the light collection system. Figure 3-51 shows an example light-collection system with a 5-m diameter collection aperture, a 20-percent obscuration, and a focal ratio of 16. The lower the focal plane loss (FPL) or the larger the fraction of encircled energy, the larger the solid angle. Imagining a vertical line through the plotted points shows that a 1.75-dB change in FPL results in a 5–7-dB change in solid angle.

For daytime optical links relying on groundbased reception, the dominant source of background is the sky radiance. Furthermore, the detected background noise is directly proportional to the solid angle. The significance of the plot of Fig. 3-51, therefore, is that by sacrificing a relatively small fraction

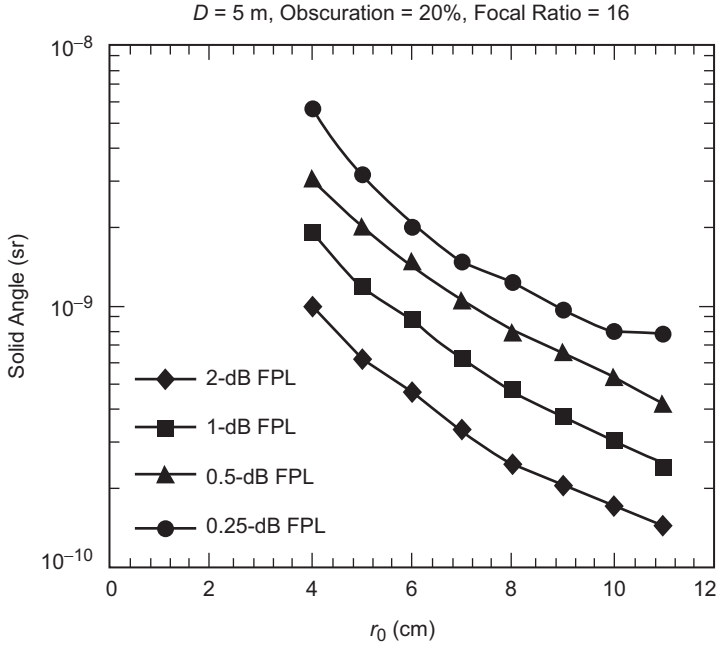


Fig. 3-51. Dependence of solid angle upon the Fried parameter r_0 for collecting 63, 80, 89, and 95 percent of encircled energy or equivalently a 2-dB focal plane loss (FPL), 1-dB, 0.25 dB, and 0.5 dB FPL.

of signal a much larger amount of background can be reduced, thereby increasing the overall channel capacity of the link. Thus, in addition to using optical filtering and polarization discrimination for rejecting background photons incident on the detector, tailoring the detector solid angle provides an additional means of discriminating against background photons. At night, on the other hand with background light levels being relatively low (except when a planet or celestial object also intersects the detector FOV), maximizing the solid angle appears to be optimal since this allows maximum signal collection with a negligible increase in background.

3.4.3 Optical Scintillation or Irradiance Fluctuations

Perturbations of the turbulent atmosphere on received laser beams cause interference so that a point detector placed in the path of the beam will alternately see patches of constructive and destructive interference causing swings in the detected irradiance. These irradiance fluctuations are characterized by a scintillation index (SI), σ_I^2 defined as:

$$\sigma_I^2 = \frac{\left\langle I^2 \right\rangle - \left\langle I \right\rangle^2}{\left\langle I \right\rangle^2} \tag{3.4-14}$$

where $\left\langle I \right\rangle$ represents the ensemble averaged irradiance of the received laser signal. As Eq. (3.4-3) shows, the SI is the variance of the irradiance normalized by its mean. The scintillation index can be expressed as:

$$\sigma_I^2 = \exp\left(4\sigma_\chi^2\right) - 1 \tag{3.4-15}$$

where σ_χ^2 is the variance of the log amplitude that can be expressed in terms of C_n^2 by the relation:

$$\sigma_\chi^2 = 0.56k^{7/6}\sec(\eta)^{11/6}\int_0^L C_n^2(h)(L-h)^{5/6}dh \tag{3.4-16}$$

The relation above applies well to the downlink communications from deep space. However, for uplink, more exact Gaussian beam formulations presented in reference [40] can be used. These estimations are reasonable as long as the weak turbulence approximations are valid. For severe turbulence or large airmass, SI does not continue to increase monotonically. Saturation eventually sets in, and the SI can be estimated [52] by heuristic models.

It is important to recognize that the irradiance fluctuations predicted by Eq. (3.4-16) and its variations apply to a point detector. For viable photon-

starved optical communication links from deep space, relatively large collection apertures (5–10 m) will be utilized for gathering a sufficient number of detectable photons. As a result, the speckle features distributed over the aperture and undergoing random constructive and destructive interference average out to a large extent. Aperture averaging theory [53] predicts a factor A for a plane wave traversing a horizontal atmospheric path:

$$A = \left[1 + 1.062\left(\frac{kD^2}{4L}\right)\right]^{-7/6} \tag{3.4-17}$$

Thus, the irradiance fluctuations or σ_I^2 predicted for a point detector will be reduced by the factor A for a collection aperture diameter D and wave number $k = 2\pi / \lambda$, λ being the wavelength and L being the path traversed.

Figure 3-52 shows the dependence of A upon L for a few different values of D. It is apparent that the larger the diameter the greater the impact of aperture averaging in reducing the irradiance fluctuations or scintillation.

While one can "translate" a horizontal atmospheric path with assumed fixed turbulence structure or C_n^2 into an equivalent slant or zenith path airmass, this approach to predicting the effects of aperture averaging on a downlink signal is

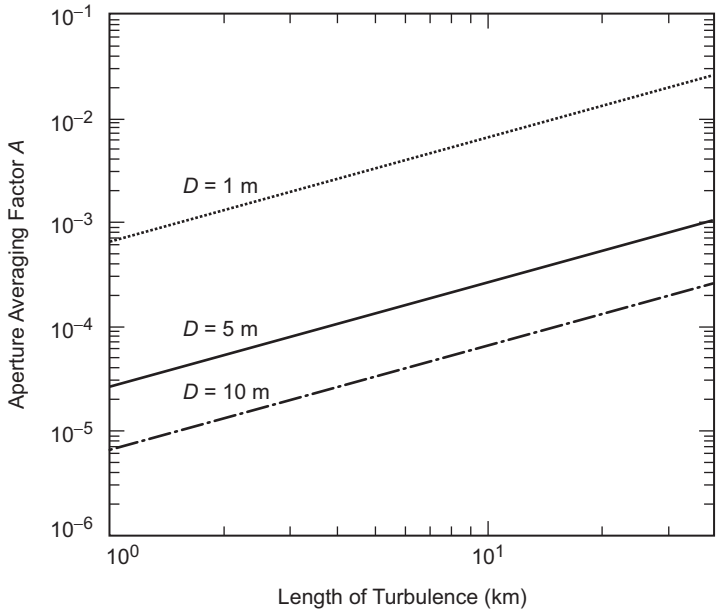


Fig. 3-52. Aperture averaging factor dependence on horizontal atmospheric path length for a number of different aperture sizes.

somewhat speculative. An expression was derived for the space-to-ground aperture averaging factor [54] that allows taking into account the C_n^2 profile and is given by:

$$A = \frac{1}{1 + A_0^{-1}\left(\frac{D^2}{\lambda h_0 \sec\theta}\right)^{7/6}} \tag{3.4-18}$$

where θ represents the zenith angle, $A_0 \approx 1.1$, and h_0 is given by

$$h_0 = \left[\frac{\int_{\text{path}} dh C_n^2(h) h^2}{\int_{\text{path}} dh C_n^2(h) h^{5/6}}\right]^{6/7} \tag{3.4-19}$$

This relation takes into account the slant range through the atmosphere and permits the modeling of the atmospheric refractive index structure function. Figure 3-53 shows the dependence of the aperture-averaging factor for a number of different aperture diameters using the approximation represented by Eq. (3.4-18). Surface C_n^2 is given by $A = 10^{-13}$ for an observer located at sea level with a 4-m/s upper-atmospheric wind speed. Changing the altitude did not provide any significant improvement in *A*, and the sea-level estimation below can be used as the worst case. Note that this prediction provides additional improvement in the aperture-averaging factor and is more representative for a true downlink path.

The net result of the predictions is that for downlink signal reception, the large collection-aperture size needed to gather sufficient photons also provides a large mitigation of the irradiance fluctuations experienced by the detector. The reduction in normalized variance of the irradiance has also been analyzed [55] as a function of the ratio of the aperture diameter D/r_0 and shown to decrease monotonically as *D* increases relative to r_0.

Comparisons of experimental measurements with theory are shown in Fig. 3-54. The measurements were performed over a 45-km mountain-top-to-mountain-top horizontal atmospheric path at an average altitude of 2 km. In Fig. 3-54, the solid lines were obtained by using theoretical expressions [40] for a plane, spherical wave propagating through turbulence. The data represented by the diamonds was obtained [56] by limiting the pupil at the Coude aperture plane of a 0.6-m telescope at Table Mountain, California, while transmitting an 811-nm laser beam with a 250-μrad beam divergence from Strawberry Peak near Lake Arrowhead, California. The triangles were obtained [57] during a 150-km laser-communication link demonstration between Haleakala and

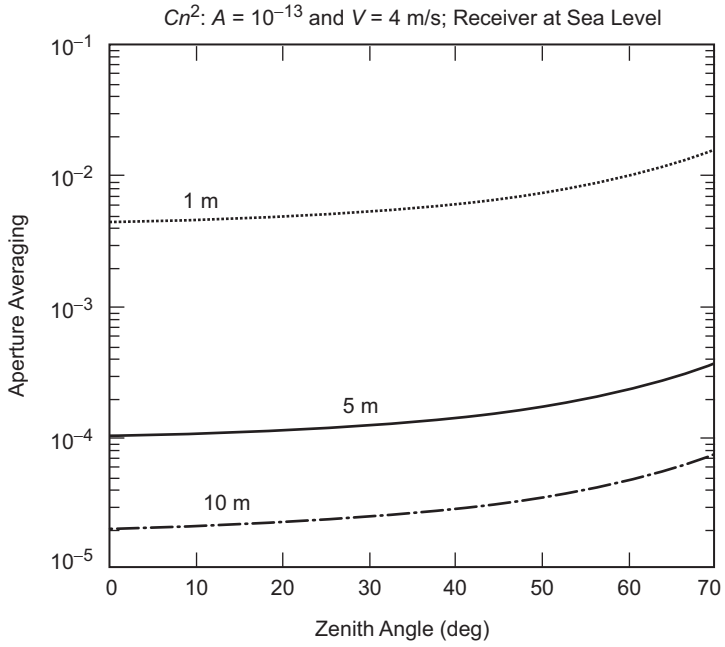


Fig. 3-53. Aperture averaging factor dependence on horizontal atmospheric path length for 1-m, 5-m, and 10-m aperture sizes for a receiver at sea level, with 4 m/s upper-atmospheric wind speed, and *A* of 10^{-13}.

Mauna Kea in Hawaii. The data labeled SP1998 (open squares) were obtained [58] using an 8-cm spotting telescope at Strawberry Mountain, California, while transmitting a 780-nm laser to Table Mountain, California. Finally, the data labeled TMF 1998 and Aug/Sep 2000 were both obtained [58, 59] using a narrow beam-divergence (22-μrad) beam transmitted from Strawberry Mountain, California and received at Table Mountain, California using the full aperture of the 60-cm telescope at Table Mountain. All the data are presented as a plot of the aperture averaging factor versus the Fresnel number. Measurements suggest that the aperture averaging achieved for the horizontal paths are generally better than theory predicts. The wide spread in the TMF 1998 and August–September 2000 factors is partially attributed to the narrowness of the beam so that the beam footprint at Table Mountain just overfills the telescope (0.9-m beam footprint versus 0.6-m aperture) and beam jitter contributes to additional irradiance fluctuations.

An interesting approach being considered for providing large effective aperture on the ground is to use an array of smaller telescopes [60]. In this approach, each individual smaller receiving aperture will have greater

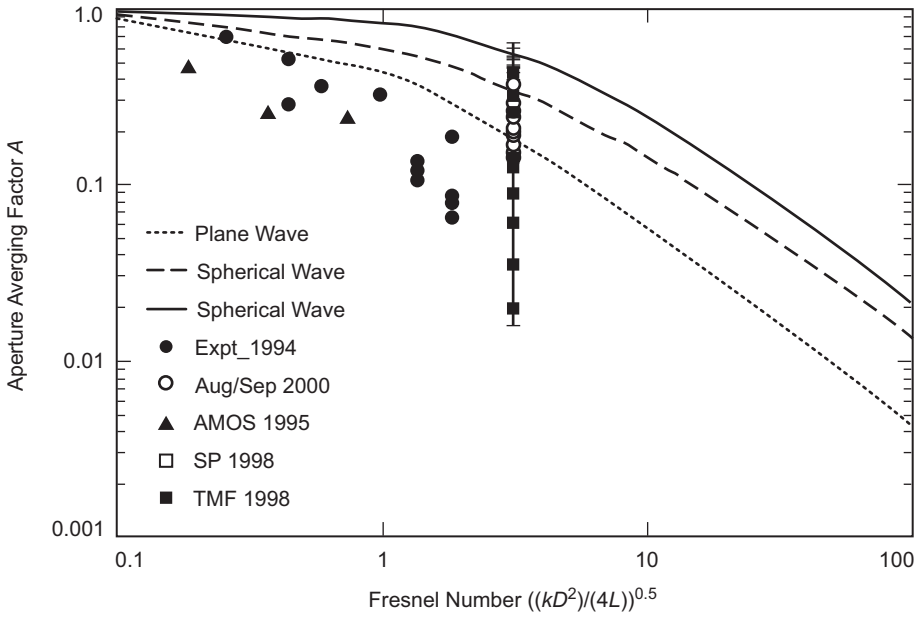


Fig. 3-54. Comparison of some horizontal path aperture averaging measurements compared to theory.

irradiance fluctuations; however, the signal-combining schemes devised for arrayed collection will average signal in the electronic domain.

Irradiance fluctuations must be mitigated on uplink lasers used for optical communications whether their purpose is to serve as a pointing reference for sending command data. At the time of writing this text, the experience with uplink lasers to deep space was very limited, with the Galileo Optical Experiment (GOPEX) [61] being the only reported demonstration. Here pulsed lasers were received by the Solid State Imaging (SSI) camera during the Galileo spacecraft's cruise to Jupiter. During this experiment, uplink scintillation of the laser beam was verified, and a nominal SI value of 0.69 was obtained over several days of transmission. This value of SI fell between those predicted by the weak-turbulence theory and the strong turbulence theory. Limited camera dynamic range coupled with uncertainties in synchronizing the laser pulses with camera framing limited data analysis and fitting measurements to the expected lognormal statistics. Furthermore the fades sensed by the camera cannot unambiguously be associated with scintillation since attitude variations of the spacecraft, and atmospheric turbulence induced beam wander (discussed below) could have contributed to the fades.

Uplink scintillation or normalized variance σ_I^2 was measured for the Low Power Atmospheric Compensation Experiment [62] (LACE) satellite where

retro-reflected links with σ_I^2 values ranging between 0.08 and 0.2 were measured. Furthermore, the temporal power spectral density showed 30 times less contribution at frequencies greater than 100 Hz compared to 1 Hz. The Relay Mirror Experiment [63] (RME) also used retro-reflectors on a near-Earth orbiting satellite to establish optical links. Scintillation indices, σ_I^2, ranging from 0.09 to 0.67 at frequencies ranging from 25 to 400 Hz were reported. The increase in higher frequency contributions in the latter demonstration is noteworthy, though no space-platform jitter results were provided in either of the two retro-reflector space-to-ground links cited.

A geostationary Earth orbit (GEO) laser communications demonstration (GOLD) reported [64,65] scintillation indices of 0.18 or better. GOLD also demonstrated that, by using multiple mutually incoherent beams, the scintillation could be reduced due to an averaging out of the fluctuations. Thus for GOLD, a four-beam 514.5-nm multi-beam uplink was demonstrated that showed a 1/N reduction in scintillation for an N-beam beacon. During the GOLD experiment, the SI values reported were 0.12 for two beams and 0.045 for four-beams.

Figure 3-55 presents a graphical summary of SI values measured during a horizontal path experiment [59]. A reduction in SI with increasing number of co-propagating beams is observed for all three nights that observations were conducted. Overall average reduction factors of 3, 3.2, and 3.7, respectively, were achieved for June 28, 29, and 30, respectively. In results reported elsewhere [66], over a 5.4-km range with nine co-propagating argon laser beams, the predicted reduction was shown to depend on the severity of atmospheric turbulence or Rytov variance. The beam-propagation simulation reported indicates as much as a 10-fold reduction for Rytov variances between 0.4 and 0.5 and a 3-fold to 4-fold reduction for Rytov variance close to 0.2.

From a systems design standpoint, the statistics of the irradiance fluctuations seen by a point detector for either the uplink or downlink obey lognormal statistics, notwithstanding deviations that may be encountered under strong turbulence. However, what is the nature of irradiance fluctuations with the mitigation strategies discussed above (namely, aperture averaging for the downlink and multi-beam averaging for the uplink)? The central limit theorem would dictate that, with enough averaging, the fluctuations should reduce to Gaussian; however, the log-normal distribution is somewhat unique in this sense and displays slow convergence [67, 68].

Consequently, assuming that the residual scintillation-induced fade statistics for an optical communications receiving system is log normal, whether in space receiving a multiple beam or on the ground receiving through a large aperture, may be adequate for design purposes.

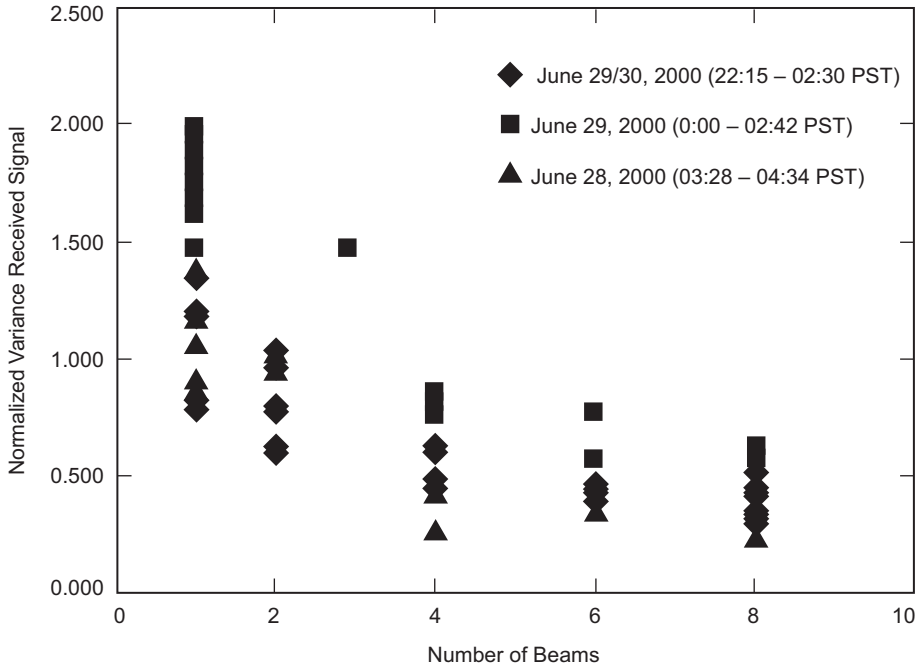


Fig. 3-55. Reduction in scintillation index as a result of multiple-beam incoherent averaging.

3.4.4 Atmospheric Turbulence Induced Angle of Arrival

The atmospheric turbulence induces distortions on the laser beam wavefront. For astronomical imaging applications, adaptive optics systems are being increasingly used to reconstruct the wavefront. For deep space optical communication, this may be a distinct possibility in the future; however, the present discussion is limited to systems that do not rely on an adaptive optical system. One of the consequences of the atmospheric distortion is "image motion" in the focal plane, and this is caused by fluctuations in the angle of arrival. In considering the effect of angle-of-arrival fluctuations on deep space optical communication links, a distinction between the downlink and uplink is needed.

The downlink is collected through a large effective aperture. Furthermore, it was pointed out in 3.4.2 that the detector FOV is restricted in order to reject excessive background from sky radiance and scattering. Imagine, therefore, the spot encircling the signal energy at the focal plane moving around so that at any instant a fraction of the encircled energy is blocked by the field stop. The situation described is shown in Fig. 3-56 where the circle represents the inner edge of the field stop, and the blurred spot represents the long-term averaged

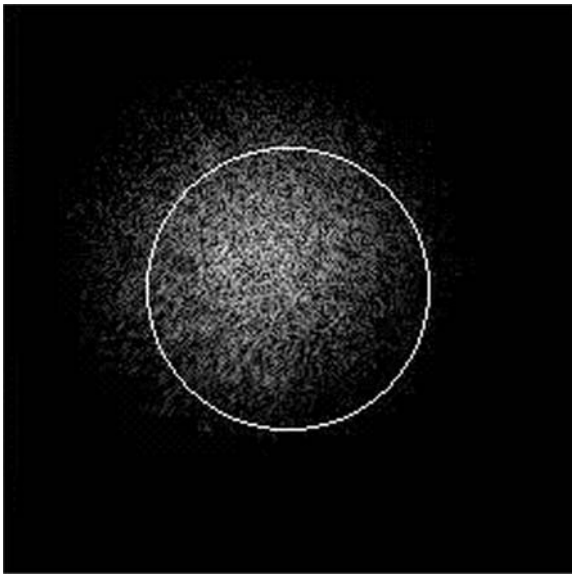


Fig. 3-56. Spatial distribution of mean laser energy averaged over many pulses with respect to the detector field-of-view depicted by the circle.

spatial distribution of the signal. A net loss of signal, with degradation of the optical link performance, will result. The situation can be greatly improved by having a simple correction system comprised of a two-axis fast-steering mirror located at the pupil image plane to actively compensate for the atmosphere-induced angle-of-arrival fluctuations. Note that correction for tilt does not compensate for higher order aberrations in the received beam; however, for optical-communication links, it is adequate since collecting the maximum possible signal energy is the intent. The implications of not having to correct for higher order aberrations is that a much lower bandwidth tracking system can be implemented.

A rather simplified form of the angle-of-arrival fluctuations that is independent of the wind profile is given by [40]

$$\sigma_\beta^2 = 2.914\left[\int_{h_0}^{H} C_n^2(h)dh\right]D^{-1/3}\sec(\eta) \qquad \text{(3.4-20a)}$$

where σ_β represents the rms angle of arrival jitter. This can be further simplified using Eq. (3.4-10) to yield:

$$\sigma_\beta^2 = 0.170\left(\frac{\lambda}{D}\right)^2\left(\frac{D}{r_0}\right)^{5/3} \qquad \text{(3.4-20b)}$$

With an r_0 of 10 cm and a wavelength of 1.064 μm, for example, the atmosphere-induced rms jitter, σ_β, will be approximately 2 μrad for a 5-m to 10-m collection aperture diameter, D, increasing to 3 μrad for D decreasing to 1 m. For example, a collection system with a 5-m diameter and an effective focal length of 80 m (i.e., an F ratio of 16 will result in the blur diameter required to enclose 84 percent of the energy of 1.7 mm according to the relations used in Eq. 3.4-12) above. At the same time, the angle-of-arrival fluctuations will cause rms spot displacements of 160 μrad. Therefore, the peak displacement (assuming a Gaussian distribution of the angle-of-arrival fluctuations) can be ± 0.48 mm, which is a significant fraction of the detector size to collect 84 percent of the energy. The need for tracking out the angle of arrival jitter or tilt error using a two-axis fast steering mirror becomes a requirement for reasonable performance. The photon-starved nature of the link may not readily allow for providing adequate signal-to-noise ratio at the desired update rate [70] required to compensate for the tilt. This problem can be further exaggerated by the presence of large amounts of background. Therefore, the design drivers here are methods of sacrificing minimal amounts of communication signal, while at the same time devising a means of getting an error signal derived from the mean spatial distribution of photons in the focal plane. This strongly suggests a class of photon-counting detectors that can be configured as a quadrant or array of detectors.

On the uplink, the atmosphere-induced beam wander can be conceived of as degrading the Strehl of the laser beam transmitted so that the on-axis energy would undergo fluctuations even if the beacon were pointed correctly at the distant spacecraft. By uplinking multiple beams as described earlier, some averaging of the random beam steering due to the atmosphere can in principal be realized, though quantifying this without running a simulation is not possible and is beyond the scope of the present discussion. The fades on the uplink are a combination of the residual scintillation and beam wander effects after multi-beam averaging, combined with superimposed irradiance fluctuations from receiver platform attitude variations and vibrations. The update rates at which the uplink beacon is sampled for command data or for the purposes of a pointing reference have a large effect on beacon design. One of the scenarios for pointing an uplink is to blind point to the spacecraft based on ephemeris predicts.

In this section an attempt was made to provide a brief introduction to the impact of the atmosphere on a groundbased optical-communications receiving and transmitting system. References cited and not cited provide much more in-depth analysis and insight into each of the atmospheric processes described, and an improved understanding will certainly bring forth better solutions. As has been mentioned already, adaptive optics offers a potentially powerful solution to mitigating many of the atmospheric degradations to the optical link. In the

near future planned demonstrations of groundbased reception of laser communication from deep space [60] hold the promise of real data to not only assess the true impact of atmospherics but also provide atmospheric statistics that lend credibility to how often links with a given performance can be operated.

References

[1] W. B. Rossow, F. Mosher, E. Kinsella, A. Arking, M. Desbois, E. Harrison, P. Minnis, E. Ruprecht, G. Seze, C. Simmer, and E. Smith, "ISCCP Cloud Algorithm Intercomparison," *Journal of Climate and Applied Meteorology*, vol. 24, pp. 877–903, 1985.

[2] United States National Oceanic and Atmospheric Administration National Climate Data Center website http://www.ncdc.noaa.gov

[3] Northrop Grumman (TASC), Chantilly, Virginia, a division of Northrop Grumman: http://www.tasc.com/

[4] *Surface Weather Observations and Reports*, Federal Meteorological Handbook (FMH) Number 1, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Silver Spring, Maryland, December 1995.

[5] S. Piazzolla and S. Slobin, "Statistics of Link Blockage Due to Cloud Cover for Free-Space Optical Communication Using NCDC Surface Weather Observation Data," *Proceedings of the SPIE*, vol. 4635, pp. 138–139, 2002.

[6] S. Piazzolla, S. Slobin, and E. Amini, *Cloud Coverage Diversity Statistics for Optical Communications in the Southern United States*, Publication 00-13, Jet Propulsion Laboratory, Pasadena, California, November 30, 2000.

[7] K. E. Wilson and J. V. Sandusky, "Development of a 1-m Class Telescope at TMF to Support Optical Communications Demonstrations," *Proceedings of the SPIE*, vol. 3266, San Jose, California, pp. 146–152, January 1998.

[8] *Cloud Free Line of Sight Study, JPL Contract No. 1248595*, Final Review, Northrop Grumman (TASC), Chantilly, Virginia, September 25, 2003.

[9] R. P. Link, R. J. Allis, and M. E. Craddock, "Mitigating the Effects of Clouds On Optical Communications," Photonics West, 24–29 January 2004, San Jose, California, *Free Space Laser Comm Technologies XVI, Proceedings of the SPIE*, vol. 5338 [G. S. Mecherle, C. Y. Young, and J. S. Stryj, editors], pp. 223–232, June 2004.

[10] A. S. Jursa (sci. ed.), *Handbook of Geophysics and the Space Environment* (NTIS Doc # ADA16700), Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1985.

[11] D. R. Bates, "Rayleigh Scattering by Air," *Planetary Space Science*, vol. 32, pp. 785–790, 1984.

[12] H. C. van de Hulst, *Light Scattering by Small Particles*, Wiley, New York, 1957. (also Dover Publications, January 1, 1982)

[13] *AFGL Atmospheric Constituent Profiles (0–120 km)*, AFGL-TR-86-0110, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1986.

[14] A. Berk, L. S. Bernstein, and D. C. Robertson, *MODTRAN: A Moderate Resolution Model for LOWTRAN 7*, Air Force Geophysics Laboratory Technical Report GL-TR-89-0122, Hanscom Air Force Base, Massachusetts, 1989.

[15] S. A. Clough, F. X. Kneizys, and R. W. Davies, "Lineshape and Water Vapor Continuum," *Atmospheric Research*, vol. 23, p. 229, 1989.

[16] J. V. Dave, "Multiple Scattering in a Non-Homogenous Rayleigh Atmosphere," *Journal of Atmospheric Science*, vol. 22, p. 273, 1965.

[17] O. Dubovik, B. N. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanre, and I. Slutsker, "2002: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations," *Journal of Atmospheric Science*, vol. 59, pp. 590–608, 2002.

[18] J. M. Rosen, "Stratospheric Dust and Its Relationship to the Meteoric Influx," *Space Science Review*, vol. 9, p. 59, 1969.

[19] R. P. Turco, R. C. Whitten, and O. B. Toon, "Stratospheric Aerosols: Observation and Theory," *Rev. Geophysics and Space Research*., vol. 20, no. 2, pp. 233–279, 1982.

[20] H. Weichel, *Laser Beam Propagation in the Atmosphere*, SPIE Optical Engineering Press, Bellingham, Washington, 1990.

[21] S. Q. Duntley, A. R. Boileau, and R. W. Preisendorfer, "Image Transmission by the Troposphere, I," *Journal of the Optical Society of America*, vol. 47, no. 6, pp. 499–506, 1957.

[22] S. Q. Duntley, R. W. Johnson, and J. I. Gordon, *Airborne Measurements of Optical Atmospheric Properties, Summary and Review III*, SIO Reference 79-5, AFGL-TR-78-0286, NTIS ADA 073-121, Scripps Institution of Oceanography, Visibility Laboratory, University of California, San Diego, California, 1978.

[23] *Atmospheric Transmittance/Radiance: Computer Code Lowtran 6*, AFGK-TR-83-0187, Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, 1983.

[24] R. C. Ramsey," Spectral Irradiance from Stars and Planets, above the Atmosphere from 0.1 to 100 Microns," *Applied Optics*, vol. 1, pp. 465–471, 1962.

[25] J. Katz, "Planets as Background Noise Sources in Free Space Optical Communications," *The Telecommunications and Data Acquisition Progress Report 42-85, January–March 1986*, Jet Propulsion Laboratory, Pasadena, California, May 15, 1986.
http://ipnpr.jpl.nasa.gov/progress_report/

[26] *Surface Models of Mars (1975)*, NASA SP-8020, National Aeronautics and Space Administration, Washington, District of Columbia, 1975.

[27] K. Lumme and E. Bowell, "Radiative Transfer in the Surfaces of Atmosphereless Bodies," Astron. J. "I Theory," vol. 86, 1694–1704, 1981; "II Interpretation of Phase Curves," vol. 86, 1705–1771, 1981.

[28] K. Shaik and M. Wilhelm, *Ground Based Advanced Technology Study (GBATS); Optical Subnet Concepts for the DSN*, Jet Propulsion Laboratory, D-11000, Release I (internal document), Jet Propulsion Laboratory, Pasadena, California, August 5, 1994.

[29] K. W. Kobayashi, J. E. Fernandez, J. H. Kobayashi, M. Leung, A. K. Oki, L. T. Tran, M. Lammert, T. R. Block, and D. C. Streit, "A DC-3 GHz Cryogenic AlGaAs/GaAs HBT Low Noise MMIC Amplifier with 0.15 dB Noise Figure," *International Electron Devices Meeting, 1999, IEDM Technical Digest*, cat. no. 99CH36318: 775–8, IEEE, Piscataway, New Jersey, 1999.

[30] M. Moszynski, W. Czarnacki, M. Szawlowski, B. L. Zhou, M. Kapusta, D. Wolski, and P. Schotanus "Performance of Large-Area Avalanche Photodiodes at Liquid Nitrogen Temperature," *IEEE Transactions on Nuclear Science*, vol. 49, no. 3, pp. 971–976, June 2002.

[31] F. Amoozegar, R. Cesarone, S. Piazzolla, L. Paal, and R. Emerson, "Performance Analysis and Comparison of Clustered and Linearly Dispersed Optical Deep Space Network," *Proceedings of 5th International Symposium on Reducing the Cost of Spacecraft Ground systems and Operations (RCSGSO)*, July 8–11, 2003, Pasadena, California, Jet Propulsion Laboratory, 2003.
http://descanso.jpl.nasa.gov/RCSGSO

[32] L. Stepp, L. Daggert, and P. Gillett, "Estimating the Costs of Extremely Large Telescopes." *Proceedings of the SPIE*, vol. 4840, *Future Giant Telescopes*, editors: J.R.P. Angel and R. Gilmozzi, pp. 309–321, 2003.

[33] W. B. Rossow, F. Mosher, E. Kinsella, A. Arking, M. Desbois, E. Harrison, P. Minnis, E. Ruprecht, G. Seze, C. Simmer, and E. Smith, "ISCCP Cloud Algorithm Intercomparison." *Journal of Climate and Applied Meteorology*, vol. 24, pp. 877–903, 1985.

[34] A. Biswas, and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report, 42-154, April–June 2003*, pp. 1–38, August 15, 2003.
http://ipnpr.jpl.nasa.gov/progress_report/

[35] I. Tatarski, *Wave Propagation in A Turbulent Medium*, McGraw-Hill, New York, translated by R. A. Silverman, 1961.

[36] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, IEEE Press, Piscataway, New Jersey, 1997.

[37] J. Stroebehn, ed. *Laser Beam Propagation in the Atmosphere*, Springer, New York, 1978.

[38] R. L. Fante, "Wave propagation in random media: a systems approach," *Progress in Optics XXII*, E. Wolf ed., Elsevier, New York, pp. 341–398, 1985.

[39] R. R. Beland, "Propagation through Atmospheric Optical Turbulence," *The Infrared & Electro-Optical Systems Handbook*, vol. 2, F. G. Smith, ed., SPIE Optical Engineering Press, Bellingham, Washington, 1993.

[40] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Optical Engineering Press, Bellingham, Washington, pp. 157–229, 1998.

[41] D. L. Walters and K. E. Kunkel, "Atmospheric Modulation Transfer Function for Desert and Mountain Locations: the Atmospheric Effects on r_0," *Journal of the Optical Society of America*, vol. 71, pp. 397–405, 1981.

[42] J. H. Brown, R. E. Good, P. M. Bench, and G. Faucher, *Sonde measurements for comparative measurements of optical turbulence*, Air Force Geophysical Laboratory, AFGL-TR-82-0079, NTIS: ADA118740, Hanscom Air Force Base, Massachusetts, 1982.

[43] R. E. Goof, R. R. Beland, E. A. Murphy, J. H. Brown and E. M. Dewan, "Atmospheric Models of Optical Turbulence," *Modeling of the Atmosphere, Proceedings of the SPIE*, vol. 928, pp. 165–186, 1988.

[44] D. L. Fried, "Statistics of Geometric Representation of Wavefront Distortion," *Journal of the Optical Society of America*, vol. 55, pp. 1427–1435, November 1965.

[45] D. L. Walters and L. W. Bradford, "Measurements of r_0 and θ_0: Two Decades and 18 Sites, *Applied Optics*, vol. 36, pp. 7876–7886, 1997.

[46] P. N. Brandt, H. A. Mauter, and R. Smartt, "Day-Time Seeing Statistics at Sacramento Peak Observatory," *Astronomy and Astrophysics*, vol. 188, pp. 163–168, 1987.

[47] E. Seykora, "Solar Scintillation and Monitoring of Solar Seeing," *Solar Physics*, vol. 145, pp. 389–397, 1993.

[48] J. M. Beckers, E. Leon, J. Mason, and L. Wilkins, "Solar Scintillometry: Calibration of Signals and Its Use for Seeing Measurements," *Solar Physics*, vol. 176, pp. 23–36, 1997.

[49] K. E. Wilson, M. Troy, M. Srinivasan, B. Platt, V. Vilarotter, M. Wright, V. Garkanian, and H. Hemmati, "Daytime Adaptive Optics for Deep Space Communications," *Space Activities and Cooperation Contributing to All Pacific Basin Countries (10th ISCOPS)*, vol. 117, *Advances in the Astronautical Sciences* [eds. P. M. Bainum, L. Furong, and T. Nakajina] American Astonautical Society, San Diego, CA 92198, pp. 481–491, 2004.

[50] V. N. Mahajan and B. K. C. Lum, "Imaging through atmospheric turbulence with annular pupils," *Applied Optics*, vol. 20, pp. 3233–3237, 1981.

[51] V. N. Mahajan, *Aberration Theory Made Simple*, vol. 6 (TT-6) of *Tutorial Texts in Optical Engineering*, series editor. D. O' Shea, SPIE Optical Engineering Press, Bellingham, Washington, p. 74, July 1991.

[52] L. C. Andrews, R. L. Phillips, C. Y. Hopen, and M. Al-Habash, "Theory of Optical Scintillation," *Journal of the Optical Society of America A*, vol. 16, pp. 1417–1479, 1999.

[53] J. Churnside, "Aperture Averaging of Optical Scintillations in the Turbulent Atmosphere," *Applied Optics*, vol. 30, no. 15, pp. 1982–1994, 1991.

[54] H. T. Yura and W. G. Mckinley, "Aperture Averaging for Space-to-Ground Optical Communications Applications," *Applied Optics*, vol. 22, pp. 1608–1609, 1983.

[55] R. M. Gagliardi and S. Karp, *Optical Communications*, Second Edition, John Wiley & Sons, p. 298, 1995.

[56] K.E. Wilson, A. Biswas, S. Bloom, and V. Chan, "Effect of Aperture Averaging on a 570 Mbps 42 km Horizontal Path optical Link," *Atmospheric Propagation and Remote Sensing IV, Proceedings of the SPIE*, ed., J. Christopher Dainty, vol. 2471, pp. 244–253, April 1995.

[57] V. Chan and S. Bloom, "Results of 150 km, 1 Gbps Lasercom Validation Experiment Using Aircraft Motion Simulator," *Free-Space Laser Communications Technologies X, Proceedings of the SPIE*, Ed., G. Stephen Mecherle, vol. 2699, pp. 60–70, 1998.

[58] A. Biswas and S. Lee, "Ground-to-Ground Optical Communications Demonstration," *Telecommunications and Mission Operations Progress Report, 42-141, January–March 2000*, pp. 1–31, May 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[59] A. Biswas and M. W. Wright, "Mountain-Top-to-Mountain-Top Optical Link Demonstration: Part I," *The Interplanetary Network Progress Report 42-149, January–March*, pp. 1–27, May 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[60] B. L. Edwards, et al., "Overview of the Mars Laser Communications Demonstration Project," Paper 2003-6417, *AIAA Space 2003 Conference* Long Beach, California, September 23–25, 2003.

[61] B. M. Levine, K. Shaik, and T.-Y.Yan, "Data Analysis for GOPEX Image Frames," *The Telecommunications and Data Acquisition Progress Report, 42-114, April–June 1993*, pp. 213–229, Jet Propulsion Laboratory, Pasadena, California, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[62] J. D. Shelton, "Turbulence-Induced Scintillation on Gaussian-Beam Waves: Theoretical and Observations from a Laser Illuminated Satellite," *Journal of the Optical Society of America A*, vol. 12, pp. 2172–2181, 1995.

[63] P. A. Lightsey, "Scintillation in Ground-to-Space and Retroreflected Laser Beams," *Optical Engineering*, vol. 33, pp. 2535–2543, 1994.

[64] M. Jeganathan, K. E. Wilson, and J. Lesh, "Preliminary Analysis of Fluctuations in the Received Uplink-Beacon-Power Data Obtained from GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report 42-124, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 20–32, February 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[65] M. Jeganathan, M. Toyoshima, K. Wilson, J. James, G. Xu and J. Lesh, "Data Analysis Results from the GOLD Experiments," *Free-Space Laser Communications Technologies IX, Proceedings of SPIE*, Ed., G. Stephen Mecherle, vol. 2990, pp. 70–81, 1997.

[66] C. Higgs, H. Barclay, D. Murphy, and C. A. Primmerman, "Multibeam Illumination," *Lincoln Laboratory Journal*, vol. 11, pp. 8–22, 1998.

[67] R. L. Mitchell, "Permanence of the Log-Normal Distribution, "*Journal of the Optical Society of America*, vol. 58, pp. 1267–1272, 1962.

[68] R. Barakat, "Sums of Independent Log-Normally Distributed Random Variables," *Journal of the Optical Society of America*, vol. 66, pp. 211–216, 1976.

[69] G. A. Tyler, "Bandwidth Considerations for Tracking Through Turbulence," *Journal of the Optical Society of America*, vol. 11, pp. 358–367, 1994.

[70] J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, Oxford, England, 1998.

Chapter 4
Optical Modulation and Coding

Samuel J. Dolinar, Jon Hamkins, Bruce E. Moision, and Victor A. Vilnrotter

4.1 Introduction

It can be argued that optical communications had its origins in ancient times, where modulated sunlight was often used to convey information over large distances. For example, mirrors have been used to create bright flashes of light in certain directions, producing a form of on–off modulation. Similarly, blankets used to cover a signal fire periodically produced puffs of smoke that could be seen for miles in the clear desert air, producing, in effect, a modulated signal. In modern times, navies throughout the world used bright incandescent light sources directed by a reflector and blanked by a manual shutter to send messages between ships often kilometers apart on the high seas. Before the invention of the laser, long-range optical communication was envisioned using bright flashes of light produced by intense pulses of electric current passing through an incandescent fiber placed in the focal-plane of an optical reflector. Even exploding wires that generated bursts of intense optical energy were considered for long-range applications. However, these sources were not effective in the production of intense, highly directional optical energy that could also be modulated at high enough data-rates to be seriously considered for deep-space optical communications.

The invention of the laser by Schawlow, Townes and Maiman [1,2] ushered in the era of deep-space optical communications. Here was a source of intense, highly directed optical energy that could produce coherent radiation, like radio frequency (RF) transmitters, but at much higher optical wavelengths. Because of its short wavelength the optical beam produced by a laser could be highly concentrated in the desired direction, constrained only by diffraction effects imposed

by the finite extent of the transmitter aperture. Therefore, much higher concentrations of signal energy could be delivered to a distant receiver than with radio frequencies, suggesting the potential for communicating at higher data rates for a given amount of expanded power. No longer are optical communications all smoke and mirrors!

It was quickly realized that lasers enabled a new type of modulation format that was not practical at radio frequencies, namely, intensity modulation. The same phase-coherent modulation techniques developed for radio frequencies could still be employed with lasers, but the potential of intensity modulation promised additional choices. This new capability is a direct result of the relatively high energy of optical photons, enabling the detection of individual photons at optical wavelengths, which is much more difficult at radio frequencies. Optical frequencies stand at a nexus of the frequency spectrum where both wave and particle views of light are useful concepts.

Varying the intensity of the laser suggests the use of on–off modulation concepts reminiscent of earlier attempts with solar radiation and incandescent sources, but potentially at a much higher rate. Surprisingly, it has been found [3] that relying entirely on the energy in the signal and ignoring the coherence of the radiation does not fundamentally limit the rate at which information can be transferred reliably over the noiseless optical channel.

In the particle view of light, one can naively imagine using individual photons of light to carry information. For example, to communicate one bit of information, simply transmit one photon or not, depending on the value of the bit. This would be intensity modulation carried to its most granular extreme. If the information bits were equally likely 0 or 1, and if all photons were noiselessly detectable, this scheme would achieve an average photon efficiency of two bits per photon.

Optical photons do not behave exactly as classical particles, and correct quantum mechanical models are needed to describe their generation, modulation, and detection. However, the particle view of light is still a very useful concept for interpreting the efficiencies of optical modulations. Practical optical modulations can communicate information at reasonable fidelity with efficiencies ranging from a handful of bits per (detected) photon to a handful of (detected) photons per bit.

In this chapter, we discuss modulation and coding for the optical communications channel. For this purpose, the optical communications channel may be

reduced to the block diagram of Fig. 4-1. User information, denoted U, is first encoded by an error correction code (ECC), mapping user bits to code bits C. The ECC introduces redundant information into the message to aid in correcting errors in the received signal. The code bits are passed though a modulator, which maps the coded bits to symbols X. The set of symbols represents the distinct messages the laser will transmit. For example, the symbols could be distinct phase, polarization, wavelength, or amplitudes of the optical carrier. These symbols are transmitted over the noisy optical channel, detected and received as noisy versions Y. The process of encoding and modulation is then inverted by the demodulator and decoder. The demodulator takes the noisy received signal and produces estimates of the transmitted symbols $\hat{X}$, or the code symbols $\hat{C}$, or both. The decoder operates on these estimates to yield estimates of the user information $\hat{U}$. In modern modulation and coding design, demodulation and decoding are executed iteratively, illustrated in Fig. 4-l by a pair of directed arrows, passing revised estimates back and forth until the two reach agreement.

In addition to the throughput of the optical communications link, measured in bits/second, we are interested in a measure of its efficiency–a measure of how well we are using the available resources to achieve a desired throughput. A commonly used measure is the photon efficiency ρ, the number of user bits transmitted per signal photon [3,4]. As in [7], the throughput, R, can be written as the product of the photon efficiency and the average signal photons received per second, n_s:

$$R = \rho n_s \quad \text{bits/second} \tag{4.1-1}$$

The usefulness of Eq. (4.1-1) stems from the fact that it separates the communications problem into two parts that can often be treated independently: one relating to the generation and transmission of photons, and one relating to modulation and coding. In cases such that ρ depends only on the modulation and coding and n_s depends only on the characteristics of the laser, the data-rate R can be maximized for a given photon rate by maximizing ρ, subject to constraints imposed by bandwidth, peak power, and other limitations of the receiving and transmitting equipment. This interpretation is valid for a large class

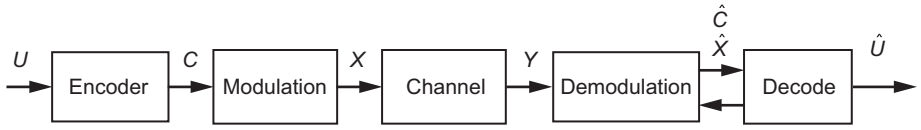


Fig. 4-1. The optimal communications channel.

of modulations, namely those that do not reduce the average photon rate at the receiver: examples include all lossless modulation formats such as phase, polarization, wavelength, and all rate-conserving intensity modulations. However, a decoupled model is often not valid. In practical intensity-modulated systems, the average amount of energy used by the laser-modulator may be a function of the timetable selected by the modulator for transmitting photons. This coupling of the characteristics of the laser and the modulator can destroy the separability of parameters expressed in Eq. (4.1-1). If it is found that while maximizing ρ a point is reached where further increases in ρ begin to reduce n_s, thus violating the independence assumption, then it becomes necessary to consider both ρ and n_s together when attempting to maximize R. In any case, it is clear that high rates of information transfer between transmitter and receiver require modulation and coding schemes capable of high information efficiencies. A thorough description of the generation, detection, and evaluation of efficient modulation and coding schemes capable of operating reliably with high information efficiencies is the subject of the rest of this chapter.

The sections of this chapter each discuss one or more components of Fig. 4-1, working roughly in order of dependence, starting with Section 4.2, which addresses the channel model. In Section 4.2, we discuss the quantum representation of optical fields and models for optical transmission and detection. This results in various statistical characterizations of a detected optical signal at the receiver. These models are the required starting point for the analysis and selection of appropriate modulation and coding, as well as determining the fundamental limits of communication with the optical channel.

Sections 4.3 and 4.4 address modulation. Section 4.3 discusses various common optical modulation schemes. Section 4.4 shows how these common optical modulations can be placed in a unified framework as cases of constrained on–off keying. This framework is used to introduce practical physical constraints on the modulation and analyze their impact. Section 4.5 addresses the performance of the demodulator, analyzing the performance of the uncoded modulation schemes on various channel models.

In Section 4.6 we determine channel capacity limits for the channel models developed in Section 4.2 when used with the modulation formats discussed in Section 4.3. The channel capacity yields bounds on the rate of information transmission as a function of the available physical resources subject to a fidelity criteria, e.g., the maximum bits per second achievable with a given average laser power and probability of error of 10^{-6}. We address physical constraints such as average power, peak power, bandwidth, and decision method at the detector (hard decision versus soft decision).

Section 4.7 discusses various types of channel codes that may be used to encode the user data. Finally, with all the pieces in place, in Section 4.8 we illustrate the performance of the coded modulation schemes on optical channel models, comparing performance to the theoretical capacity limits.

4.2 Statistical Models for the Detected Optical Field

In optical systems the statistical nature of the channel output depends on the detection method. At the receiver, light is focused onto a detector. There are two popular methods for detecting the received optical field. The first method, called *direct detection*, allows the received field to impinge directly upon the photodetector, which responds to its energy, i.e., to the squared magnitude of the incident optical field. Direct detection is non-coherent detection because any information in the phase of the received signal is lost. The second method, called *coherent detection*, adds a strong local optical field to the received field prior to photodetection.

In either case, the detector output may be either discrete or continuous. In most practical detectors—including photo-multiplier tubes (PMTs), avalanche photo-diode (APD) detectors, and PIN diode detectors—the output is a real-valued voltage or current that arises from the detector input as well as from random processes within the detector and follow-on circuitry.

In this section we give a brief formulation of the quantum optical field, and the resulting statistical models of the received optical signal when it is detected coherently or non-coherently.

4.2.1 Quantum Models of the Optical Field

Unlike familiar RF systems, optical systems often operate in regimes where the communication performance is susceptible to the fundamental measurement uncertainties implied by quantum theory. This observation has two ramifications for analyzing optical systems. First, mathematical models that accurately describe physically realizable optical fields and optical detectors must be firmly rooted in quantum theory. Second, the quantum theory permits additional abstract measurements formulated as mathematical operators in a Hilbert space that may in principle outperform the more conventional measurements that can be implemented by direct detection or coherent detection. Considerable attention in the literature [3,5] has been devoted to formulating and analyzing optical communications problems in abstract quantum Hilbert spaces.

The performance of some communications systems is ultimately limited by thermal noise entering the receiver along with the signal. This idea can

be illustrated with Shannon's capacity formula for bandlimited classical channels, according to which error-free communication is possible at rates less than $W \log_2(1 + [S/N])$ bits per second (bps) [6], where W is the bandwidth in hertz, S is the average signal power, and N is the power of additive thermal noise. The above expression shows that if there is no thermal noise, then error-free communication should be possible at arbitrarily high rates, even with bandlimited channels. However, the Shannon formulation implicitly assumes that arbitrarily precise measurements are possible, since if this were not the case then there would have to be an effective noise term associated with the noisy signal measurements in the denominator of the capacity expression in addition to the thermal noise, and hence the denominator would never actually approach zero.

The classical model assumes that deterministic signals are observed in the presence of additive Gaussian noise. This model is perfectly adequate for describing communications systems operating at radio frequencies, where quantum effects are not readily detectable. However, at optical frequencies quantum effects tend to be the dominant source of error, and therefore must be considered in the communications system model. The approach most consistent with the principles of quantum mechanics starts out by quantizing the received electromagnetic field, and seeks to determine those measurements on the received field that achieve the best results such as, for example, minimizing the average probability of detection error. The best measurements may not be readily realizable with physically available devices; however, it is often possible to determine the performance of the "quantum optimum" receiver analytically. Therefore, if these measurements could somehow be made and incorporated into a communications receiver, the performance of the "quantum optimum" receiver would represent the achievable limit on communications system performance consistent with the principles of quantum mechanics.

Another approach for evaluating optical communications systems assumes a classical instead of a quantized received field, but models the response of physically realizable detectors using the same statistics that a quantum mechanical model would provide [7]. This "quantum mechanically correct" detector response is then used as the fundamental observable on which the decisions are based. Receivers using this approach are often called "semi-classical" and have the advantage of employing well-known detection techniques; however, such receivers generally cannot match the performance of the optimum quantum receiver.

4.2.1.1 Quantization of the Electric Field. The application of detection theory to a quantum mechanical model of the aperture field was originally developed by Helstrom and summarized in the IEEE Proceedings article "Quantum Mechanical Communications Theory" [5], as well as the subsequent monograph

Quantum Detection and Estimation Theory, Mathematics in Science and Engineering and the references therein [3]. This section relies heavily on results presented therein and on the numerous references cited in these publications. A detailed derivation of quantum communication theory is beyond the scope of this chapter; therefore, we concentrate on summarizing key results, with a minimum of discussion and derivation.

The receiver in a quantum optical communications system is often modeled as a large cubical box of volume V, with perfectly conducting walls [5]. The received field is admitted into this box, or cavity, through a receiving aperture assumed normal to the direction of propagation, during the time interval $(0,T)$. After this time the aperture is closed, and measurements are made on the "received field" inside the cavity. The received field can be represented as a superposition of normal modes of the cavity, where each mode behaves much like a harmonic oscillator with radian frequency ω_k. The classical waveform $\varepsilon(\mathbf{r},t)$ can be expanded in terms of standing-wave normal mode functions $u_k(\mathbf{r})$ within the cavity as [3,5]

$$\varepsilon(\mathbf{r},t) = -\varepsilon_0^{-1/2} \sum_k p_k(t)\, u_k(\mathbf{r}) \tag{4.2-1}$$

where ε_0 is the dielectric constant for free space, and $p_k(t)$ and $u_k(\mathbf{r})$ describe the temporal and spatial variation of the kth field mode, respectively. If the received field is represented in terms of plane-wave mode functions (instead of standing waves), the spatial variation of each mode takes the form

$$u_k(\mathbf{r}) = V^{-1/2}\, \mathbf{e}_k \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{4.2-2}$$

where $\mathbf{k}$ is the propagation vector and $\mathbf{e}_k$ is a unit polarization vector perpendicular to $\mathbf{k}$. The complex amplitude $\alpha_k(t)$ of each traveling-wave mode can be expressed as

$$\alpha_k(t) = \alpha_k \exp(-\,i\,\omega_k t) \tag{4.2-3}$$

In the quantum formulation, the coordinates and momenta of the harmonic oscillators are replaced by the corresponding operators $Q_k(t)$ and $P_k(t)$, while the complex amplitudes are replaced by the operators a_k. The operator $a_k^+ a_k$ is called the "number operator" because its eigenvectors are the "number states" denoted by the ket $|n\rangle$, and its eigenvalues are the non-negative integers:

$$a_k^+ a_k |n_k\rangle = n_k |n_k\rangle \tag{4.2-4}$$

The operator a_k converts the eigenvector $|n_k\rangle$ to $|n_k - 1\rangle$, whereas the operator a_k^+converts $|n_k\rangle$ to $|n_k + 1\rangle$: for this reason, these operators are often called "annihilation" and "creation" operators.

The electric field operator can be expressed in terms of the annihilation and creation operators [3,5] as

$$E(\mathbf{r},t) = i\sum_k \frac{\hbar\omega_k}{\sqrt{2\varepsilon_0 V}}\mathbf{e}_k \left\{a_k \exp[-i\,(\omega_k t - \mathbf{k}\cdot\mathbf{r}] - a_k^+ \exp[i\,(\omega_k t - \mathbf{k}\cdot\mathbf{r})\right\} \tag{4.2-5}$$

4.2.1.2 The Coherent State Representation of a Single Field Mode. When the k-th mode of the electric field is in a state that is the right eigenvector of the annihilation operator, $a_k|\alpha_k\rangle = \alpha_k|\alpha_k\rangle$, it is said to be in a coherent state. The coherent states are denoted by kets $|\alpha_k\rangle$, and can be expressed in terms of the number eigenstates as [8,9]

$$|\alpha_k\rangle = \exp[-|\alpha_k|^2/2] \sum_{n_k=0}^{\infty} (n_k!)^{-1/2}\, \alpha_k^{n_k} |n_k\rangle \tag{4.2-6}$$

The coherent states are normalized so that $\langle \alpha_k|\alpha_k\rangle = 1$. The overlap between two coherent states, $|\alpha_k\rangle$ and $|\beta_k\rangle$ is not zero; hence, the coherent states are not orthogonal. Denoting the average number of photons in the k-th normal mode by $K_{sk} \equiv |\alpha_k|^2$, the probability that n photons are contained in that mode can be found as

$$\Pr(n) = |\langle n|\alpha_k\rangle|^2 = K_{sk}^n \exp[-K_{sk}]/n! \tag{4.2-7}$$

which is recognized as the Poisson probability mass function (pmf).

Although the coherent states are not orthogonal, they are complete, and hence can be used to expand a large class of density operators, including those of interest in communication theory, as follows [3,5,9]:

$$\rho = \int P(\alpha_k) \prod_k |\alpha_k\rangle\langle\alpha_k|\, d^2\alpha_k \tag{4.2-8}$$

An expansion of this form is called the P representation, where $P(\{\alpha_k\})$is called the weight function. The P representation will be used to describe the density operators of fields consisting of thermal noise, and signal plus thermal noise, in the following development.

4.2.1.3 Quantum Representation of Thermal Noise. When the field inside the cavity consists of thermal radiation alone, in equilibrium with the cavity at an absolute temperature of T_{th} kelvins, the density operator describing the state of the k-th normal mode can be expressed in terms of the P representation as [3,5,9]

$$\rho_k = (\pi K_k)^{-1/2} \int \exp[-|\alpha|^2/K_k]|\alpha\rangle\langle\alpha| d^2\alpha \tag{4.2-9}$$

where K_k is the average number of photons in the k-th mode. Making use of the number-state expansion of the coherent states in Eq. (4.2-6), this density operator can also be expressed in terms of the number states as

$$\rho_k = \sum_{n_k=0}^{\infty} (1 - v_k)\, v_k^{n_k}\, |n_k\,\rangle\,\langle\, n_k| \tag{4.2-10}$$

where $v_k = K_k/(K_k + 1)$. Note that this density operator is diagonal in the number representation. The probability of finding a given number of photons in the kth mode is simply

$$\Pr(n) = \langle\, n|\rho|n\,\rangle = \sum_{n_k=0}^{\infty} (1 - v_k)\, v_k^{n_k}\, \langle\, n|n_k\,\rangle\,\langle\, n_k|\, n\,\rangle = (1 - v_k)\, v_k^n \tag{4.2-11}$$

which is recognized as a Bose–Einstein probability. In the classical limit, when the inequality $\kappa T \gg \hbar\,\omega_k$ holds (here κ is Boltzmann's constant), the weight function of the k-th mode becomes

$$P(\alpha_k) = (\pi K_k)^{-1/2} \exp[-|\alpha_k|^2/K_k] \tag{4.2-12}$$

This is recognized as a Gaussian probability density, representing the probability density of the complex envelope of the kth normal mode.

4.2.1.4 Quantum Representation of Signal Plus Thermal Noise. When the received field mode contains both thermal noise and a coherent-state signal component $|\mu_k\,\rangle$, the density operator is of the form

$$\rho_k = (\pi K_k)^{-1/2} \int \exp[-|\alpha - \mu_k|^2/K_k]|\alpha\rangle\langle\alpha| d^2\alpha \tag{4.2-13}$$

This density operator can also be expressed in terms of number states, as shown in [3,5]. With the number state representation, this density operator can be interpreted as an infinite-dimensional matrix with elements

$$\langle n|\rho_k|m\rangle = (1-v_k)\sqrt{\frac{n!}{m!}} v_k^m \left(\frac{\mu_k^*}{K_k}\right)^{m-n} e^{-(1-v_k)|\mu_k|^2} L_n^{m-n}\left[\frac{-(1-v_k)^2|\mu_k|^2}{v_k}\right] \tag{4.2-14}$$

where $L_n^{m-n}(x)$ is the generalized Laguerre polynomial. If no attempt is made to maintain coherence between the transmitter and the receiver, the density operator must be averaged with respect to phase. If the phase is taken to be uniformly distributed between $(0, 2\pi)$, the averaged density operator $\bar{\rho}_k$ becomes diagonal in the number representation:

$$\bar{\rho}_k = \sum_{k=0}^{\infty} (1-v_k)\, v_k^n \exp[-(1-v_k)K_{sk}]\, L_n[-(1-v_k)^2 K_{sk}/v_k]\, |n_k\rangle\, \langle n_k| \tag{4.2-15}$$

where $L_n(x)$ is the ordinary Laguerre polynomial, and $K_{sk} \equiv |\mu_k|^2$ is the average number of signal photons in the normal mode. As before, the probability of obtaining exactly n photons can be found as

$$\Pr(n) = \langle n|\bar{\rho}_k|n\rangle = (1-v_k)\, v_k^n \exp[-(1-v_k)K_{sk}]\, L_n[-(1-v_k)^2 K_{sk}/v_k] \tag{4.2-16}$$

This is recognized as specifying the well-known Laguerre probabilities for the number of photons in a mode containing both signal fields and noise fields of thermal origin.

4.2.2 Statistical Models for Direct Detection

From a communications systems perspective, it is often convenient to state the various stochastic properties for each detection method in terms of a common framework. For many of the modulations and classical detection methods considered in this chapter, this may be done with the general binary-input discrete time channel model shown in Fig. 4-2. This model can be used for a direct

4.2.2.1.2 Limiting Form of Thermal Noise Fields. Thermal noise fields, and combinations of coherent signal and thermal noise fields, also generate approximate Poisson-distributed random variables at the output of an ideal photon counter, under certain limiting conditions. Whereas the number of photons, hence the modal count, for a single mode of thermal radiation is Bose–Einstein distributed as shown in Eq. (4.2-11), the count distribution for thermal light filtered by a narrowband rectangular optical filter of bandwidth W Hz around the optical carrier frequency, and observed for T seconds, is given by the negative binomial density [7]

$$\Pr(k) = \binom{k + WT - 1}{WT - 1} (1 - v)^{WT} \, v^k \tag{4.2-19}$$

where $v = N/(N + 1)$, and N is the average number of photons in a single mode of the thermal field. If the time-bandwidth product approaches infinity, $WT \to \infty$, and at the same time N decreases such that NWT remains constant, the distribution for the total number of photons approaches the Poisson limit:

$$f_{Y|X}(k|x = 0) = \lim_{\substack{WT \to \infty \\ NWT = c}} \Pr(k) = \frac{K_b^k e^{-K_b}}{k!} \tag{4.2-20}$$

where K_b is the sum of the average numbers of modal background photon counts. This can be extended to the case of signal plus thermal noise: K_b is replaced by $K_s + K_b$ in Eq. (4.2-20) when $x = 1$, as shown in [7].

4.2.2.2 The McIntyre–Conradi Model for APD Detectors. The average number of photons absorbed over the active surface of an APD illuminated with optical power $P(t)$ in T seconds can be expressed as [32]

$$K = \frac{\eta}{h\nu} \int_0^T P(t) \, dt \tag{4.2-21}$$

where h is Planck's constant, ν is the optical frequency, and η is the detector's quantum efficiency, defined as the average number of photons absorbed by the APD's photosensitive surface divided by the average number of incident photons. The actual number of absorbed photons is a Poisson-distributed random variable with mean K (where $K = K_b$ or $K = K_s + K_b$, as in Eqs. (4.2-17) and (4.2-18)).

In an APD, the density of the output electrons in response to absorbed photons was modeled accurately by McIntyre [40] and verified experimentally

4.2.2.1.2 Limiting Form of Thermal Noise Fields. Thermal noise fields, and combinations of coherent signal and thermal noise fields, also generate approximate Poisson-distributed random variables at the output of an ideal photon counter, under certain limiting conditions. Whereas the number of photons, hence the modal count, for a single mode of thermal radiation is Bose–Einstein distributed as shown in Eq. (4.2-11), the count distribution for thermal light filtered by a narrowband rectangular optical filter of bandwidth W Hz around the optical carrier frequency, and observed for T seconds, is given by the generalized Bose–Einstein density [7]

$$\Pr(k) = \binom{k + WT - 1}{WT - 1} (1 - v)^{WT} v^k \tag{4.2-19}$$

where $v = K/(K + 1)$, and K is the average number of photons in a single mode of the thermal field. If the time-bandwidth product approaches infinity, $WT \to \infty$, and at the same time K decreases such that KWT remains constant, the distribution for the total number of photons approaches the Poisson limit:

$$f_{Y|X}(k|x) = \lim_{\substack{WT \to \infty \\ KWT = c}} \Pr(k) = \frac{K^k e^{-K}}{k!} \tag{4.2-20}$$

where c is a constant and K is the sum of the average number of modal photon counts, i.e., $K = K_b$ when $x = 0$ and $K = K_s + K_b$ when $x = 1$.

4.2.2.2 The McIntyre–Conradi Model for APD Detectors. The average number of photons absorbed over the active surface of an APD illuminated with optical power $P(t)$ in T seconds can be expressed as [32]

$$K = \frac{\eta}{h\nu} \int_0^T P(t)\, dt \tag{4.2-21}$$

where h is Planck's constant, ν is the optical frequency, and η is the detector's quantum efficiency, defined as the average number of photons absorbed by the APD's photosensitive surface divided by the average number of incident photons. The actual number of absorbed photons, call it N, is a Poisson-distributed random variable with mean K (where $K = K_b$ or $K = K_s + K_b$, as in Eqs. (4.2-17) and (4.2-18)).

In an APD, the density of the output electrons in response to absorbed photons was modeled accurately by McIntyre [40] and verified experimentally

by Conradi [40]. We refer to it as the McIntyre–Conradi distribution. The conditional density of obtaining k electrons at the APD output in response to $N = n$ absorbed photons is given by

$$f_{Y|N}(k|n) = \frac{n\Gamma\left(\frac{k}{1-k_{eff}}+1\right)}{k(k-n)!\Gamma\left(\frac{k_{eff}k}{1-k_{eff}}+n+1\right)}\left[\frac{1+k_{eff}(G-1)}{G}\right]^{\frac{n+k_{eff}k}{1-k_{eff}}}$$

$$\times\left[\frac{(1-k_{eff})(G-1)}{G}\right]^{k-n} \tag{4.2-22}$$

where G is the average gain of the APD, and k_{eff} is the ionization ratio, $0 < k_{eff} < 1$, a property of the semiconductor. Averaging Eq. (4.2-22) over the number of absorbed photons, n, yields

$$f_Y(k) = \sum_{n=1}^{k} f_{Y|N}(k|n)\,\frac{K^n}{n!}\,e^{-K}, \quad k \geq 1 \tag{4.2-23}$$

where the summation limit is k instead of infinity because according to the model there can never be more absorbed photons than released electrons. Thus, for $k \in \mathbb{N}$ (where $\mathbb{N}$ is the set of natural numbers),

$$f_{Y|X}(k|x) = \sum_{n=1}^{k} \frac{n\Gamma\left(\frac{k}{1-k_{eff}}+1\right)\left[\frac{1+k_{eff}(G-1)}{G}\right]^{\frac{n+k_{eff}k}{1-k_{eff}}} K_x^n e^{-K_x}}{k(k-n)!\Gamma\left(\frac{k_{eff}k}{1-k_{eff}}+n+1\right)n!}$$

$$\times\left[\frac{(1-k_{eff})(G-1)}{G}\right]^{k-n} \tag{4.2-24}$$

where $K_0 = K_b$ is the average number of photons detected when $x = 0$ and $K_1 = K_s + K_b$ is the average number of photons detected when $x = 1$. In determining K for an APD, there are some subtleties to Eq. (4.2-21) that should be noted. The APD bulk leakage current, I_b, is multiplied by the APD gain and can be modeled artificially as part of the background optical power $P(t)$ entering the telescope, even when no actual background light is present. The APD surface leakage current, I_s, is not multiplied by the APD gain and can be modeled as a direct current (DC) at the output.

4.2.2.3 The Webb, McIntyre, and Conradi Approximation to the McIntyre–Conradi Model. An approximation to Eq. (4.2-24) that is simpler to evaluate has been derived by Webb, McIntyre, and Conradi [41], which we refer to as the WMC density, and is given by the continuous conditional density function

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi KG^2F}}\left(1+\frac{(y-GK)(F-1)}{KGF}\right)^{-3/2} \times \exp\left(\frac{-(y-GK)^2}{2KG^2F\left(1+\frac{(y-GK)(F-1)}{KGF}\right)}\right) \quad (4.2\text{-}25)$$

for $y \geq K$, where the excess noise factor, F, is defined as $F = k_{eff}G + (2-[1/G])(1-k_{eff})$. Unfortunately, the constraint $y \geq K$ causes this to be an invalid density function–it does not integrate to one. If we extend the domain artificially, this difficulty is avoided. Denoting the zero-mean, unit-variance WMC probability density function (pdf) (with valid domain indicated) by

$$p_w(y;\delta) = \frac{1}{\sqrt{2\pi}}(1+y/\delta)^{-3/2}\exp\left[\frac{-y^2}{2(1+y/\delta)}\right], \quad y > -\delta \quad (4.2\text{-}26)$$

the conditional density at the output of an APD can be written as

$$f_{Y|X}(y|x) = \frac{1}{\sigma_x}p_w\left(\frac{y-m_x}{\sigma_x};\delta_x\right) \quad (4.2\text{-}27)$$

where $x \in \{0,1\}$, $m_x = K_xG$, $\sigma_x^2 = K_xG^2F$, $\delta_x^2 = K_xF/(F-1)^2$, $K_0 = K_b$, and $K_1 = K_s + K_b$.

Unlike a Gaussian distribution, the WMC distribution is not determined solely by its mean and variance; it also depends on the "skewness" parameter δ. As $\delta \to \infty$, the WMC distribution reduces to the normal distribution, as illustrated in Fig. 4-3. However, the model does not allow independently varying skewness and variance parameters. Both the skewness and variance are signal dependent, being proportional to the average number of absorbed photons, K_x. Thus, this model imposes the constraint $\sigma_0/\sigma_1 = \delta_0/\delta_1$. We shall use this constraint for the WMC channel in the remainder of this chapter.

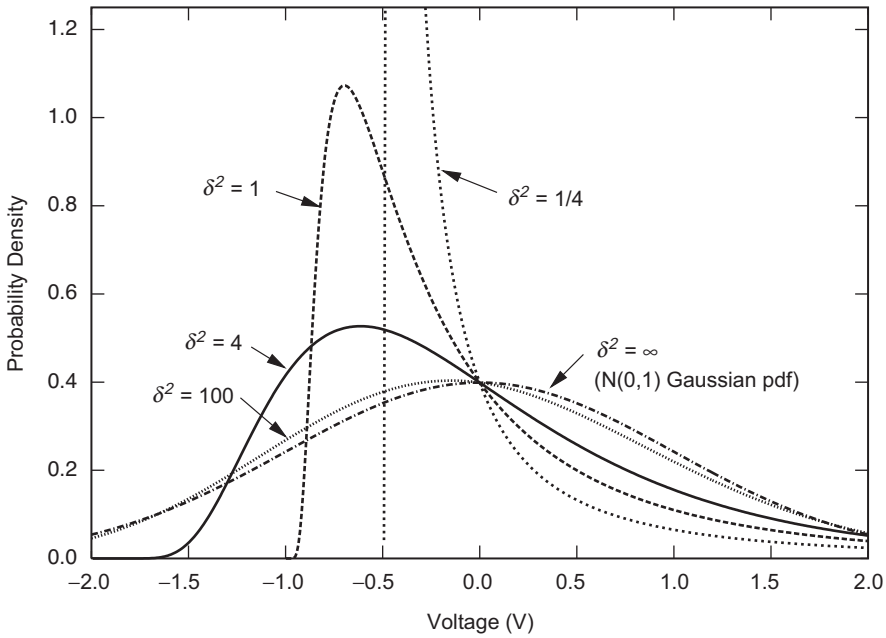


Fig. 4-3. The standardized WMC pdf, for various δ.

4.2.2.4 The WMC Plus Gaussian Approximation. In addition to the avalanche electrons produced by the APD, the follow-on amplifier and resistance within the follow-on electronics also generate electrons of thermal origin, which can be taken into account [42]. That is, the detector output Y is the sum of the McIntyre–Conradi component, with the conditional density function given by Eq. (4.2-24) and an independent normal (Gaussian) component with zero mean and variance σ_n^2. If we use the WMC density in Eq. (4.2-27) to approximate Eq. (4.2-24), the WMC plus Gaussian conditional probability density function becomes

$$f_{Y|X}(y|x) = \int_{m_x - \delta_x \sigma_x}^{\infty} \underbrace{\frac{1}{\sigma_n}\phi\left(\frac{y-z}{\sigma_n}\right)}_{\text{Gaussian pdf at } y-z} \cdot \underbrace{\frac{1}{\sigma_x} p_w\left(\frac{z-m_x}{\sigma_x}; \delta_x\right)}_{\text{WMC pdf at } z} dz$$

where again $x \in \{0, 1\}$ indicates the condition that either a 0 or a 1 was sent, and $\phi(x)$ is the normal density, $\phi(x) = 1/\sqrt{2\pi} e^{-x^2/2}$.

4.2.2.5 Additive White Gaussian Noise Approximation. Under certain conditions, the additive white Gaussian noise (AWGN) channel model can be used to model direct detection [42], and it is a model that has been used

in free-space optical communications link budget software at NASA [34]. The conditional pdf at the output of the channel is given by

$$f_{Y|X}(y|x) = \frac{1}{\sigma_x}\phi\left(\frac{y-m_x}{\sigma_x}\right) = \frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-(x-m_x)^2/2\sigma_x^2} \tag{4.2-28}$$

where $x \in \{0,1\}$, and m_x and σ_x^2 are the mean and variance, respectively, when $X = x$. This model is often only an approximation, and it may over- or underestimate error probabilities [49,50]. In this subsection, we discuss several ways in which this model may arise.

4.2.2.5.1 AWGN Approximation for an APD. The probability density of Y is approximately Gaussian when operating under high background conditions, and in the presence of additive Gaussian noise [49–51]. In this case, the conditional density is given by Eq. (4.2-28), with [51]

$$m_0 = G\,K_b + I_s T_s/q \tag{4.2-29}$$

$$m_1 = G(K_s + K_b) + I_s T_s/q \tag{4.2-30}$$

$$\sigma_0^2 = \left[G^2 F K_b + \frac{I_s T_s}{q} + \frac{2\kappa T T_s}{q^2 R_L}\right] 2BT_s \tag{4.2-31}$$

$$\sigma_1^2 = \left[G^2 F (K_s + K_b) + \frac{I_s T_s}{q} + \frac{2\kappa T T_s}{q^2 R_L}\right] 2BT_s \tag{4.2-32}$$

where G is the average APD gain, K_b and K_s are the average number of absorbed background and signal photons, respectively, I_s is the surface leakage current of the APD, T_s is the slot width, and q is the electron charge. The other parameters are the "excess noise" factor of the APD F, the equivalent noise temperature of the device T, the Boltzmann's constant κ. The noise bandwidth B is assumed to be matched to the slot duration as $B = 1/2T_s$. As mentioned earlier, the bulk dark current of the APD, I_b, can be artificially modeled as part of the background radiation, by absorbing the quantity I_b/q into K_b.

4.2.2.5.2 AWGN Approximation for the Ideal PMT. The conditional probability density of the output voltage Y, given X, of a PMT can be expressed as the sum of conditional densities representing the output voltage for a given number of absorbed photons, n:

$$f_{Y|X}(y|x) = \sum_{n=0}^{\infty} f_{Y|N}(y|n) f_{N|X}(n|x) \tag{4.2-33}$$

where $f_{N|X}(n|x)$ is a Poisson-probability mass function with mean K_b when $x = 0$ and mean $K_s + K_b$ when $x = 1$. The conditional density of the output voltage may be modeled as Gaussian [32]:

$$f_{Y|N}(y|n) = \phi\left(\frac{y - Gn}{\sigma}\right) \tag{4.2-34}$$

where G is the average gain of the PMT and $\sigma = \xi G n$, where ξ is the spreading factor of the PMT [32].

During daytime operation, or with a bright planet (Mars) in the field of view, the Poisson-distributed number of absorbed photons, N, can be much greater than one, which justifies a continuous approximation. A Gaussian approximation of $f_{N|X}(n|x)$ may be made [52], making Eq. (4.2-33) a sum of Gaussians, and hence Gaussian as well:

$$f_{Y|X}(y|x) = \phi\left(\frac{y - G\bar{N}_x}{\sigma_x}\right) \tag{4.2-35}$$

where $\sigma_x = \xi G \bar{N}_x$, with $N_x = K_b$ when $x = 0$ and $N_x = K_s + K_b$ when $x = 1$.

4.2.3 Summary of Statistical Models

Table 4-1 lists the conditional density (or mass) function $f_{Y|X}(y|x)$ associated with each model discussed earlier in this section. In each case, m_0 and σ_0^2 denote the mean and variance conditioned on $X = 0$ (see Fig. 4-2), and m_1 and σ_1^2 denote the corresponding quantities conditioned on $X = 1$. We define the slot signal-to-noise ratio (SNR) as $\beta = (m_1 - m_0)^2/\sigma_0^2$, the "excess SNR" as $\gamma = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$, and the "bit SNR" as $\beta_b = \beta/(2R_c)$, where R_c is the number of bits per symbol.

4.3 Modulation Formats

Optical direct detection effectively measures the energy in the optical signal impinging on the detector. Since direct detection does not respond to phase, but is capable of distinguishing only between different intensity levels, intensity modulation is required. The most common forms of digital intensity

Table 4-1. The pmf's and pdf's of several optical channel models. In each case, for $x \in \{0, 1\}$, m_x and σ_x^2, respectively, are the mean and variance when $X = x$.

Channel Model	pmf or pdf
Poisson	$f_{Y\|X}(k\|x) = \dfrac{m_x^k e^{-m_x}}{k!}$, where $m_0 = K_b, m_1 = K_s + K_b, k \in \mathbb{N}$
AWGN	$f_{Y\|X}(y\|x) = \dfrac{1}{\sigma_x}\phi\left(\dfrac{y - m_x}{\sigma_x}\right) = \dfrac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\dfrac{(y-m_x)^2}{2\sigma_x^2}\right], y \in \mathbb{R}$ (where $\mathbb{R}$ is the set of real numbers)
McIntyre-Conradi	$f_{Y\|X}(k\|x) = \sum_{n=1}^{k} \dfrac{n\Gamma\left(\dfrac{k}{1-k_{eff}} + 1\right) \left[\dfrac{1 + k_{eff}(G-1)}{G}\right]^{n + k_{eff}k/(1-k_{eff})} \left[\dfrac{(1-k_{eff})(G-1)}{G}\right]^{k-n} K_x^n e^{-K_x}}{k(k-n)!\Gamma\left(\dfrac{k_{eff}k}{1-k_{eff}} + n + 1\right) n!}$
WMC	$f_{Y\|X}(y\|x) = \dfrac{1}{\sqrt{2\pi\sigma_x^2}} \left(1 + \dfrac{y - m_x}{\sigma_x \delta_x}\right)^{-3/2} \exp\left[-\dfrac{(y-m_x)^2}{2\sigma_x^2\left(1 + \dfrac{y-m_x}{\sigma_x\delta_x}\right)}\right], y \in \mathbb{R}$
WMC+Gaussian	$f_{Y\|X}(y\|x) = \int_{m_x - \delta_x\sigma_x}^{\infty} \dfrac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\dfrac{(y-t)^2}{2\sigma_n^2}\right] \cdot \dfrac{1}{\sqrt{2\pi\sigma_x^2}} \left(1 + \dfrac{t - m_x}{\sigma_x\delta_x}\right)^{-3/2} \exp\left[-\dfrac{(y-m_x)^2}{2\sigma_x^2\left(1 + \dfrac{t-m_x}{\sigma_x\delta_x}\right)}\right] dt$

modulation are on–off keying (OOK) and pulse-position modulation (PPM). In addition, wavelength modulation (WM) and pulse intensity modulation (PIM) can also be envisioned to increase the information throughput, but have not achieved the level of acceptance enjoyed by OOK and PPM. Wavelength modulation is a form of spatial PPM, hence its performance can be determined directly from PPM detection. Therefore, we shall concentrate on these modulation formats here, and only briefly describe other schemes.

Most of the common modulation formats suitable for direct detection can be decomposed into a simple binary on–off "pulsed" form of modulation. At a given wavelength or during a given period of time an optical pulse is either present or absent. The simplest of these is binary OOK, for which each modulated bit consists of one of these on–off pulses. More complex pulsed modulations, such as PPM, can be conceptually regarded as coded versions of binary OOK, as discussed in Section 4.4.

Coherent reception can be used with any pulsed modulation such as PPM, but additionally there are many non-pulsed modulations suitable for coherent reception. Examples include M-ary phase shift keying (M-PSK) and M-ary quadrature amplitude modulation (M-QAM), i.e., the whole suite of modulations used for coherent RF communication.

Pulsed modulations, such as OOK and PPM, share a common characteristic that the laser is either "on" or "off" during every slot time interval. Similarly, for wavelength-shift keying (WSK) laser energy is either present or absent in each frequency slot. Any such modulation can be conveniently regarded as a mapping from a binary sequence of 0's and 1's into a discrete set of optical pulses. A pulse is present in the ith slot if the ith bit is a 1, and the pulse is absent if the ith bit is a 0.

Pulsed modulations can be decomposed into a coded sequence of binary OOK pulses. Each pulsed modulation format, such as PPM, differential pulse-position modulation (DPPM), etc., simply imposes a different set of constraints on the binary coded sequence. For example, the coded sequences for M-ary PPM are the M binary sequences of length M containing exactly one 1 bit and $M-1$ 0 bits.

4.3.1 On–Off Keying (OOK)

With on–off keying, binary data is represented by the presence or absence of a single light pulse in each T-second symbol interval. The binary information sequence can be mapped directly to a sequence of light pulses at the transmitter

according to the rule: if the information bit is 1, transmit a laser pulse; if it is 0, transmit nothing. Therefore, there is a one-to-one correspondence between 1's in the data-stream, and the occurrence of light pulses emanating from the transmitter. This is illustrated in Fig. 4-4. For comparison with other modulation schemes that use more than one slot per bit, alternate symbols are shaded.

4.3.2 Pulse-Position Modulation (PPM)

Optical PPM is well suited to existing laser modulation techniques (such as Q-switching, mode-locking, and cavity-dumping), requires low average power, attains reasonably high information efficiencies, and is resistant to background radiation.

A k-bit source $\mathbf{U} = (U_1, \cdots, U_k) \in \{0,1\}^k$ is modulated with M-ary PPM, $M = 2^k$, to yield a signal $\mathbf{X} = (0, \ldots, 0, 1, 0, \ldots, 0) \in \{0,1\}^M$, which contains a single one in the position indicated by the binary representation of $\mathbf{U}$. The transmission channel is a binary-input unconstrained-output memoryless channel ($X = 0$ or 1 in Fig. 4-5). One use of the overall PPM-symbol channel consists of M serial uses of the binary-input channel, and it produces the received vector $\mathbf{Y} = (Y_1, \cdots Y_M) \in \mathbb{R}^M$ This is illustrated in Fig. 4-6.

Fig. 4-4. OOK uses one slot per bit. Modulation of message 101001 is shown.

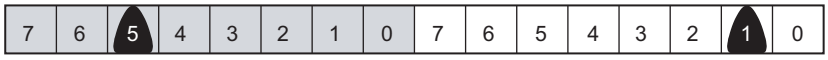


Fig. 4-5. 8-PPM uses an eight slot symbol for each three bits. Modulation of message 101001 is shown. The order in which slots are labeled is not consequential; here, label 7 is assigned to the left-most slot of a symbol, and label 0 is assigned to the right-most slot.

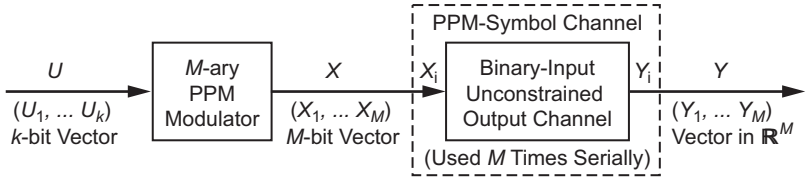


Fig. 4-6. PPM signaling.

This modulation format consists of a fixed number of symbols, M, of equal duration, T seconds. Assuming for convenience that M is a power of two, $M = 2^L$, and recalling that the number of bits of information contained in a PPM symbol is $\log_2 M$, we can view the mapping from information-bits to PPM symbols as a one-to-one assignment of symbols to each of L consecutive information bits. To illustrate, consider the following sequences of $L = 4$ information bits and a particular mapping of a PPM pulse into one of $M = 2^L$ possible time-slots. Suppose each consecutive L-bit information sequence is mapped into a unique PPM symbol according to the rule, "1 plus the numerical value of the sequence, when the sequence is viewed as an L-digit binary number." According to this mapping, the sequence [0,0,0,0] is mapped to the first PPM slot at the transmitter since the numerical value of this sequence is 0. At the receiver, after slot and symbol synchronization has been achieved, this transmitted laser pulse gives rise to an average signal energy of K_s photons. Similarly, the sequence [1,0,0,1] corresponds to a laser pulse in the 10th slot, whereas the sequence [1,1,1,1] is mapped into the PPM symbol with a single pulse in the last, or 16th, time slot.

4.3.3 Differential PPM (DPPM)

In differential PPM [28], also called truncated PPM (TPPM), throughput is increased by beginning a new PPM symbol immediately following the slot containing the pulse. That is, non-pulsed slots of a PPM symbol which follow a pulsed slot are flushed. This imposes a more challenging synchronization problem because the symbols vary in length and error propagation could occur at the receiver or decoder. However, it also increases the throughput per unit time by a factor of two, since symbols are on average half as long as they would be with ordinary PPM. DPPM is shown in Fig. 4-7.

A Q-switched laser works well with the PPM format [53,54] because it can successfully confine a large pulse energy to a narrow slot. One side effect of Q-switched lasers, however, is a required delay, or dead time, between pulses during which the laser is recharged. This delay is significant relative to the pulse duration. PPM may be modified to satisfy the dead-time constraint by following each frame by a period during which no pulses are transmitted. However, this affects the optimality of PPM as a modulation format.

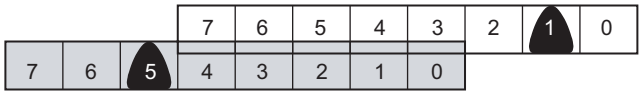


Fig. 4-7. 8-DPPM is the same as 8-PPM, except all slots following the pulse are flushed and a new symbol immediately follows. Modulation of message 101001 is shown.

DPPM is an attractive low-complexity, high-throughput scheme. However, it has two implementation issues that are common to variable-rate schemes. The first is the difficulty of adapting decoding algorithms to function with variable-rate codes. It may be particularly difficult to accommodate the code as an inner code in a concatenated coding scheme since a block of data from the outer code would map to a variable length block of transmitted symbols.

The second problem is the possibility of catastrophic error propagation in decoding due to loss of frame synchronization. Assume the DPPM decoder operates in a manner similar to a PPM decoder, by choosing the maximum slot count in an appropriate window. If an error is made in the estimation of the pulse position, the location of the following window will be incorrect. The detector will, however, re-synchronize with the next correctly detected pulse position. The probability of re-synchronizing in the frame following a pulse-position estimation is $\approx 2/3$ for q small and M large.

There may be methods of averting the problems with TPPM by buffering data and performing an appropriate sequence-detection algorithm. However, given lower-complexity options with similar performance, we did not pursue implementing DPPM.

4.3.4 Overlapping PPM (OPPM)

Overlapping PPM is a generalization of PPM proposed in [55]. In OPPM, each symbol interval of length T is divided into NM chips of duration $T_c = T/(NM)$. A pulse occupies N chips, and is constrained to be entirely contained within the symbol epoch. When $N = 1$, we have ordinary PPM discussed above, in which $\log_2 M$ bits are transmitted per T seconds. When $N > 1$, the pulse can be in one of $NM - N + 1$ positions, and we have $\log_2[NM - N + 1]$ bits per T seconds, i.e., nearly an additional $\log_2 N$ bits per T seconds for large M. OPPM signaling is shown in Fig. 4-8. OPPM imposes more stringent synchronization requirements, and special synchronizable codes may used to aid in this [29].

Fig. 4-8. OPPM with N = 3 and M = 2 has six chips per two bits. The four possible starting positions for three-chip-wide pulses are indicated. Modulation of message 101001 is shown.

4.3.5 Wavelength Shift Keying (WSK)

From the viewpoint of communication theory, wavelength-shift keying is similar to PPM of the same dimension. Instead of placing a single pulse of laser light of a given wavelength into one of M time-disjoint time-slots, WSK places a single laser pulse of one of M disjoint wavelengths into one of M optical detectors, as shown in Fig. 4-9.

4.3.6 Combined PPM and WSK

One way to avoid the linear increase in required average laser power with data-rate that is characteristic of WSK, is to use a combination of PPM and WSK, as shown in Fig. 4-10. By restricting the laser pulse to one of M time-slots, but allowing the laser pulse to take on any of N wavelengths, the low average laser power of conventional PPM can be maintained with the added advantage of increased data-rate. This can be demonstrated by observing that the dimensionality of the signal-space has been increased from M-dimensions (PPM) to NM dimensions.

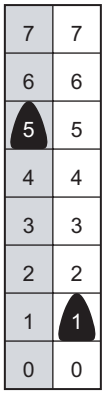


Fig. 4-9. 8-ary WSK uses eight frequencies in one time slot for each three bits. Modulation of message 101001 is shown.

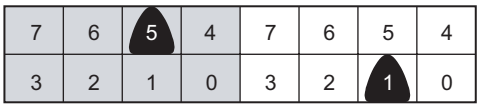


Fig. 4-10. Binary WSK with 4-PPM uses two frequencies and four time slots for each three bits. Modulation of message 101001 is shown.

4.4 Rate Limits Imposed by Constraints on Modulation

A modulation format enforces constraints on the relative location of pulses and describes a mapping of information bits to the sequence of pulses. The constraints are driven by the limitations of the physical devices and follow-on electronics. For example, Q-switched lasers require a minimum delay between pulses to allow the laser to recharge, while timing synchronization algorithms benefit from limiting the maximum delay between pulses. The modulation format is designed to satisfy such constraints while maximizing the throughput, or transmitted bits per second. The upper bound on the throughput for a set of constraints is referred to as the *Shannon capacity* of the constraint, to distinguish it from the channel capacity. The Shannon capacity may be thought of as the capacity of a channel with constraints on the inputs but no distortion from the channel. Given a set of practical constraints, we would like to determine the Shannon capacity and find modulations that satisfy the constraint in an efficient manner, i.e., with throughputs close to the Shannon capacity. This section treats those two problems.

If we represent a pulsed slot by a binary 1 and a non-pulsed slot by binary 0, the modulation format is seen to be a binary code mapping unconstrained binary sequences to constrained binary sequences. For example, PPM enforces the constraint that there is exactly one pulse within each frame of M slots and maps $\log_2 M$ bits to each of the M possible frames. It is, in essence, a binary code of rate $(\log_2 M)/M$. To reflect this, we refer to the modulation as a *modulation code*, a code designed primarily to satisfy certain signalling constraints, as opposed to correcting errors, although the distinction is muddled when both are involved in error correction, or are treated as a single code, as with some iterative decoding techniques. Any constraints on the sequence of 0's and 1's imposed by the physical devices and implemented by some modulation limits the amount of information per slot that can be conveyed. An unconstrained channel can support 1 bit/slot. In comparison, M-ary PPM conveys only $(\log_2 M)/M$ bits/slot, which decreases monotonically in M, for $M > 3$.

The study of Shannon capacity and modulation codes has taken on additional significance recently with the advent of new techniques in quantum mechanics capable of generating orthogonal quantum states. As we shall show in Section 4.5.4, when optimum quantum detection is employed, the probability of error for both binary and higher dimensional modulation depends on the overlap between the quantum states that represent the various hypotheses. The overlap approaches zero as the states become orthogonal, yielding error probabilities that also approach zero. Hence, error-free detection is possible when orthogonal quantum states are used. For this class of signals, the relevant channel capacity is the Shannon capacity.

Recent developments in the generation of number states, $|n\rangle$, suggest their use in future communications systems. Since number states are orthogonal, in principle they can be detected without error. Furthermore, since number states represent energy, classical energy detectors, such as photon counters, provide optimum quantum measurements. Therefore, once the number states are generated and transmitted, their detection via classical photon-counting detectors is certain provided the channel does not introduce any appreciable losses. If the channel is lossy, then the number-state channel transforms into a classical erasure channel, where the only source of error is loss of the transmitted state.

For the case of number-state communications, capacity can be determined using a two-step process, where the external noise is first reduced to zero, resulting in a classical erasure channel. Then the channel losses are further reduced to zero, yielding the Shannon capacity.

In this section we consider several physical constraints and show their impact on the achievable throughput. We also show methods of constructing modulation codes to implement the constraint.

4.4.1 Shannon Capacity

Constraints on the allowable sequences of 0's and 1's may be described in a concise manner by a labelled graph. Figure 4-11 illustrates a graph describing the deadtime constraint in which each 1 is separated by at least one 0.

A graph $\mathcal{G}$ consists of a set of states $\mathcal{V}$ and a set of labeled edges $\mathcal{E}$. Each edge e is directed, with initial state $i(e) \in \mathcal{V}$ and terminal state $t(e) \in \mathcal{V}$. Sequences are generated by traversing edges in the graph and reading off, or concatenating, their labels to form words or finite strings of bits. The collection of all sequences presented by a graph is referred to as a *constrained system* L^S. We assume throughout that the graphs are *irreducible*, meaning that there is a path between any two states in the graph; *right-resolving*, meaning that the labels of paths beginning in the same state generate distinct words for all sufficiently

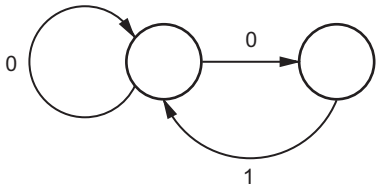


Fig. 4-11. A one-slot deadtime constraint graph.

long paths; and *primitive*, meaning that for some m, there is a path between any two states of length m. Most constraints of practical interest have an irreducible, right-resolving, primitive presentation.

A constrained code is an invertible mapping of unconstrained binary sequences into constrained binary sequences. The maximum rate of an invertible code that maps source sequences into sequences from the constrained system is referred to as the Shannon capacity of the system. It is the noiseless capacity of a channel defined by the graph in the sense that sequences are transmitted noiselessly, but are constrained to be those generated by paths on the graph. Letting $N(n, S)$ denote the number of distinct words of length n occurring in a constrained system S, the Shannon capacity is given by

$$C(S) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \log_2 N(n, S) \text{ bits/slot} \tag{4.4-1}$$

For example, the Shannon capacity of the system illustrated in Fig. 4-11 is 0.6942 bits/slot, which tells us that, if pulses must be separated by at least one non-pulsed slot, we are strictly limited to data rates less than 0.6942 bits/slot. There are several approaches to determining the capacity. We will present them here without proof.

4.4.1.1 Characterizing Capacity: Fixed Duration Edges. Suppose we describe the constraint with a graph wherein all edges correspond to a duration of one slot. The *adjacency matrix* of $\mathcal{G}$ is the matrix $A_\mathcal{G}$ with the u, vth entry containing the number of edges from state u to state v. For example $A_\mathcal{G}$ corresponding to Fig. 4-11 is

$$A_\mathcal{G} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{4.4-2}$$

Let S be the system presented by $\mathcal{G}$. One can show

$$C(S) = \log_2 \big(\rho(A_\mathcal{G})\big) \text{ bits/slot} \tag{4.4-3}$$

where $\rho(A_\mathcal{G})$ is the spectral radius of $A_\mathcal{G}$, the largest magnitude eigenvalue of the adjacency matrix. For any non-negative matrix such as an adjacency matrix, the spectral radius is non-negative. For $\mathcal{G}$ that is irreducible and right-resolving, the spectral radius has a multiplicity of one. The spectral radius of Eq. (4.4-2) is $(1 + \sqrt{5})/2$, the golden mean. Hence the capacity is 0.6942 bits/slot.

To determine the spectral radius for matrices of small dimension, we may find the eigenvalues and choose the largest. However, this is computationally prohibitive for matrices of large dimensions. In this case, one can find a good approximation to the spectral radius via the power method, e.g., [56, Exercise 8.5.16].

4.4.1.2 Characterizing Capacity: Variable Duration Edges. The previous characterization required each edge in the graph to correspond to one slot, or a constant duration. We may describe the constraint more concisely, and allow a broader class of constraints, by allowing edges to have variable durations. For example, Fig. 4-12 is an equivalent presentation of the constraint presented in Fig. 4-11. Let $\tau(e)$ be the duration of edge e, measured in slots. The *partition matrix* of $\mathcal{G}$ is the matrix with i,jth entry

$$B_{i,j}(s) = \sum_{e \in \mathcal{E}, i(e)=i, t(e)=j} e^{-s\tau(e)}$$

For example, the partition matrix corresponding to Fig. 4-12 is $B(s) = e^{-s}+e^{-2s}$. It can be shown that [57]

$$C(S) = \frac{s_0}{\ln(2)} \text{ bits/slot} \tag{4.4-4}$$

where s_0 is the solution to $\rho(B(s)) = 1$. The spectral radius $\rho(B(s))$ is strictly decreasing in s; hence, the solution is unique. For $B(s) = e^{-s} + e^{-2s}$, $s_0 = \ln((1+\sqrt{5})/2)$. Hence, the capacity of Fig. 4-12 is, again, 0.6942 bits/slot. Note that if all edges have duration 1, then $B(s) = \exp(-s)A_{\mathcal{G}}$, $s_0 = \ln(\rho(A_{\mathcal{G}}))$, and $C(S) = \log_2 \rho(A_{\mathcal{G}})$, in agreement with Eq. (4.4-3).

4.4.1.3 Characterizing Capacity: Probabilistic Characterization. We can consider the graph as describing a Markov source with each edge assigned a *transition probability* $p(e)$ such that

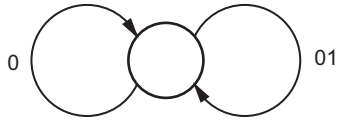


Fig. 4-12. A one-slot deadtime constraint graph, variable duration edges.

$$\sum_{e,i(e)=v} p(e) = 1 \; \forall v \in \mathcal{V}$$

Let $P_{u,v} = \sum_{e,i(e)=u,t(e)=v} p(e)$ be the *transition probability matrix* with the u,vth entry denoting the probability of a transition from u to v, given the source was in state u. For an irreducible graph, the $p(e)$'s define a unique stationary distribution μ on the states, the left eigenvector of the transition probability matrix

$$\mu^T = \mu^T P$$

normalized such that $\sum_{v\in\mathcal{V}} \mu(v) = 1$.

The *absolute transition probabilities* $q(e) = \mu(i(e))p(e)$ denote the probability that edge e is selected at an arbitrary time and satisfy $\sum_{e\in\mathcal{E}} q(e) = 1$. The information conveyed by transmitting the label of e is $-\log_2 q(e)$ bits, and the time required to send this information is $\tau(e)$. The average information that may be conveyed by this source per edge is the entropy rate of the source

$$H(\mathbf{q}) = -\sum_{e\in\mathcal{E}} q(e) \log_2 p(e) \text{ bits/edge}$$

However, edges have variable durations. The average duration of an edge is

$$T(\mathbf{q}) = \sum_{e\in\mathcal{E}} q(e)\tau(e) \tag{4.4-5}$$

hence, the information rate per slot is

$$\frac{H(\mathbf{q})}{T(\mathbf{q})} \text{ bits/slot}$$

Let $\mathbf{q}^*$ be the distribution that maximizes the information rate in bits per slot. Shannon [6] proved that this probabilistic notion of capacity is equivalent to the combinatorial notion given in Eq. (4.4-1), i.e.,

$$C(S) = \frac{H(\mathbf{q}^*)}{T(\mathbf{q}^*)} \tag{4.4-6}$$

It is informative to determine the maximizing distribution $\mathbf{q}^*$. Let ξ and η be the unique left and right eigenvectors of $B(s_0)$ associated with eigenvalue 1,

$$\xi B(s_0) = \xi$$

$$B(s_0)\eta = \eta$$

normalized such that $\xi\eta = 1$. The maximizing distribution is given by

$$p^*(e) = \frac{\eta_{t(e)}}{\eta_{i(e)}} e^{-s_0\tau(e)}$$

$$q^*(e) = \xi_{i(e)}\eta_{t(e)} e^{-s_0\tau(e)}$$

$$\mu^*(u) = \eta_u \xi_u$$

4.4.1.4 Characterizing Capacity: Energy Efficiency. In the prior sections, we defined capacity as the least upper bound on the efficiency of the modulation in bits per slot. We may also define capacity as the least upper bound on the efficiency of the modulation in bits per photon. The characterization from Sections 4.4.1.2 and 4.4.1.2 extend to this case in a straightforward manner by assigning edges a cost $\tau(e)$ in the appropriate units.

Redefine $\tau(e)$ to be the number of photons transmitted when edge e is traversed. Then the average number of photons per edge is given by Eq. (4.4-5), and the capacity in bits per photon is given by Eq. (4.4-6). For example, in Fig. 4-12, let the edge labeled '01' cost K_s photons, and the edge labeled '0' cost no photons. Then $B(s) = 1 + e^{-s}$, and s_0 is the solution to $e^{-s_0} = 0$. In this case, we can transmit at arbitrarily large bits/photons. One can see this is accomplished by letting the probability of using edge '01' go to zero.

4.4.2 Constraints

We consider several constraints on laser transmission. Certain lasers require a recovery time after transmission of a pulse, imposing a minimum delay, or deadtime, between pulses. Timing recovery will impose a constraint on the maximum time that may elapse between pulses. Finally, any deep-space optical system will be subject to an average power constraint, or, equivalently for equal-power pulses, a duty-cycle constraint.

4.4.2.1 Dead Time. We first consider the general form of the deadtime constraint, presented in Fig. 4-13. The capacity, from Eq. (4.4-4) is

$$C(d) = \log_2(\lambda)$$

where λ is the largest positive root of

$$\lambda^{-(d+1)} + \lambda^{-1} - 1 = 0 \tag{4.4-7}$$

For small d, exact solutions may be found efficiently for Eq. (4.4-7). For large d, substitute $\lambda = e^{C(d)}$ and use the approximation $e^{-C(d)} \approx 1 - C(d)$, which yields $C(d)e^{(d+1)C(d)} \approx 1$ and thus

$$d + 1 \approx (d + 1)C(d)e^{(d+1)C(d)}$$

or

$$C(d) \approx \frac{W(d+1)}{(d+1)\ln 2} \text{ bits/slot} \tag{4.4-8}$$

where $W(z)$ is the *productlog* function, which gives the solution for w in $z = we^w$.

Denote the two edges in the graph e_0 and e_{10^d}. Each transmission of e_0 can be thought of as an unpulsed slot in excess of the minimum required. For large d, the optimal probability of this transmission is

$$p(e_0) \approx e^{W(d+1)/(d+1)}$$

$$= \exp\big(-\exp\big(-W(d+1)\big)\big)$$

The productlog function grows roughly like the log, so, to a rough approximation, the probability of inserting excess zeros grows as $e^{-1/(d+1)}$.

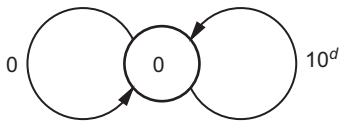


Fig. 4-13. Deadtime constraint.

4.4.2.2 Runlength. Any practical system will also limit the duration between pulses to aid in timing recovery. Let k denote the maximum number of consecutive slots without a pulse, or the *runlength* constraint. Figure 4-14 illustrates the deadtime and runlength constraint. The capacity, from Eq. (4.4-4), is

$$C(d,k) = \log_2(\lambda) \text{ bits/slot}$$

where λ is the largest positive root of

$$\lambda^{-(d+1)} + \lambda^{-(d+2)} + \cdots + \lambda^{-(k+1)} - 1 = 0$$

The subject of deadtime and runlength constraints has been extensively studied in the literature pertaining to magnetic and optical recording, e.g., [58], where modulation codes of this type are extensively used.

The energy efficient case has a simple solution. Let each transmission of a 1 cost a single photon, and let $C_E(d,k)$ be the energy-efficient deadtime and runlength capacity. We have

$$C_E(d,k) = \log_2 k - d + 1 \text{ bits/photon}$$

achieved by choosing each of the edges with equally likely probability.

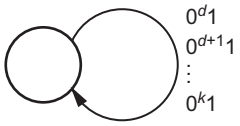


Fig. 4-14. Deadtime and runlength constraint.

4.4.3 Modulation Codes

A modulation code is an invertible mapping from unconstrained binary sequences into a system S. Let $R_{\mathcal{C}}(S)$ denote the rate of a modulation code $\mathcal{C}$ into the system S. Then $E_{\mathcal{C}(d)} \stackrel{\text{def}}{=} R_{\mathcal{C}}(S)/C(S)$ is the relative *efficiency* of the code, measuring how close the code rate is to the limit. Note that the efficiency of a code is measured relative to the constraint it is designed to satisfy.

There are well-known techniques to construct codes into a modulation system at rates arbitrarily close to capacity, e.g., [59,60]. However, for our parameter

range, a straight-forward application of these approaches may be prohibitively complex. In the following sections we present some approaches that trade efficiency for complexity.

4.4.3.1 *M*-ary PPM with Deadtime. PPM may be made to satisfy a deadtime constraint by appending d non-pulsed slots after each PPM frame as illustrated in Fig. 4-15. A graph describing the allowable sequences for deadtime-PPM is illustrated in Fig. 4-16.

The rate is

$$R_{\text{ppm}}(d, M) = \frac{\log_2(M)}{M + d} \text{ bits/slot}$$

We will allow non-integer M in analysis to simplify expressions, since rounding has a negligible effect on rate for large d. For a given value of d we find M^*, the argument that maximizes $R_{\text{ppm}}(d, M)$, by solving $\partial R_{\text{ppm}}(d, M)/\partial M = 0$, which yields

$$\frac{\ln(M^*)}{M^* + d} = \frac{1}{M^*}$$

or

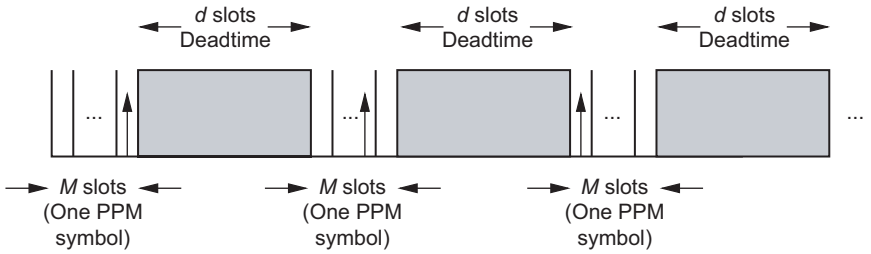


Fig. 4-15. PPM signaling with a deadtime constraint.

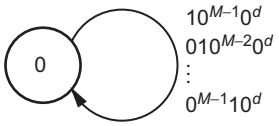


Fig. 4-16. Deadtime-PPM.

$$M^* = \frac{d}{W(d/e)} \tag{4.4-9}$$

Noting that $R_{\text{ppm}}(d, M^*) = 1/(M^* \ln 2)$, the maximum rate is

$$R_{\text{ppm}}(d) = \frac{W(d/e)}{d \ln 2} \text{ bits/s}$$

By an application of L'Hôpital's rule, one can show $R_{\text{ppm}}(d)/C(d) \to 1$ as $d \to \infty$, i.e., deadtime-PPM achieves deadtime capacity in the limit of large d. However, for small to moderate d, significant gains in throughput over PPM are available.

4.4.3.2 *M*-ary DPPM with Deadtime. With deadtime-PPM, there are d unused slot positions in the transmitted signal following each frame—unused in the sense that they neither convey information nor are always necessary for satisfying the deadtime constraint. It would be more efficient to map $\log_2 M$ bits to a pulsed slot position and follow each pulse by exactly d non-pulsed slots as illustrated in Fig. 4-17. This signaling scheme is referred to as deadtime-differential-PPM (deadtime-DPPM), presented by Fig. 4-18.

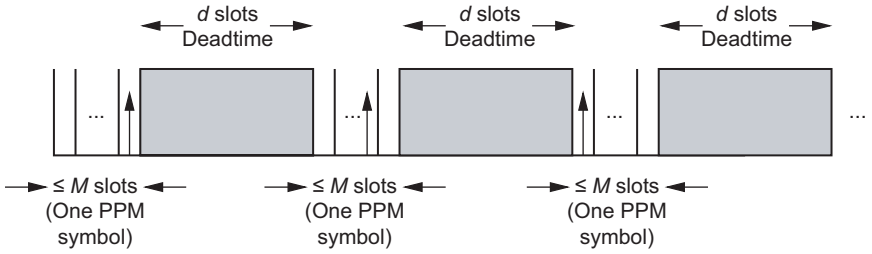


Fig. 4-17. In DPPM signaling, the designated deadtime begins immediately after the pulse of the PPM symbol.

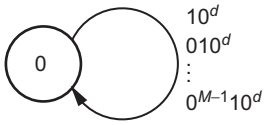


Fig. 4-18. Deadtime-DPPM.

Since the duration of a codeword mapping to $\log_2 M$ bits is variable, deadtime-DPPM has an *average* rate

$$R_{\text{dppm}}(d, M) = \frac{\log_2(M)}{\frac{M+1}{2} + d} \text{ bits/slot}$$

For a given value of d we find M^*, the argument that maximizes $R_{\text{dppm}}(d, M)$, by solving $\partial R_{\text{dppm}}(d, M)/\partial M = 0$, which yields

$$M^* = \frac{2d+1}{W\left(\frac{2d+1}{e}\right)}$$

and maximum achievable rate

$$R_{\text{dppm}}(d) = \frac{2W\left(\frac{2d+1}{e}\right)}{(2d+1)\ln 2} \text{ bits/slot}$$

Since $R_{\text{dppm}}(d)$ is bounded above by $C(d)$ and below by $R_{\text{ppm}}(d)$, DPPM also achieves capacity in the limit of large d. DPPM is a low-complexity scheme that demonstrates significant throughput gains over PPM. However, there are practical issues with implementing variable-rate decoders. Hence we investigate an intermediate solution.

4.4.3.3 Synchronous Variable-Length Codes. The encoders considered so far have been either fixed rate or variable rate. Allowing a variable rate adds a degree of freedom in design, resulting in higher efficiency and/or lower complexity encoders. However, variable rate encoding and decoding has practical drawbacks. A compromise is to allow a synchronous rate, namely mappings of mp bits to mq bits, where p, q are fixed positive integers, and m is a positive integer that can vary. Methods of constructing synchronous variable-length codes were initially described in [61], and reviews of various approaches may be found in [58,60].

We describe a systematic procedure to construct synchronous encoders and decoders for (d, ∞) constraints. The procedure may be interpreted as a practical method of approaching rates of deadtime-DPPM.

Choose a rate $p/q < C(d)$ bits/slot. We desire a set of variable-length codewords $\mathcal{C} = \{c_1, c_2, \cdots, c_N\}$ such that any sequence formed by freely concatenating the codewords satisfies the constraint, the codeword lengths $l(c_i)$ are

multiples of q, no codeword is the prefix of another (sufficient but not necessary to guarantee decodability), and the collection satisfies the Kraft (In)equality:

$$\sum_{c_i \in \mathcal{C}} 2^{-l(c_i)p/q} = 1 \tag{4.4-10}$$

We can use such a set to construct a synchronous variable-length code mapping unconstrained binary sequences into the constraint.

We detail one method to construct such a set that leads to a low-complexity encoder and decoder. The codewords are constructed as nodes on a binary tree. The root of the tree is the pattern 0^d. Branches with a label 1 are extended with zeros to the first length that is a multiple of q. At this point, the branch label is taken as a codeword. The tree is expanded until we have a set of codewords that satisfies Eq. (4.4-10). Figure 4-19 illustrates the procedure for $(d, k) = (16, \infty)$, $q = 7$.

The all-zeros pattern is not allowed as a codeword since allowing it reduces the minimum Euclidean distance from 2 to 1, the small gain in throughput does not offset the loss due to the smaller distance (allowing the all zeros codeword does yield significant throughput gains for small d), and a finite k constraint is desired for synchronization. The encoding and decoding may be done at a fixed rate by using encoders and decoders with appropriate memory. A simple encoder implementation exists if we allow variable-out-degree states.

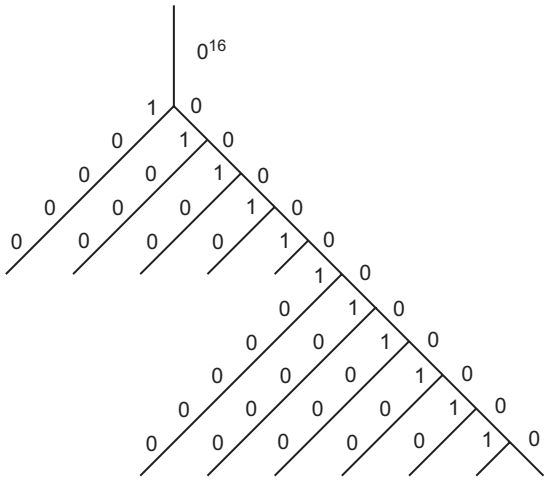


Fig. 4-19. Synchronous variable-length construction for $(d,k) = (16,\infty)$, $q = 7$.

4.5 Performance of Uncoded Optical Modulations

We now analyze the performance of the modulation schemes discussed in Section 4.3 when used on channels governed by the models described in Section 4.2. As in Fig. 4-2, let X denote the message to be sent on the channel and let Y denote the output of the detector. If we observe $Y = y$, the maximum-likelihood (ML) detector[1] produces the output

$$\hat{X} = \underset{x}{\operatorname{argmax}}\, f_{Y|X}(y|x) \tag{4.5-1}$$

where $f_{Y|X}(y|x)$ is the conditional probability density function or probability mass function of Y, given $X = x$. If X is a priori uniformly distributed, as it is for most cases of interest, the ML detector minimizes the probability of detector error $\Pr(\hat{X} \neq X)$. We shall assume a uniform a priori distribution throughout the rest of the chapter.

When X can take on only the values 0 or 1, it is often convenient to write the ML detector in terms of the log-likelihood ratio $\Lambda(y) = \ln L(y) = \ln[f_{Y|X}(y|1)/f_{Y|X}(y|0)]$. The log-likelihood ratio test is an equivalent way to write Eq. (4.5-1) [62]:

$$\hat{X} = \begin{cases} 1 & \text{if } \Lambda(y) > 0 \\ 0 & \text{if } \Lambda(y) \leq 0 \end{cases} \tag{4.5-2}$$

When $\Lambda(y) = 0$, Eq. (4.5-2) defines $\hat{X} = 0$, but the error rate would not be affected by defining $\hat{X}$ to be randomly selected from 0 and 1. As we shall see in the following sections, the log-likelihood ratio test often reduces to a simple threshold test of the form

$$\hat{X} = \begin{cases} 1 & \text{if } y > \tau \\ 0 & \text{if } y \leq \tau \end{cases} \tag{4.5-3}$$

[1] The use of "detector" here does not refer to just the physical recording of photons, but rather the more general algorithmic problem of identifying a transmitted signal given one or more observables.

4.5.1 Direct Detection of OOK on the Poisson Channel

The Poisson channel with $f_{Y|X}(k|0) = K_b^k e^{-K_b}/k!$ and $f_{Y|X}(k|1) = (K_s + K_b)^k e^{-(K_s+K_b)}/k!$ results in a log-likelihood ratio of

$$\Lambda(k) = \ln\left(\frac{(K_s+K_b)^k e^{-(K_s+K_b)}}{k!} \cdot \frac{k!}{K_b^k e^{-K_b}}\right) = k\ln\left(1+\frac{K_s}{K_b}\right) - K_s \tag{4.5-4}$$

from which we can see that the ML rule becomes a threshold test as in Eq. (4.5-3), with threshold

$$\tau = \frac{K_s}{\ln\left(1+\frac{K_s}{K_b}\right)} \tag{4.5-5}$$

Denoting the expression "greatest integer less than or equal to x" by $\lfloor x \rfloor$, the bit error rate (BER) can be expressed as

$$P_b = \frac{1}{2} - \frac{1}{2}\sum_{k=0}^{\lfloor \tau \rfloor} \big(f(k|0) - f(k|1)\big) \tag{4.5-6}$$

which is a finite sum of easily computed terms. When no background is present, then $K_b = 0$, the threshold in Eq. (4.5-5) becomes $\tau = 0$, and the summation reduces to a single term:

$$P_b = \frac{1}{2}e^{-K_s}$$

The performance of OOK on a Poisson channel is shown in Fig. 4-20 for various values of K_b. Discontinuities in the slope of the performance curves for $K_b > 0$ occur when the optimum threshold changes by an integer amount. A more general derivation of the optimum threshold for the case of unequal a priori probabilities can be found in [63].

The BER may be expressed in terms of the photon efficiency, as this provides a direct way to determine system efficiency when operating at a prescribed error probability [64]. Since each OOK symbol corresponds to one bit and the average photons per symbol is $K_s/2$, the photon efficiency of uncoded OOK signaling can be expressed as

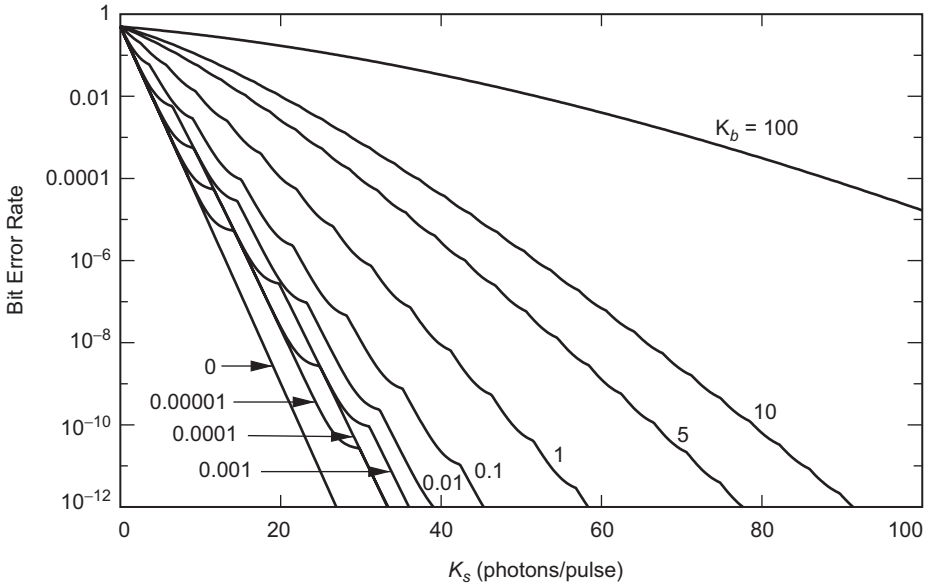


Fig. 4-20. BER versus signal level for uncoded OOK signaling on a Poisson channel, for various background levels.

$$\rho = \frac{2}{K_s} \text{ bits/photon} \tag{4.5-7}$$

This leads to the following alternate expression for BER, in terms of photon efficiency, when $K_b = 0$:

$$P_b = \frac{1}{2} e^{-2/\rho} \tag{4.5-8}$$

The result shows that operating at high values of ρ tends to result in high BERs. By generating plots of the error probability as a function of ρ, the photon efficiency of the system can be determined at any desired uncoded BER.

4.5.2 Direct Detection of PPM

At the receiver, the detected photons in each slot are counted. If hard decisions are used, the slot with the greatest photon count is declared to be the signal slot. This has been shown to be the ML decision for the channel models described in Section 4.2 [65]. On a continuous-output channel, the PPM symbol error probability P_s is the well-known performance [47,66] of an ML detector for M-ary orthogonal signaling:

$$P_s = 1 - \Pr\left(Y_1 = \max\{Y_1, \cdots, Y_M\} | X_1 = 1\right)$$

$$= 1 - \int_{-\infty}^{\infty} f_{Y|X}(y|1) \left[\int_{-\infty}^{y} f_{Y|X}(y'|0)dy'\right]^{M-1} dy \tag{4.5-9}$$

In Eq. (4.5-9), observations in the $M-1$ nonsignal slots are assumed to be independent. Equation (4.5-9) may be evaluated numerically by first producing a table lookup for the bracketed term, and then computing the outer integral numerically in the usual way.

When the detector outputs take on discrete values, there is a possibility of a tie for the maximum count. The solution for this problem has been derived in [7]. Suppose k photons are detected in the slot containing the pulse, l nonsignal slots also have count k, and the remaining nonsignal slots have count strictly less than k. Then the correct decision is made with probability $1/(l+1)$. Otherwise, an error is made. By summing over all possible values of k and l, it follows that

$$P_s = 1 - \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} \Pr\begin{bmatrix}\text{correct decision when}\\ l \text{ nonsignal slots tie}\\ \text{the signal slot for the}\\ \text{maximum count}\end{bmatrix} \times \Pr\begin{bmatrix}\text{exactly } l \text{ of } M-1\\ \text{nonsignal slots have}\\ \text{value } k, \text{ all others}\\ \text{smaller}\end{bmatrix} \tag{4.5-10}$$

$$\times \Pr\begin{bmatrix}\text{signal slot}\\ \text{has value } k\end{bmatrix} \tag{4.5-11}$$

$$= 1 - \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} \frac{1}{l+1} \binom{M-1}{l} f_{Y|X}(k|0)^l F_{Y|X}(k-1|0)^{M-l-1} f_{Y|X}(k|1) \tag{4.5-12}$$

where $f_{Y|X}(k|1)$ and $f_{Y|X}(k|0)$ denote the conditional probabilities that a received slot $Y_i = k$ when a 1 (pulse) or 0 (no pulse), respectively, is transmitted in the slot, and $F_{Y|X}(k|1)$ ($F_{Y|X}(k|0)$) denote the cumulative distributions. An extension of Eq. (4.5-12) to n-pulse PPM, $n \geq 2$, is straightforward, and involves a triple summation in place of the double summation. After some algebraic manipulation, this can be rewritten in a single summation as [67]

$$P_s = 1 - \frac{1}{M}\sum_{k=0}^{\infty} L(k)\left(F_{Y|X}(k|0)^M - F_{Y|X}(k-1|0)^M\right) \tag{4.5-13}$$

where $L(k)$ is the likelihood ratio. On a noiseless channel, $f_{Y|X}(0|0) = 1$, and the erasure probability is $f_{Y|X}(0|1)$, and thus, Eq. (4.5-13) can be simplified to

$$P_s = \frac{(M-1)f_{Y|X}(0|1)}{M} \tag{4.5-14}$$

Once the PPM symbol is detected, it is mapped to a string of $\log_2(M)$ bits via the inverse of the encoding mapping (see Fig. 4-6). There are $M/2$ symbol errors that will produce an error in a given bit in the string, and there are $M-1$ unique symbol errors. Thus, assuming all symbol errors are equally likely, the resulting bit error rate is

$$P_b = \frac{M}{2(M-1)}P_s \tag{4.5-15}$$

where P_s is given by Eq. (4.5-9) or Eq. (4.5-13).

4.5.2.1 Poisson Channel. For $K_b > 0$, the SER in Eq. (4.5-13) becomes

$$P_s = 1 - \sum_{k=0}^{\infty}\left(1 + \frac{K_s}{K_b}\right)^k \frac{e^{-K_s}}{M}\left(F_{Y|X}(k|0)^M - F_{Y|X}(k-1|0)^M\right) \tag{4.5-16}$$

where $F_{Y|X}(k|0) = \sum_{m=0}^{k} K_b^k e^{-K_b}/k!$. When $K_b = 0$, from Eq. (4.5-14) we have

$$P_s = \frac{(M-1)e^{-K_s}}{M} \tag{4.5-17}$$

and from Eq. (4.5-15) we see that $P_b = (1/2)e^{-K_s}$, which is independent of M and equal to that of OOK. This is shown in Fig. 4-21(a). As K_b increases, the dependence on M grows and performance for each M degrades, as seen in Figs. 4-21(b) and 4-21(c). It is not appropriate to interpret performance versus K_s as a measure of power efficiency, however. The average transmitter power is proportional to K_s/M photons per slot. Whereas low values of M in Figs. 4-21(a) through 4-21(c) produce a lower BER compared to high values of M, the situation is reversed in Figs. 4-21(d) through 4-21(f). The performance is also shown for bits/photon in Figs. 4-21(g) through 4-21(i), where it is also seen that

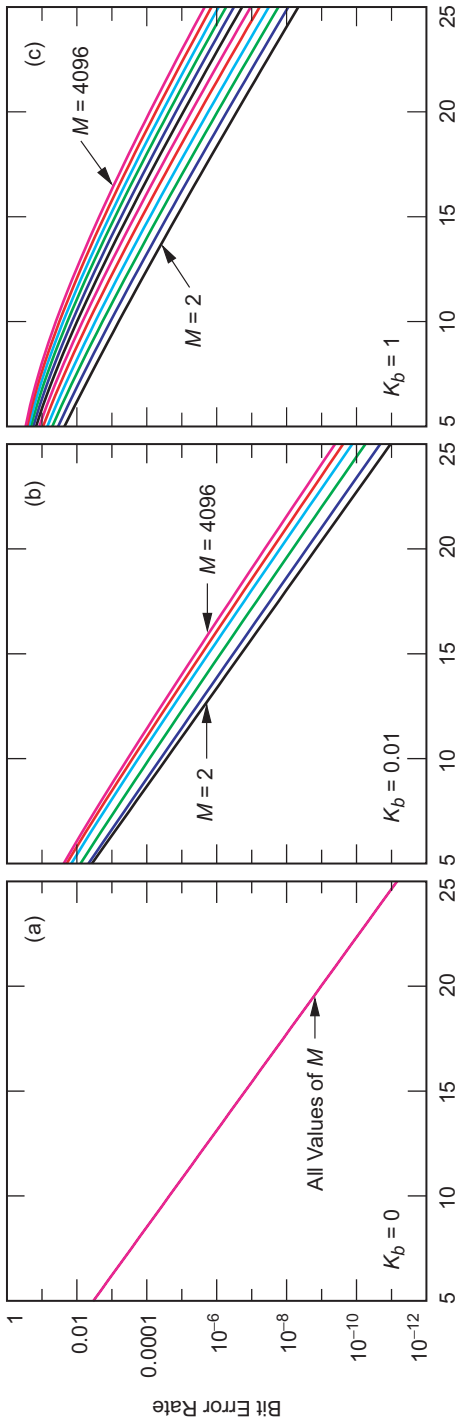


Fig. 4-21. BER of uncoded M-PPM signaling on a Poisson channel, versus (a-c) K_s photons/PPM symbol, (d-f) $P_{av} = K_s/M$ photons/slot, and (g-i) $\log_2 M/K_s$ bits/photon, for M = 2 to 4096 and K_b = 0, 0.01, 1.

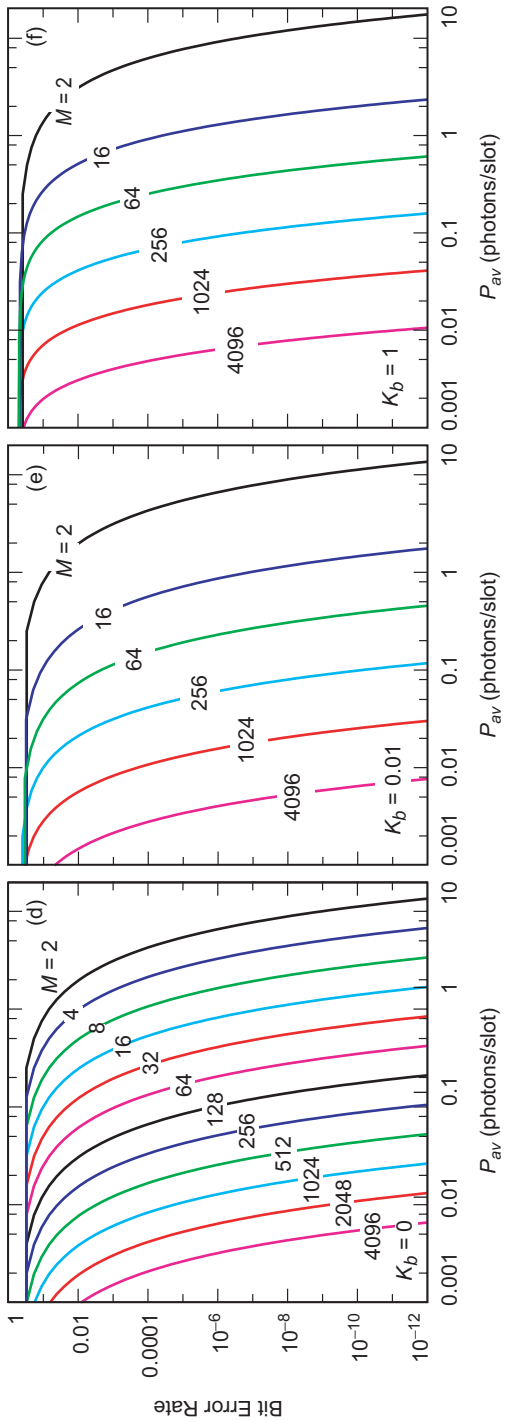


Fig. 4-21 (cont'd.).

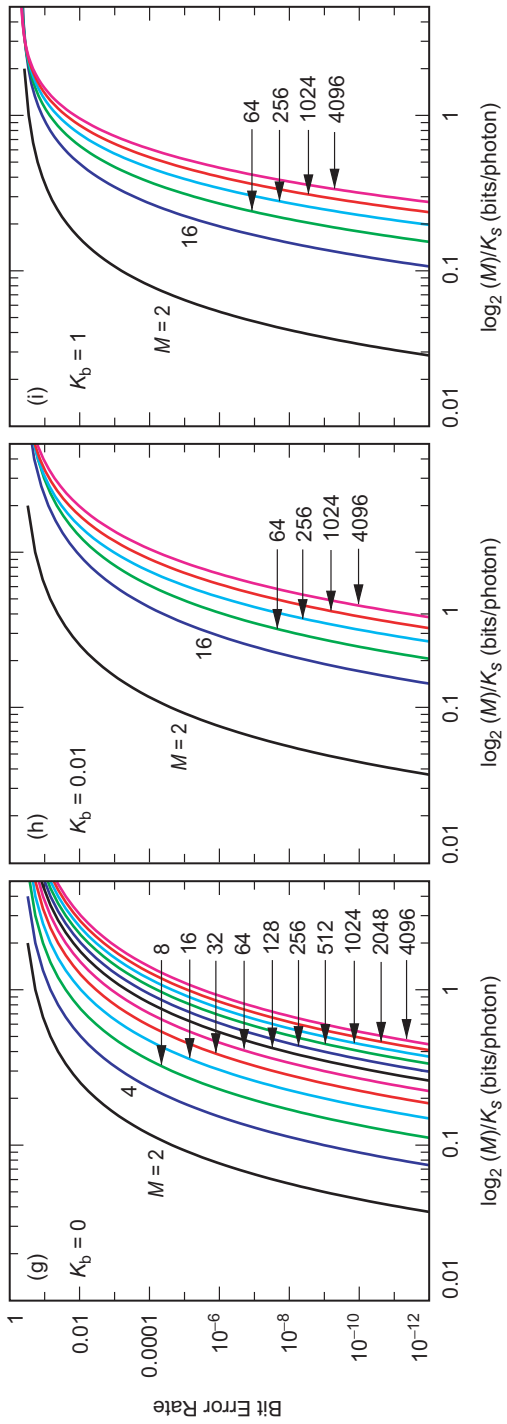


Fig. 4-21 (cont'd.).

high values of M are more photon efficient. Note that there is an average power difference of about 30 dB between $M = 2$ and $M = 4096$ [Figs. 4-21(d) through 4-21(f)], and a photon efficiency difference of about 10 dB [Figs. 4-21(g) through 4-21(i)]. The performance is also shown in terms of the slot SNR $\beta = K_s^2/K_b$ in Figs. 4-22(a) and 4-22(b), where we see little dependence on M; when plotted in terms of the bit SNR $\beta_b = K_s^2/(2K_b \log_2 M)$, the 30-dB gap manifests itself again, as seen in Figs. 4-22(c) and 4-22(d).

Expressing the BER in terms of the photon efficiency can be accomplished analytically when $K_b = 0$, by writing K_s in terms of ρ and substituting into Eq. (4.5-17), as originally shown in [64]:

$$P_b = \frac{1}{2} e^{-(\log_2 M/\rho)} = \frac{1}{2} M^{-(\log_2 e/\rho)} = \frac{1}{2} M^{-1.44/\rho} \tag{4.5-18}$$

4.5.2.2 AWGN Channel. The probability of symbol error is given by Eq. (4.5-9), which becomes

$$P_s = 1 - \int_{-\infty}^{\infty} \frac{1}{\sigma_1} \phi\left(\frac{x - m_1}{\sigma_1}\right) \Phi\left(\frac{x - m_0}{\sigma_0}\right)^{M-1} dx \tag{4.5-19}$$

$$= 1 - \int_{-\infty}^{\infty} \sqrt{\frac{\gamma}{\beta + \gamma}} \phi\left(\sqrt{\frac{\gamma}{\beta + \gamma}} \left(v - \sqrt{\beta}\right)\right) \Phi(v)^{M-1} dv \tag{4.5-20}$$

where $\phi(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$ is the standard normalized Gaussian probability density function, $\Phi(x)$ is its cumulative distribution function, and, as defined in Section 4.2.3, $\beta = (m_1 - m_0)^2/\sigma_0^2$ and $\gamma = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$. If $\sigma_1^2 = \sigma_0^2$, then $\gamma = \infty$ and Eq. (4.5-20) simplifies to

$$P_s = 1 - \int_{-\infty}^{\infty} \phi\left(x - \sqrt{\beta}\right) \Phi(x)^{M-1} dx \tag{4.5-21}$$

The BER P_b is then given by Eq. (4.5-15) and can be expressed in terms of the bit SNR $\beta_b = \beta/(2R_c)$, where in uncoded M-PPM there are $R_c = \log_2 M$ bits per PPM symbol.

For the case of homodyne detection, the BER can also be expressed in terms of ρ, as shown in [68]:

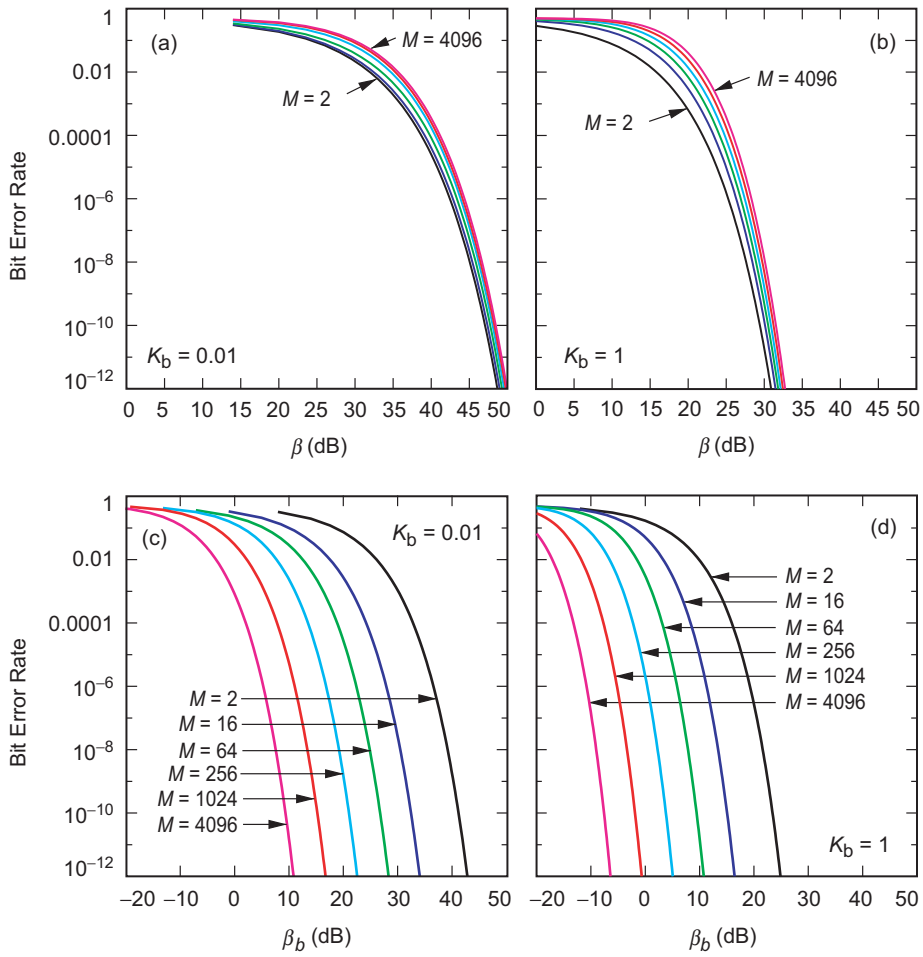


Fig. 4-22. BER of uncoded *M*-PPM signaling on a Poisson channel versus (a-b) slot SNR $\beta = K_s^2/K_b$ and (c-d) bit SNR $\beta_b = K_s^2/(2K_b \log_2 M)$, for M = 2 to 4096 and K_b = 0.01, 1.

$$P_s = 1 - \int_{-\infty}^{\infty} dx \frac{e^{-x^2}}{\sqrt{\pi}} \left[1 - Q\left(x\sqrt{2} + 2\sqrt{\frac{\log_2 M}{\rho}}\right)\right]^{M-1} \tag{4.5-22}$$

The limiting form of the bit error probability, as the number of symbols (hence the dimension of the symbol set) grows without bound, is in [68]. Using Eq. (4.5-15) to relate P_s to P_b, we get

$$\lim_{M\to\infty} P_b = \begin{cases} \dfrac{1}{2}; & \dfrac{2}{\rho} < \ln 2 \\ 0; & \dfrac{2}{\rho} > \ln 2 \end{cases} \tag{4.5-23}$$

It follows that arbitrarily low error probabilities can be achieved as M approaches infinity, as long as the inequality $\rho < 2/\ln 2 \approx 2.89$ bits/photon is satisfied.

4.5.3 Direct Detection of Combined PPM and WSK

Combined PPM and WSK is an orthogonal signaling scheme. Therefore, its performance is the same as PPM, but with NM dimensions instead of M. The probability of correct detection is given by Eqs. (4.5-9) and (4.5-13), but with M replaced by NM.

The information efficiency of this combined modulation scheme is $\rho = (\log_2 NM)/K_s = (\log_2 N + \log_2 M)/K_s$ bits/photon, but the photon rate remains $n_s = K_s/T$ photons/second. Hence the data rate is

$$R = \frac{\log_2 NM}{K_s}\,\frac{K_s}{T} = \frac{\log_2 N + \log_2 M}{T} \text{ bits/second} \tag{4.5-24}$$

For a given bandwidth, the throughput of combined PPM and WSK is higher than PPM alone. However, this modulation scheme requires N detectors, one for each wavelength, instead of just one.

4.5.4 Performance of Modulations Using Receivers Based on Quantum Detection Theory

We now present a description of receivers based on quantum detection theory, along with some specific examples of optical modulation formats that are of potential interest for deep-space applications.

4.5.4.1 Receivers Based on Quantum Detection Theory. In 1973, Kennedy described an ideal coherent receiver with significant communication advantages over ideal heterodyning, homodyning, or direct detection receivers [69]. Kennedy's receiver adds a local field prior to photodetection that exactly negates

the received field corresponding to one of the possible messages. The photodetector is guaranteed (ideally) to record zero photons in the case of this message. For the other message(s), the photodetector responds to the coherent sum of the signal field and the local field. For the binary coherent state detection problem, Kennedy's receiver obtains a 3-dB improvement over ideal homodyne or direct detection and a 6-dB improvement over ideal heterodyne detection.

Dolinar proposed a generalization of Kennedy's receiver structure [70] and optimized it to obtain a realization of an exactly optimum quantum measurement for the case of binary coherent state signals [70]. The optimum receiver for the binary case eventually nulls the field corresponding to the message that is ultimately selected as more likely. Initially, however, the receiver does not preselect one of the two messages for nulling. The added local field changes abruptly with each observed count from the photodetector as the message likelihoods are updated.

Dolinar also described a near-optimum conditionally nulling receiver for the M-ary PPM detection problem [71]. This receiver sequentially, slot by slot, based on prior observed counts, decides either to add or not add a local field designed to null the signal field that corresponds to a pulse in the current slot.

Since an ideal nulling or conditionally nulling receiver is a mathematical artifact that can never be precisely realized in practice, non-ideal versions that permit phase errors in the nulling signal have also been modeled and analyzed by Dolinar [71] and Vilnrotter and Rodemich [63].

The performance of the optimum quantum receiver has been determined in the literature, particularly for the cases of noiseless reception of binary and higher-dimensional signals and of noisy reception of binary signals. The performance literature has recently extended to higher dimensional signals operating in the presence of noise. A summary of these results follows.

4.5.4.1.1 Quantum Detection with Binary Decisions. The problem of determining which of two possible quantum states is present, representing binary hypotheses, has been addressed in detail for the case of both noisy and noiseless reception [3,5,75]. Under hypothesis H_0, the received electromagnetic field is in a mixture of states governed by the density operator ρ_0, and under hypothesis H_1, it is governed by ρ_1. Assume equal a priori probabilities, so that $\Pr(H_0) = \Pr(H_1) = 1/2$. Suppose a receiver applies the "detection operators" defined in terms of an appropriate set of orthonormal basis states $\{|b_n\rangle\}$:

$$\Pi_0 = \sum_{n:\, b_n \in I_0} |b_n\rangle\langle b_n|$$
$$\Pi_1 = \sum_{n:\, b_n \in I_1} |b_n\rangle\langle b_n| \tag{4.5-25}$$

These detection operators effectively partition the outcome space R into two disjoint regions I_0 and I_1, such that $R = I_0 \cup I_1$, $I_0 \cap I_1 = \emptyset$. The detection operators are projection operators that partition the outcome space into an exhaustive set of disjoint decision regions. For binary detection the outcome space contains only two regions, $\Pi_0 + \Pi_1 = \mathbf{1}$; hence only one of the detection operators needs to be applied. If the outcome of the measurement falls within the region I_1, H_1 is selected. Otherwise, the receiver chooses H_0.

The conditional probabilities of a correct decision are [3,5,75]

$$\Pr(C|H_1) = Q_d = \mathrm{Tr}\,[\rho_1\Pi_1]$$
$$\Pr(C|H_0) = 1 - Q_0 = 1 - \mathrm{Tr}\,[\rho_0\Pi_1] \tag{4.5-26}$$

It follows that, for equally probable signals, the average probability of correct decision is

$$P_c = \frac{1}{2}\big\{\mathrm{Tr}\,[\rho_1\Pi_1] + 1 - \mathrm{Tr}\,[\rho_0\Pi_1]\big\}$$
$$= \frac{1}{2}\big\{1 + \mathrm{Tr}\,[(\rho_1 - \rho_0)\Pi_1]\big\} \tag{4.5-27}$$

and the average probability of error can be expressed as

$$P_b = 1 - P_c = \frac{1}{2}\big\{1 - \mathrm{Tr}\big[(\rho_1 - \rho_0)\Pi_1\big]\big\} \tag{4.5-28}$$

It is clear from the form of Eq. (4.5-28) that the probability of error is minimized by choosing the projection operator Π_1 to maximize the quantity $\mathrm{Tr}\,[\rho_1 - \rho_0)\Pi_1]$.

4.5.4.1.2 The Optimum Measurement for Binary Detection. Helstrom demonstrated that the detection operator which maximizes the quantity $\mathrm{Tr}\,[(\rho_1 - \rho_0)\Pi_1]$ is of the form [5]

$$\Pi^* = \sum_{k:\eta_k \geq 0} |\eta_k\rangle\langle\eta_k| \tag{4.5-29}$$

where $|\eta_k\rangle$ are the eigenvectors and η_k the eigenvalues of the operator $(\rho_1 - \rho_0)$, that is,

$$(\rho_1 - \rho_0)|\eta_k\rangle = \eta_k|\eta_k\rangle \tag{4.5-30}$$

The probability of error for this optimum projector is

$$P_b^* = \frac{1}{2}\left\{1 - \mathrm{Tr}\left[(\rho_1 - \rho_0)\Pi^*\right]\right\}$$

$$= \frac{1}{2}\left\{1 - \sum_{k:\eta_k > 0} \eta_k\right\} \tag{4.5-31}$$

as shown in [3,5,75]. For "pure states" represented by $|\psi_0\rangle$ and $|\psi_1\rangle$, the density operators are $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$. The eigenvectors of the operator $(\rho_1 - \rho_0)$ are now linear combinations of $|\psi_0\rangle$ and $|\psi_1\rangle$, yielding an expression of the form

$$|\eta_k\rangle = z_{k0}|\psi_0\rangle + z_{k1}|\psi_1\rangle, \qquad k = 0,\ 1 \tag{4.5-32}$$

where the coefficients are determined by substituting into Eq. (4.5-30) and equating coefficients. The optimum projection operator is simply $\Pi^* = |\eta_1\rangle\langle\eta_1|$, and its application yields the detection and false-alarm probabilities

$$Q_d = \mathrm{Tr}\left[\rho_1\Pi^*\right] = \frac{1}{2}\left[1 + \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-33}$$

$$Q_0 = \mathrm{Tr}\left[\rho_0\Pi^*\right] = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-34}$$

The probability of correct detection becomes

$$P_c^* = \frac{1}{2}\left[1 + \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-35}$$

while the probability of error is

$$P_b^* = 1 - P_c^* = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-36}$$

Thus, for the case of binary decisions, the error probability associated with the optimum projection measurement depends on the overlap between the two quantum states.

4.5.4.1.3 Orthogonal Signal States. With orthogonal signal states, $\langle\psi_0|\psi_1\rangle = 0$. Substituting into Eq. (4.5-36), we find that $P_b^* = 0$. Therefore, truly error-free communication could be achieved if practical orthogonal signal states could be generated, for example, if the number states $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = |1\rangle$ could somehow be prepared [75].

4.5.4.2 Performance of Representative Modulations. Performance of the following detection schemes for binary modulation has been described in [75], repeated here for reference.

4.5.4.2.1 On–Off Keying. For on–off keying we let

$$|\psi_0\rangle = |0\rangle$$

and

$$|\psi_1\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{(n!)^{1/2}}|n\rangle$$

where the average number of photons is $K_s = |\alpha|^2$ for the optical pulse, and zero for the ground state. The squared magnitude of the overlap between the two states is $|\langle\psi_0|\psi_1\rangle|^2 = e^{-|\alpha|^2}$, yielding $P_b^* = 1/2\left[1 - \sqrt{1 - e^{-|\alpha|^2}}\right] = 1/2\left[1 - \sqrt{1 - e^{-K_s}}\right]$ for optimum quantum detection. In comparison, the error probability for photon-counting detection was shown to be $P_b = (1/2)e^{-|\alpha|^2} = (1/2)e^{-K_s}$.

4.5.4.2.2 Optical BPSK. In this case of optical BPSK, we have two coherent states with the same average photon energy, but π radians out of phase. The signal states are defined as $|\psi_0\rangle = |\alpha\rangle$ and $|\psi_1\rangle = |\beta\rangle$ where the complex amplitudes are related as $\beta = -\alpha$, and the average number of signal photons

is $K_s = |\alpha|^2$. Therefore, $|\langle\psi_0|\psi_1\rangle|^2 = e^{-4K_s}$, and the error probability becomes $P_b^* = 1/2\left[1 - \sqrt{1 - e^{-4K_s}}\right]$. A physically realizable receiver structure that achieves this error probability has been devised and analyzed by Dolinar [70].

4.5.4.2.3 Near-Optimum Optical BPSK. Exponentially optimum (or "near optimum") performance can be obtained by adding a local field of the same amplitude, in phase, to the received field followed by photon counting, as shown by Kennedy [69]. With this technique, the field amplitude under one hypothesis is shifted to the ground state, but doubled under the other. The error probability for this near-optimum detection scheme is $P_b = (1/2)e^{-4|\alpha|^2} = (1/2)e^{-4K_s}$.

4.5.4.2.4 Coherent Detection: The Classical Limit. If we add a local field of great amplitude, in phase, to the received field and detect the resulting sum field using classical energy detection, then the performance of the classical coherent optical detector is obtained. For optical BPSK modulation, the error probability for this receiver is given by [72] $P_b = Q\left(\sqrt{4|\alpha|^2}\right) = Q\left(\sqrt{4K_s}\right)$, where $Q(x) \equiv (1/\sqrt{2\pi})\int_x^\infty e^{-y^2/2}dy$. Performance curves for the various modulation formats and detection options are shown in Fig. 4-23.

4.5.4.2.5 Optimum Binary Detection in the Presence of Background Radiation. When background radiation is present, the received field is in a mix-

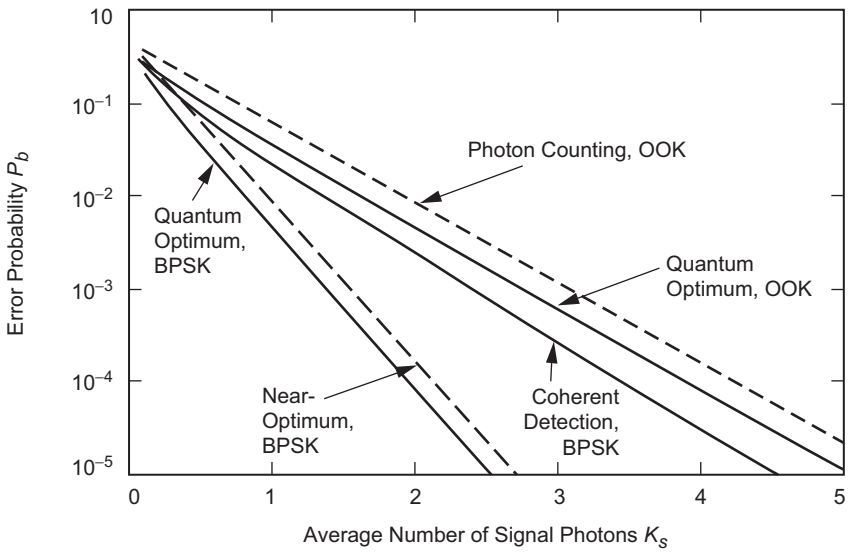


Fig. 4-23. Performance of binary optical receivers versus average signal photons per slot, assuming no background radiation.

ture of states described by an appropriate density operator. The probability of error can be calculated by finding the significant eigenvalues of the difference operator $(\rho_1 - \rho_0)$ and applying the formula for the error probability given by Eq. (4.5-31). Equivalently, the matrix representation of the difference operator can be diagonalized, the projection operator that selects only the positive eigenvalues applied, and the trace of the resulting diagonal matrix determined.

The following analysis parallels the derivation in [3] for on–off keying with noise. For this scenario, a density matrix ρ_1 can be defined for the signal and noise case. The controllable values are the size of the matrix, the average number of noise photons, and the average number of signal photons. The elements of this density matrix in the number state basis is, for $m \geq n$,

$$\langle n|\rho_1|m\rangle = (1-v)\sqrt{\frac{n!}{m!}}v^m \left(\frac{\mu^*}{N}\right)^{m-n} e^{\left(-(1-v)|\mu|^2\right)} L_n^{m-n}\left[-(1-v)\frac{|\mu|^2}{v}\right]$$

and for $m < n$,

$$\begin{aligned} \langle n\,|\rho_1|\,m\rangle &= \langle m\,|\rho_1|\,n\rangle^* \\ v &= \frac{N}{(N+1)} \end{aligned} \tag{4.5-37}$$

where N represents the average number of noise photons, μ is the complex envelope of the signal, $|\mu|^2$ is the average number of signal photons, and $L_n^{m-n}(x)$ is a Laguerre polynomial. A density operator ρ_0 for the null hypothesis, or noise-only case, can be generated using

$$\rho_0 = \sum_{n=0}^{\infty} (1-v)v^n \,|n\rangle\,\langle n| \tag{4.5-38}$$

Next, the resulting difference matrix is diagonalized using the formula $[\langle\eta_k|\rho_1 - \rho_0|\eta_k\rangle]$. The probability of error is found from this diagonalized matrix. The probability of correct detection is found by adding all the positive diagonal terms, yielding the probability of error

$$P_b = 1 - P_c = \frac{1}{2}\left\{1 - \sum_{k:\eta_k>0} \eta_k\right\} \tag{4.5-39}$$

For comparison, the performance of classical coherently detected BPSK signals observed in the presence of noise is called "threshold detection," with error probability $P_b = Q\left(\sqrt{4K_s/(2N+1)}\right)$, which is seen to be similar to the expression for the noiseless case, but with K_s replaced by $K_s/(2N+1)$. Note that for high background levels, the performance of the classical threshold detector is nearly as good as that of the optimum quantum detector. Hence, the physically realizable classical detector is a good approximation to the optimum quantum detector under conditions of high background radiation.

4.5.4.2.6 Multiple Hypotheses: Orthogonal Envelopes. The performance of the optimum quantum receiver for the case of equal-energy, equally probable signals is considered [3,5,75]. The signals are assumed to have orthogonal classical envelopes, which means that the classical complex envelopes $S_k(t)$ obey the condition

$$\int_0^T S_k^*(t) S_m(t) dt = 0, \quad k \neq m \tag{4.5-40}$$

In this case there are assumed to be M hypotheses, represented by M orthogonal classical envelopes modulating electromagnetic plane waves normally incident on the receiving aperture, and with temporal variation proportional to $S_k(t)$. PPM is an example of this modulation format where a single pulse is placed in one of M consecutive slots. The kth aperture-field mode is assumed to be matched to the kth signal, such that when the kth signal is present, the state of the aperture field is in a coherent state described by the "product state"

$$|\alpha_k\rangle = \prod_{j=1}^{M} |\alpha_{k,j}\rangle = |\alpha_{k,1}\rangle |\alpha_{k,2}\rangle \cdots |\alpha_{k,M}\rangle \tag{4.5-41}$$

where each of the $|\alpha_{k,j}\rangle$ is a coherent state associated with an individual mode, with $|\alpha_{k,j}| = |\alpha|\delta_{k,j}$, and where $K_s = |\alpha|^2$ is the average number of signal photons in each signal.

Since the M product states are linearly independent, the optimum strategy for minimizing the average probability of error is to project the received signal state onto M orthonormal measurement states spanning the same subspace and select the signal corresponding to the measurement state with the greatest projection. The state-space interpretation is similar to the binary case; because of symmetry, the M orthogonal measurement states are aligned with the M signal states in such a way as to maximize the projection of each state onto the

corresponding measurement state, thus minimizing the probability of error. It has been shown [70] that for equally likely signals, the minimum symbol error probability for the optimum quantum receiver is

$$P_s = \frac{M-1}{M^2} = \left[\sqrt{1+(M-1)e^{-K_s}} - \sqrt{1-e^{-K_s}}\right]^2 \tag{4.5-42}$$

The error probability is given by $P_s = 1/2\left[1-\sqrt{1-|\langle\psi_1|\psi_2\rangle|^2}\right]$ as before, but now $|\langle\psi_1|\psi_2\rangle|^2 = e^{-2K_s}$, as compared to e^{-K_s} for the corresponding overlap with on–off keying, and e^{-4K_s} for BPSK signals. For comparison, the symbol error probability achieved by the photon-counting receiver in the absence of background radiation is given by Eq. (4.5-17). These results are summarized in [75].

For high-dimensional signaling and modest error probabilities on the order of $P_s \cong 10^{-3}$, conventional photon counting performs approximately 1.3 dB worse than the optimum quantum receiver in the absence of background radiation, but this performance gap increases at the lower error probabilities.

The performance of PPM and other high-dimensional signal envelopes in the presence of noise is not known in general; however, numerical techniques have been developed recently and applied to three-dimensional signals [73–75]. The performance of the optimum quantum receiver for the case of ternary signals in the presence of noise has been obtained. This signaling scheme is defined by three signal states, namely, $|\alpha\rangle$, $|-\alpha\rangle$, and $|0\rangle$. Further work is required to extend this approach to arbitrary dimensional signal sets operating in the presence of external noise.

4.6 Optical Channel Capacity

Shannon demonstrated in [76] that for any communications channel, as long as the rate of transmitting information is less than some constant, C, it is possible to make the average error probability arbitrarily small by coding messages into a large enough class of signals. The constant C is called the *capacity* of the channel. Characterizing the capacity of the optical channel provides a useful bound on the data rates achievable with any modulation and coding scheme, thus serving as a benchmark for assessing the performance of a particular design [10–20,22–27,43,48].

The capacity will be a function of the received optical signal and noise powers, the modulation, and the detection method. In this section, we assume the

modulation is PPM. The loss in capacity by restricting the modulation to PPM is small (at most a few tenths of a dB) in the low average power regime where the deep-space optical channel currently operates.

We divide the capacity into two categories depending on the type of information provided to the decoder by the receiver. In one case, the receiver makes estimates of each PPM symbol, passing these estimates, or *hard* decisions, on to the decoder. In this case, the (hard-decision) capacity may be expressed as a function of the probability of symbol error, P_s, derived for several channel models in Section 4.5.

In the second case, the receiver makes no explicit symbol decision, but passes on slot counts (integrals of the received signal in each slot), or *soft* decisions, directly to the decoder. In this case, the (soft-decision) capacity may be expressed as a function of the channel statistic $f_{Y|X}$, presented for several channel models in Section 4.2.2. The soft-decision capacity is at least as large as the hard-decision capacity because the slot counts provide additional information to the decoder.

4.6.1 Capacity of the PPM Channel: General Formulas

The hard-decision PPM channel is an M-ary input, M-ary output, symmetric channel with capacity given by [66]

$$C = \log_2 M + (1 - P_s)\log_2(1 - P_s) + P_s \log_2 \left(\frac{P_s}{M-1} \right) \text{ bits/PPM symbol} \tag{4.6-1}$$

where P_s is the probability of incorrect PPM symbol detection. The function P_s is provided for several channel models in Section 4.5.

The soft-decision capacity is given by

$$C = E_{\mathbf{Y}} \log_2 \left[\frac{ML(Y_1)}{\sum_{j=1}^{M} L(Y_j)} \right] \text{ bits/PPM symbol} \tag{4.6-2}$$

an expectation over $\mathbf{Y}$, where $L(y) = f_{Y|X}(y|1)/f_{Y|X}(y|0)$ is the channel likelihood ratio, and the Y_j have density $f_{Y|X}(y|1)$ for $j = 1$ and density $f_{Y|X}(y|0)$ otherwise.

The M-fold integration in Eq. (4.6-2) is often intractable. However, it is straightforward to approximate the expectation by a sample mean. A quick approximation follows from the lower bound

$$C \geq E \log_2 \left[\frac{M}{1 + \dfrac{M-1}{L(Y_1)}} \right] \text{ bits/PPM symbol} \tag{4.6-3}$$

which is a good approximation for large M, reducing the M-fold integration (or set of M-dimensional vector samples) needed to evaluate Eq. (4.6-2) to a one-dimensional integral (or set of scalar samples).

4.6.2 Capacity of Soft-Decision PPM: Specific Channel Models

4.6.2.1 Poisson Channel. We consider first the Poisson channel. The behavior of the case $K_b = 0$ has a particularly simple form. When $K_b = 0$, we have from Eqs. (4.2-17) and (4.2-18)

$$L(k) = \begin{cases} e^{-K_s} & k = 0 \\ \infty & k > 0 \end{cases}$$

and Eq. (4.6-2) reduces to

$$C = (\log_2 M)\left(1 - e^{-K_s}\right) \text{ bits/PPM symbol}$$

When $K_b = 0$, only signal photons are detected. If any signal photons are detected, the signal is known exactly. If no photons are detected, all M candidate symbols are equally likely. Since the received statistic takes binary values, the soft- and hard-decision capacities are equal.

When $K_b > 0$, we have $L(k) = e^{-K_s}\left(1 + [K_s/K_b]\right)^k$, and Eq. (4.6-2) becomes

$$C = \log_2(M)\left(1 - \frac{1}{\log_2 M} E_{Y_1,\cdots,Y_M} \log_2 \left[\sum_{i=1}^{M} \left(1 + \frac{K_s}{K_b}\right)^{(Y_j - Y_1)} \right] \right)$$

which is expressed in bits/slot. The case $K_b = 1$ is illustrated in Fig. 4-24 as a function of average power $P_{av} = K_s/M$ for a range of M. In the plot, an average

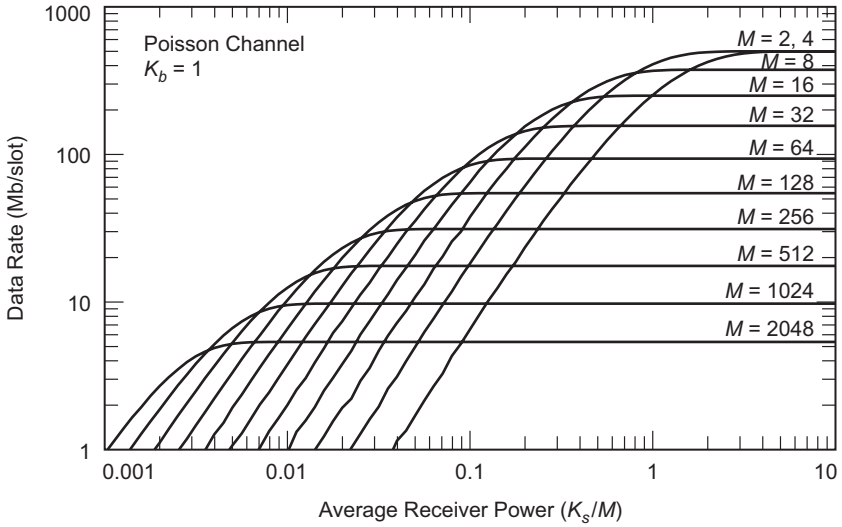


Fig. 4-24. Capacity of *M*-PPM on a Poisson channel, when $K_b = 1$.

power constraint would be represented by a vertical line. A peak constraint can be shown [77] to result in an upper limit on the PPM order. Hence, the maximum data rate subject to both peak and average power constraints can be identified using Fig. 4-24.

4.6.2.2 AWGN Channel. For the AWGN channel, the likelihood ratio reduces to

$$L(y) = \sqrt{\frac{\gamma}{\beta+\gamma}} \exp\left[\frac{\beta v^2 + 2\gamma\sqrt{\beta}v - \gamma\beta}{2(\beta+\gamma)}\right]$$

where $v = (y - m_0)/\sigma_0$ (recall β, γ are defined in Section 4.2.3). The capacity reduces to

$$C = \log_2 M - E\log_2 \sum_{j=1}^{M} \exp\left[\frac{(Y_j - Y_1)\big(\beta(Y_j + Y_1) + 2\gamma\sqrt{\beta}\big)}{2(\beta+\gamma)}\right]$$

bits per PPM symbol, or, when $\sigma_1^2 = \sigma_0^2$,

$$C = \log_2 M - E\log_2 \left[\sum_{j=1}^{M} e^{\sqrt{\beta}(Y_j - Y_1)}\right] \text{ bits/PPM symbol} \qquad (4.6\text{-}4)$$

4.6.3 Hard-Decision Versus Soft-Decision Capacity

The gap between hard- and soft-decision capacity for the optical channel is not fixed, but varies with the channel model and operating conditions. For example, the gap for the Poisson channel is zero when $K_b = 0$ (hard and soft decisions are equivalent) and increases to several dB with increasing K_b. In this section, we illustrate the gap for several channel models.

Figure 4-25 compares capacities for hard- and soft-decision AWGN channels for the case of $M = 256$. A similar comparison of capacities is shown in Fig. 4-26 for the hard-output and soft-output WMC channels. The capacity curves for both the AWGN and the WMC channels show that a minimum value of ρ_b is reached at a nonzero code rate. Unlike the soft-output channels, which exhibit monotonically better efficiency in terms of the bit-normalized SNR parameter ρ_b as the code rate (and hence the capacity per channel use) is reduced toward zero, the bit-normalized SNR efficiency of the hard-output channel worsens if the capacity is lowered below about 4 bits per channel use. This implies that an optimum code rate of about 1/2 will achieve the lowest ρ_b for the hard-output channel, while the soft-output channel achieves lowest ρ_b in the limit as the code rate goes to 0.

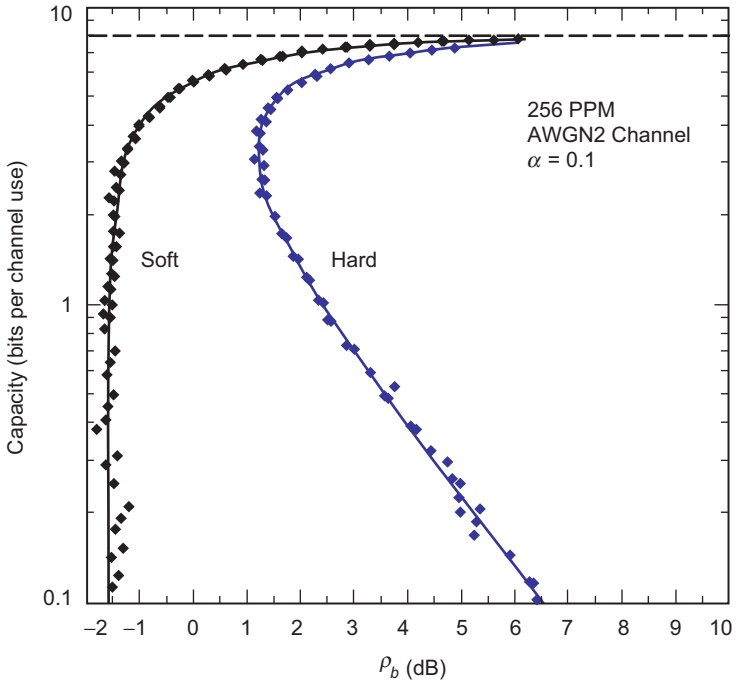


Fig. 4-25. Capacity of 256-PPM on hard-output and soft-output AWGN channels.

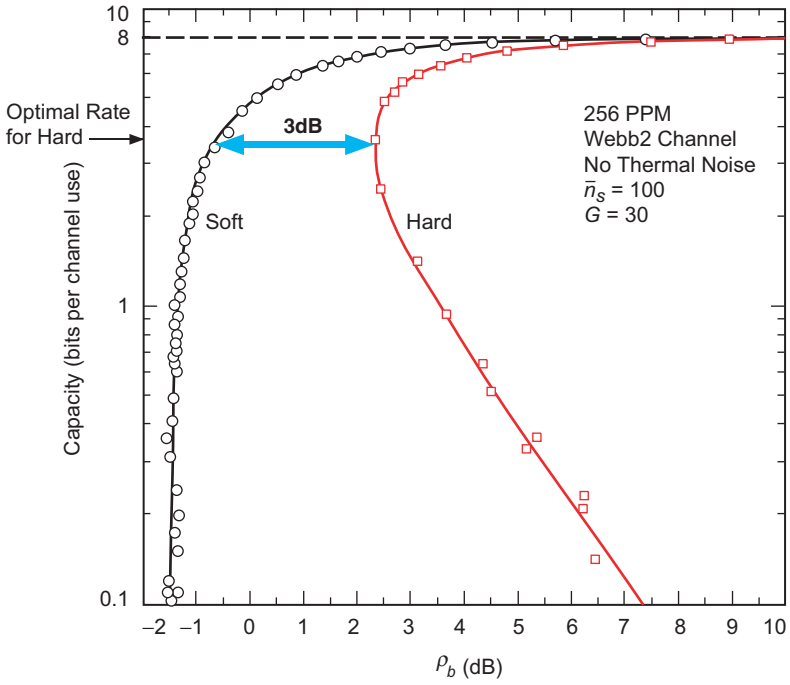


Fig. 4-26. Capacity of 256-PPM on hard-output and soft-output WMC channels.

Another comparison of capacity for the hard-output and soft-output WMC models is shown in Fig. 4-27, this time plotted versus $\bar{n}_b$. The hard-output capacity in this figure is based on Eq. (4.6-1) and was computed in [14] for a general WMC+Gaussian channel that also models the effects of thermal noise.

4.6.4 Losses Due to Using PPM

What loss is incurred by restricting the modulation to PPM? PPM is essentially a binary modulation code with a duty cycle $1/M$, and a single pulse (binary 1) in each (synchronized) window of M slots. Suppose we were to replace PPM with a binary modulation code with duty cycle $1/M$ but no constraint on the distribution of pulses. What gains are available by allowing an arbitrary pulse distribution?

The capacity of a memoryless channel with input restricted to duty cycle $1/M$ is

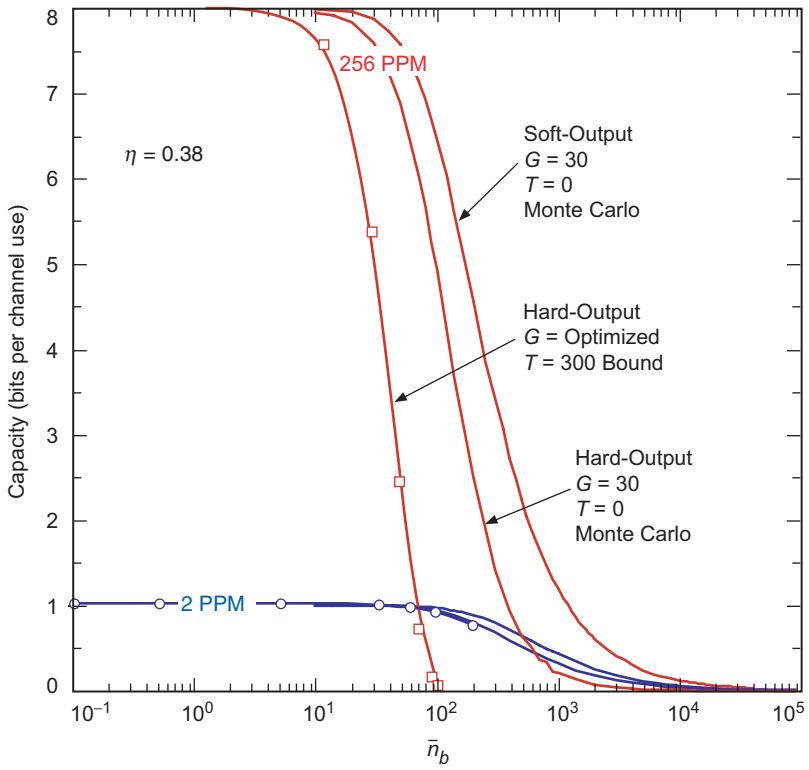


Fig 4-27. Comparison of capacity for hard-output and soft-output WMC model.

$$C_{\text{OOK}} = \frac{1}{M} E_Y \log \frac{f_{Y|X}(Y|0)}{f_Y(Y)} + \frac{M-1}{M} E_Y \log \frac{f_{Y|X}(Y|1)}{f_Y(Y)} \tag{4.6-5}$$

where

$$f_Y(y) = \frac{1}{M} f_{Y|X}(y|0) + \frac{M-1}{M} f_{Y|X}(y|1)$$

is the probability mass function for a randomly chosen slot.

Let C^* be the capacity of the PPM channel optimized over the choice of PPM order (PPM orders are implicitly constrained to be powers of 2), and C^*_{OOK} the capacity of the duty-cycle-constrained channel optimized over the duty cycle (the duty cycle may take any positive real value). Figure 4-28 illustrates C^*_{OOK}/C^* for the Poisson channel as a function of the average power for a range of back-

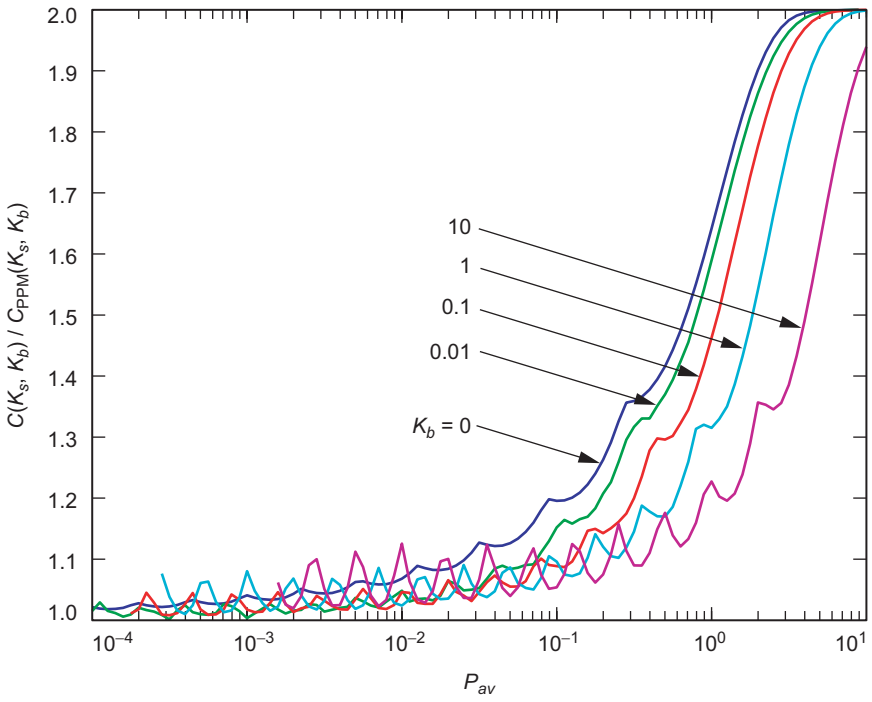


Fig. 4-28. Relative loss due to using PPM, C^*_{OOK}/C^*, Poisson channel, $K_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

ground noise levels. This represents the potential gain in using an arbitrary duty-cycle constraint relative to PPM. The gains are larger for high average power, corresponding to small PPM orders, and for smaller background noise levels. We can potentially double the capacity for moderate to high average power. We note, however, in this discussion we have not specified codes that achieve arbitrary duty cycles. There are systematic methods to construct such codes, e.g., [78], but we will not explore their use here. We provide the results shown in Fig. 4-28 to demonstrate regions where their use should be explored. Since the deep-space optical channel is typically in the lower average power regime and losses in constraining the modulation to PPM are small in this area, in the remainder of the chapter we focus on results for PPM.

4.6.5 Capacity of the Binary Channel with Quantum Detection

The capacity of the binary OOK channel with quantum detection and with no external noise has been determined in [75]. Note that for the "noiseless" quantum model photon counting leads to an erasure channel whereas optimum quantum decoding results in a binary symmetric channel (BSC). For an arbi-

trary rotation of the measurement states with respect to the signal states, as described in [75], the transition probabilities are not equal, and, hence, a generalized (asymmetric) binary channel model must be considered. The capacity of the binary channel can be found by computing the mutual information between input and output for each rotation of the measurement states, starting with photon counting where one of the measurement states is aligned with the ground state, and computing the mutual information as a function of symbol input probability, β, for each rotation away from this configuration. For each rotation, the maximum of the mutual information as a function of β is recorded. The global maximum of the mutual information over all input probabilities and rotations is the capacity of the binary channel.

The input alphabet is denoted by A, and the output alphabet is denoted by B. The input alphabet consists of the two symbols $a_1 = 0$ and $a_2 = 1$. Likewise, the output can take on one of two values, namely $b_1 = 0$ and $b_2 = 1$. The probability that a 0 is transmitted is β, whereas the probability of a transmitted 1 is $1 - \beta \equiv \overline{\beta}$. The probability that b_2 is received (given that a_1 was transmitted) is p, while the probability that b_1 is received (given that a_2 was transmitted) is q.

The mutual information for this binary channel can be expressed as

$$I(A;B) = \left[\left(\beta\bar{p} + \overline{\beta}q\right)\log\left(\frac{1}{\beta\bar{p} + \overline{\beta}q}\right) + \left(\beta p + \overline{\beta}\bar{q}\right)\log\left(\frac{1}{\beta p + \overline{\beta}\bar{q}}\right)\right]$$

$$- \left[\beta\left(p\log\frac{1}{p} + \bar{p}\log\frac{1}{\bar{\beta}}\right) + \overline{\beta}\left(q\log\frac{1}{q} + \bar{q}\log\frac{1}{\bar{q}}\right)\right] \qquad (4.6\text{-}6)$$

Note that the mutual information of the erasure and the BSC can be obtained by setting $p = 0$ and $p = q$, respectively.

The capacity of the quantum channel was determined by starting with a rotation angle of zero between the ground state and its measurement state (corresponding to photon counting, as we have shown above), and computing the mutual information defined in Eq. (4.6-6) as a function of β, $0 \leq \beta \leq 1$ for each rotation in the signal plane, until the measurement state corresponds to the signal state. Since different rotations yield different projections onto the measurement states, the values of p and q change with each rotation.

Examples of mutual information and capacity for the binary channel with OOK modulation are shown in Fig. 4-29, as a function of the input probability β,

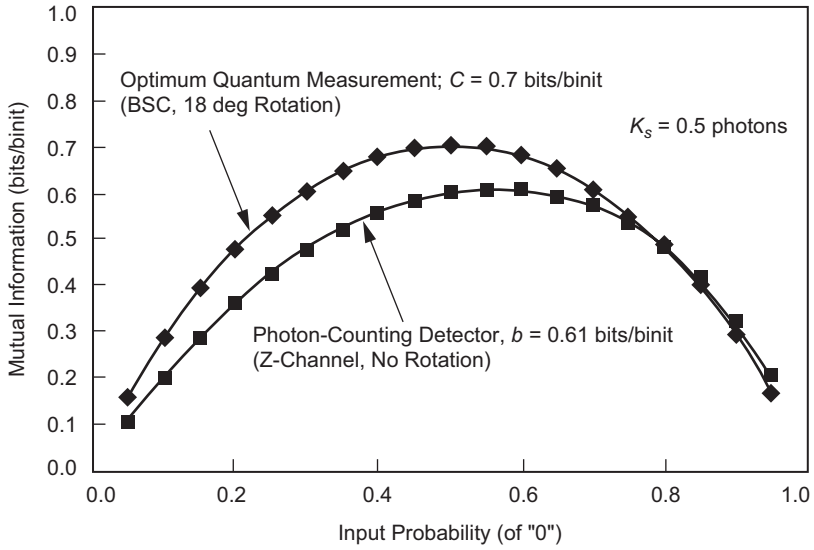


Fig. 4-29. Mutual information and capacity of the binary channel, with quantum and "classical" detection.

for an average value of one photon per symbol (or two photons per signal pulse). Only the limiting cases of optimum quantum measurement and photon counting are included. The error probabilities are approximately 0.1025 and 0.18 for quantum and direct detection, respectively. The global maximum value of mutual information was found to occur with optimum quantum measurement, at an input probability of $\beta = 0.5$. With photon-counting detection, for which the asymmetric z-channel is the correct representation, the maximum mutual information occurs at a higher value of input probability, namely at $\beta = 0.55$. The value of the maximum mutual information was found to be 0.7 bits/binit for quantum detection and 0.61 bits/binit for photon counting, verifying that optimum quantum detection achieves higher capacity.

4.7 Channel Codes for Optical Modulations

The constrained codes, or modulations, introduced in Section 4.3 enforce physical constraints and achieve desired peak-to-average power ratios, but their performance is far from the theoretical limits given in Section 4.6. In this section, we examine the application of Reed–Solomon, convolutional, turbo-like serial and parallel concatenated codes, and low-density parity check (LDPC) codes to the optical channel.

4.7.1 Reed–Solomon Codes

Since their introduction in 1960, Reed–Solomon (RS) codes [79] have become one of the most ubiquitous error-correcting codes. They have found applications in storage devices (tape, compact disk, digital video disc), bar codes, wireless communications (cellular telephones, deep-space RF communications), digital television, and high speed modems (digital subscriber line, DSL), in addition to optical communications.

An RS(n, k) code is a linear block code which encodes every block of k data symbols into n code symbols, where each symbol is an element of the Galois field with $q = n + 1$ elements, denoted GF(q) [79]. Most commonly, q is a power of 2, $q = 2^s$, in which case each symbol is conveniently represented by s bits. Thus, the code can also be viewed as a (sn, sk) binary code.

In systematic form, a codeword of RS(n, k) contains k systematic (unchanged data) symbols and $n - k$ parity symbols. RS codes are maximum distance separable, meaning that they have the largest minimum distance, $d_{min} = n - k + 1$, among all linear (n, k) codes defined over GF(q). In some sense then, RS codes are optimal for their block length.[2] An RS(n, k) code can correct any pattern of $t = (d_{min} - 1)/2 = (n - k)/2$ errors; alternatively, it can correct any pattern of $2t$ erasures.

The encoder for RS codes operates as follows. The sk-bit message at the input to the RS encoder can be written as coefficients of a polynomial of degree $k - 1$:

$$u(X) = u_0 + u_1 X + u_2 X^2 + \cdots + u_{k-1} X^{k-1} \tag{4.7-1}$$

where each u_i is an element of GF(q), i.e., an s-bit block. An RS code has an associated generator polynomial

$$\begin{aligned} g(X) &= (X - \alpha)(X - \alpha^2) \cdots (X - \alpha^{n-k}) \\ &= g_0 + g_1 X + g_2 X^2 + \cdots + g_{n-k-1} X^{n-k-1} + X^{n-k} \end{aligned}$$

[2] However, an equivalent block-length code could have a higher minimum distance. For example, RS(255,223) has $d_{min} = 33$, the largest possible for this length code over GF(256). The same RS code could be viewed as a $(255 \times 8, 223 \times 8) = (2040, 1784)$ binary code, which is not necessarily a maximum distance separable code over GF(2), i.e., there could be other (2040,1784) binary codes with larger minimum distances.

where α is a primitive element of GF(q), and for each i, g_i is an element of GF(q). The parity symbols are the coefficients of the remainder

$$p(X) = p_0 + p_1 X + p_2 X^2 + \cdots + p_{n-k-1} X^{n-k-1} \tag{4.7-2}$$

that results from dividing $u(X)X^{n-k}$ by $g(X)$. There are efficient hardware implementations of this operation using shift registers [79]. To decode RS codes, $n-k$ syndromes are computed by dividing the received polynomial by $X+\alpha^i$, $i \in \{1, \cdots, n-k\}$. Then Berlekamp's algorithm is used to find the error-location polynomial, and the corrected values in the error-locations can be computed [79].

RS codes naturally fit the nonbinary nature of PPM signaling [21,30,31]. One can use RS(n,k) with M-PPM, $M = n+1$, by assigning each RS code symbol to one PPM symbol. When system constraints push one towards small M, this leads to small block length codes, which have limited coding gain, but this problem can be overcome by using a longer RS code and splitting RS code symbols across multiple PPM symbols.

If each of the n PPM symbols contains a received laser pulse in one of M slots, of average photon-energy K_s photons, the information rate for RS codewords is given by $\rho = ks/nK_s = r(s/K_s)$ bits/photon, where $r = k/n$ is the code rate.

4.7.2 Turbo and Turbo-Like Codes for Optical Modulations

For a fixed rate k/n, the performance of a code (e.g., the achievable bit error rate as a function of signal power), roughly speaking, increases with n, the block length of the code. However, the complexity of ML decoding of the code increases with the block length as well. For example, complexity increases quadratically in the block length with RS codes, making their implementation prohibitively complex for very large block lengths.

Turbo-like serial and parallel concatenated codes, decoded iteratively, achieve a large effective block length while providing low complexity near ML decoding. In this section, we discuss recent studies applying turbo and turbo-like codes to the deep-space, or free space, optical channel.

4.7.2.1 Parallel Concatenated (Turbo) Codes. A turbo code, or parallel concatenated convolutional code, preceding a PPM modulator has been proposed in [33,49]. The turbo code is a binary code, and its outputs are gathered and mapped to PPM symbols using the method described in Section 4.3.2. The turbo code can be viewed as an outer code C_o and the modulation as an inner code C_i.

To decode such a code effectively, the decoder needs more than the hard M-PPM symbol decisions that are used for RS-coded PPM. A turbo decoder takes from the receiver the set of M soft statistics corresponding to the slots, from which it can compute the likelihood of any symbol, and ultimately, the likelihood of each bit used to form the M-ary symbol. This is used to initialize the binary symbol likelihoods that a turbo decoding algorithm would use for conventional BPSK, as described in [80–82].

Turbo-coded PPM has been shown to offer improvements of 0.5 to 1 dB over RS-coded PPM, when the PPM size is 256 and WMC-plus-Gaussian statistics are assumed. Some additional improvement could be expected by updating the binary symbol likelihoods that are input to the turbo decoder with each new iteration. That is the approach taken in the next section, which also uses a single convolutional code instead of a parallel concatenation of convolutional codes.

4.7.2.2 Serially Concatenated Codes with Iterative Decoding. Modulation is a mapping of bits to symbols transmitted on the channel. This mapping may be considered a code and demodulation a decoding of the code. Conventionally, the modulation and ECC are decoded independently, with the demodulator sending its results to the ECC decoder. However, we may consider the combination of the modulation and the ECC as a single large code that maps user information bits directly to the symbols transmitted on the channel. We could gain several dB in performance by decoding the ECC and modulation jointly as a single code relative to decoding them independently.

An exact ML decoding of the joint modulation–ECC code would, in most cases of practical interest, be prohibitively complex. However, we may approximate true ML decoding while limiting the decoder complexity by iteratively decoding the modulation and the ECC. Iterative decoding is described in more detail in [80]. Applications of iterative decoding to the deep-space optical channel may be found in [84].

The PPM mapping may be preceded by a binary accumulator, making the inner code recursive. We refer to the inner code that is formed by the concatenation of a binary accumulator and PPM mapping as accumulate-PPM (APPM), illustrated in Fig. 4-30. The accumulator performs an exclusive-OR operation with the input bit and the previous output of the accumulator. This introduces memory in the modulator, which can be helpful in conjunction with an outer code.

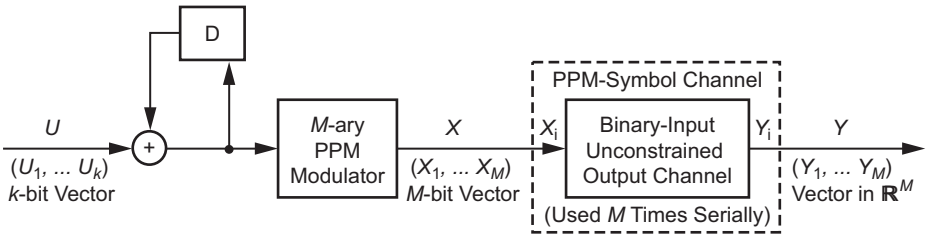


Fig. 4-30. APPM signaling.

4.8 Performance of Coded Optical Modulations

The prior sections have provided us with statistical models of the channel (Section 4.2), practical modulation formats for these channels that are physically realizable (Sections 4.3 and 4.4), the performance of the uncoded modulation (Section 4.5), the capacity of the channel when using this modulation (Section 4.6), and error correction codes suitable for concatenation with the modulation (Section 4.7). We now have all the pieces necessary to design a coded modulation for a particular channel and to measure its performance. In this section we provide general guidelines for this choice and illustrate performance measurements. Our running example will be the Poisson channel. The analysis would carry through in an analogous manner for other channel models.

4.8.1 Parameter Selection

In the absence of other constraints, we first choose to use the PPM order that maximizes the capacity for the available signal power. Other constraints are considered elsewhere [35–39,44–46]. For example, suppose our channel is Poisson with $K_b = 1$ and we have an average signal power of $K_s/M = 0.0541$ signal photons per slot. From Fig. 4-24, we find the the maximum capacity is 56 megabits (Mb) per slot, achieved by taking $M = 64$. Let M^* be the optimum choice of M. Figure 4-31 illustrates M^* constrained to be a power of two as a function of average power for the Poisson channel with $K_b \in \{0, 0.01, 0.1, 1.0, 10\}$. Discontinuities correspond to switching the order and are shown as vertical lines. The optimal order is increasing in K_b and decreasing in P_{av}, showing that as channel conditions become worse, the channel is used efficiently by moving to higher orders, increasing the peak power for a given average power. As channel conditions improve, we may increase the throughput by lowering the peak power and transmitting pulses more frequently (for the same average power).

Given the PPM order M^*, we choose the ECC code rate to satisfy

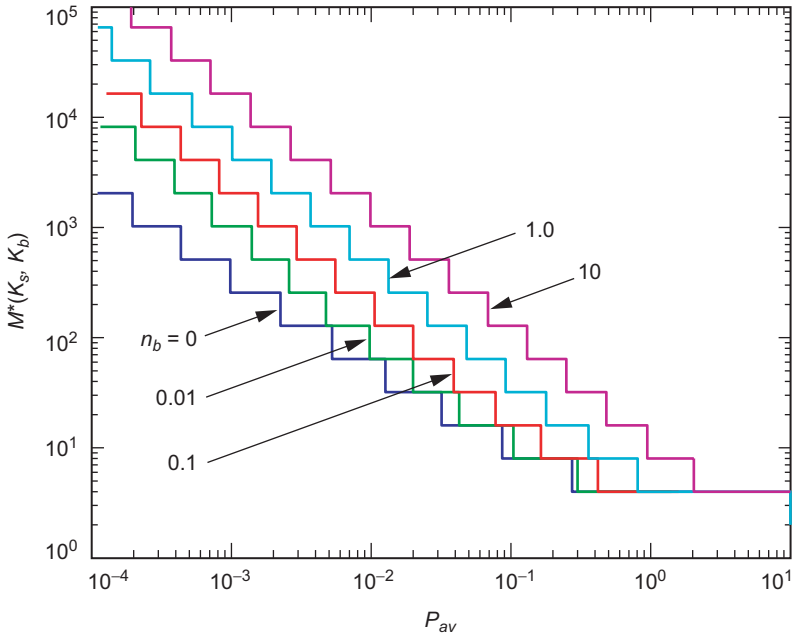


Fig. 4-31. Optimal PPM order, M^*, $K_b \in \{0, 0.01, 0.1, 1, 10\}$.

$$R \approx \frac{CM^*}{\log_2 M^*}$$

Figure 4-32 illustrates the ECC rate as a function of average power for several background levels on the Poisson channel. The discontinuities, corresponding to switching the order, obscure the general behavior. The behavior of the rate can be seen more clearly by allowing M^* to be real-valued. Let M^*_{OOK} specify the duty cycle of the modulation that maximizes Eq. (4.6-5). The upper bound on the rate of a modulation with duty cycle $1/M^*_{\mathrm{OOK}}$ is given by

$$h\left(M^*_{\mathrm{OOK}}\right) = \frac{1}{M^*_{\mathrm{OOK}}} \log_2 M^*_{\mathrm{OOK}} + \frac{M^*_{\mathrm{OOK}} - 1}{M^*_{\mathrm{OOK}}} \log_2 \left(\frac{M^*_{\mathrm{OOK}}}{M^*_{\mathrm{OOK}} - 1} \right)$$

bits per slot. The capacity and duty cycle may be used to specify an ECC data rate as

$$R_{\mathrm{OOK}} = \frac{C^*_{\mathrm{OOK}}}{h\left(M^*_{\mathrm{OOK}}\right)}$$

which is shown in Fig. 4-33 for the Poisson channel with $K_b \in \{0, 0.01, 0.1, 1, 10\}$.

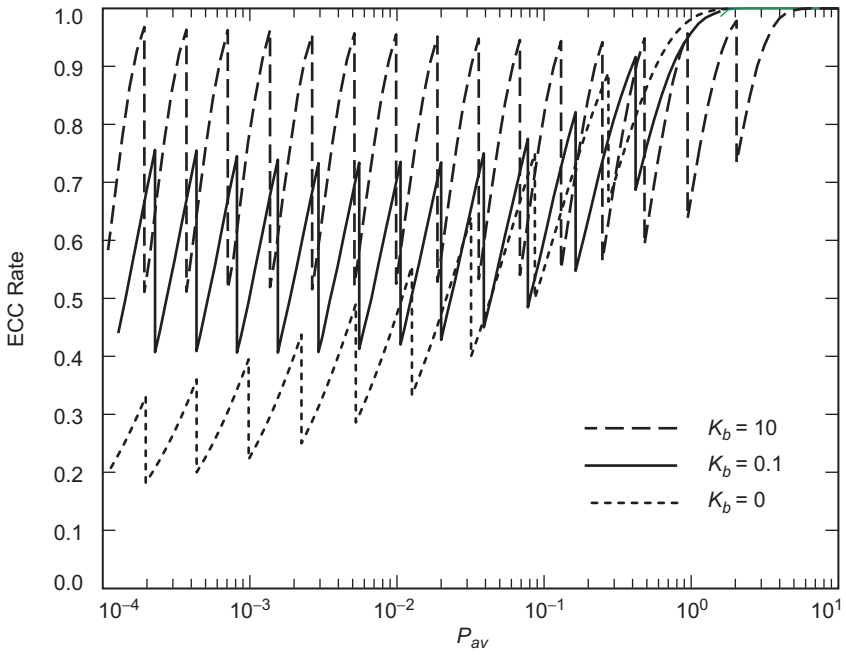


Fig. 4-32. PPM ECC rates, Poisson channel, $K_b \in \{0, 0.1, 10\}$.

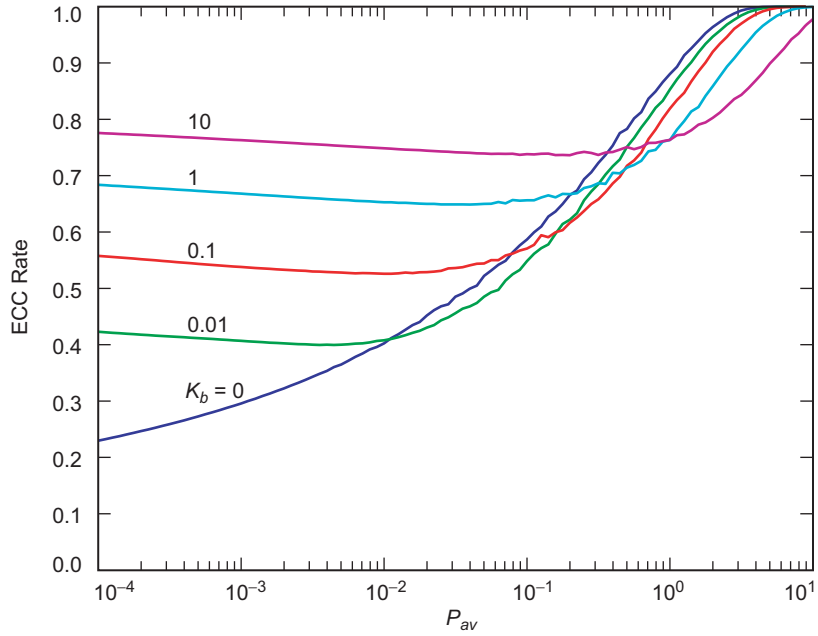


Fig. 4-33. Duty-cycle-constrained ECC rates, Poisson channel, $K_b \in \{0, 0.01, 0.1, 1, 10\}$.

4.8.2 Estimating Performance

We can save simulation time and gain insight by analytically determining the performance of coded modulation schemes. Such analysis is well known for RS codes and has been recently developed for iterative decoding schemes. An extensive discussion is beyond the scope of this chapter. We give a brief overview of the analysis in the following.

4.8.2.1 Reed–Solomon Codes. With hard-decision decoding, symbol decisions are made and sent to the RS decoder, which corrects all patterns of t or fewer errors. With p denoting the input channel symbol error probability, the probability of symbol error is approximated by the following expression:

$$P_s \approx (2^m - 1)^{-1} \sum_{k=t+1}^{N} k \binom{N}{k} p^k (1-p)^{N-k} \tag{4.8-1}$$

This expression holds for channel conditions dominated by symbol errors, rather than erasures. However, when channel conditions are dominated by symbol erasures, rather than symbol errors, the performance of the RS decoder improves.

An erasure RS decoder does not make undetected errors. That is, either the number of erasures in a received word is d_{min} or less, in which case the codeword is decoded correctly, or there are more than d_{min} erasures, in which case the decoder announces that the number of erasures is too large to decode properly.

The codeword error rate of an (n, k) RS code used with M-PPM on an erasure channel is (see, e.g., [21,79])

$$P_w = \sum_{j=d_{min}}^{n} \binom{n}{j} P_s^j (1 - P_s)^{n-j} \tag{4.8-2}$$

where $d_{min} = n - k + 1$ is the minimum distance of the code, and P_s is the symbol erasure rate as discussed in Section 4.5.2.

When a decoder error occurs, with high probability a minimum distance codeword is chosen. Since code symbols are equally likely to be in error with an RS code, approximately $(n - k + 1)/n$ information symbols are decoded incorrectly. On average, half the bits that map to a symbol will be in error, hence the bit error rate may be approximated by [20]

$$P_b \approx \frac{1}{2}\frac{n-k+1}{n}P_w \tag{4.8-3}$$

Performance results for RS-encoded PPM symbols on the erasure channel have been computed in [4] and reproduced in Fig. 4-34 for the case $M = 16, 32, 64$, and 128, with $k = M/2, N = M - 1$, and $d_{\min} = (M/2) - 1$.

Conventionally, RS decoders have operated on hard decisions from the receiver due to the complexity of computing maximum-likelihood estimates from soft decisions. However, recent results have demonstrated efficient decoding of RS codes using soft-decision inputs, improving their performance.

With soft-decision decoding, the ML decoder selects the RS codeword with the highest metric. For the case of Poisson detection, the optimum metric consists of the sum of slot counts corresponding to each codeword. For example, if a particular codeword consisted of a laser pulse in the first slot of every PPM symbol, then the metric corresponding to that codeword would be obtained by adding up the detected counts from the first slot of every symbol. Although exact error probabilities are difficult to compute, useful bounds on the probability of codeword error have been derived [83]. For the case of soft-decision decoding, an upper bound on the probability of codeword error (PWE) can be found by a direct application of the union bound. Since the premise of the union

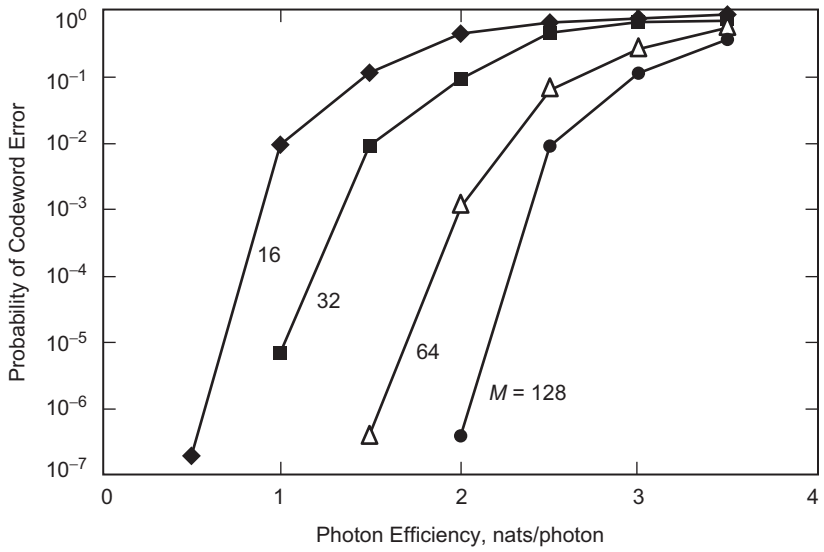


Fig. 4-34. RS decoding: codeword error probabilities for the erasure channel, M = 16, 32, 64, 128.

bound is that the exact error probability never exceeds the sum of pairwise error probabilities, we can write

$$P_w \leq \sum_{k=d_{\min}}^{n} L(k,n) P_k \tag{4.8-4}$$

where P_k is the probability of committing an error when attempting to decide between two codewords a distance k apart, and $L(k,n)$ is the codeword enumerator function that specifies the number of codewords that are a distance k apart, with a code of length n. For RS codes the enumerator function is of the form

$$L(k,n) = \sum_{j=n-k}^{n-d_{\min}} (-1)^{j+k-n} \binom{j}{N-k} \binom{n}{j} \left(M^{n-d_{\min}+1-j} - 1\right) \tag{4.8-5}$$

where the index k takes on the values $k = d_{min}, d_{min}+1, \cdots n$. Plots of P_b as a function of n, for various average signal counts, and with $K_b = 1$ photon/slot, have been computed in [83].

4.8.2.2 Iterative Codes. The bit-error-rate versus average-power curve of a typical iterative code may be roughly divided into three regions (moving from left to right, or from low average power to high average power): a flat high error rate region, a "waterfall" region, and an "error floor" region. This behavior arises since typical iterative codes have a small minimum distance, hence an error floor, but low multiplicity, hence the waterfall region when (approximately) ML decoded. Recent results have illustrated methods to accurately predict the location of the waterfall and error floor.

The waterfall region may be predicted by SNR input–output diagrams, e.g., [85], or extrinsic information transfer (EXIT) curve charts, e.g., [86]. The error floor may be predicted from the first few terms of the weight enumerator polynomial and is ultimately dominated by the minimum distance and multiplicity of the minimum-weight codewords, e.g., [87].

4.8.3 Achievable Data Rates Versus Average Signal Power

In this section we present a sample code and modulation design, leading to a family of trade-offs of achievable data rate versus average signal power.

Suppose we have a system with a slot width of 1 ns and background noise $K_b = 1.0$. We would like to find the power required to achieve 56 megabits

per second (Mbps) and choose appropriate coding and modulation. From Fig. 4-24, we find the optimum PPM order to achieve this data rate is $M = 64$, and the minimum required $K_s/M = 0.0541$. To achieve 56 Mbps, we choose a rate $R = 0.6 \approx 0.056/(\log_2(M)/M)$ ECC and concatenate it with 64-PPM. The performance of two candidate ECCs for this operating point, a serially concatenated convolutionally coded PPM, with outer code rate 3/5, and a 16,410-bit interleaver [SCPPM$(3/5, 2, 64, 16410)$] and Reed–Solomon coded PPM with $n = 4095$, $k = 2457$, and $M = 64$ [RSPPM$(4095, 2457, 64)$], are illustrated in Fig. 4-35.

Performance is compared with the lower bound on the bit error rate for a code with rate R, arrived at by application of the converse to the coding theorem [66, Theorem 7.2.2]:

$$P_b \geq \mathcal{H}^{-1}\left(1 - \frac{C}{R}\right) \tag{4.8-6}$$

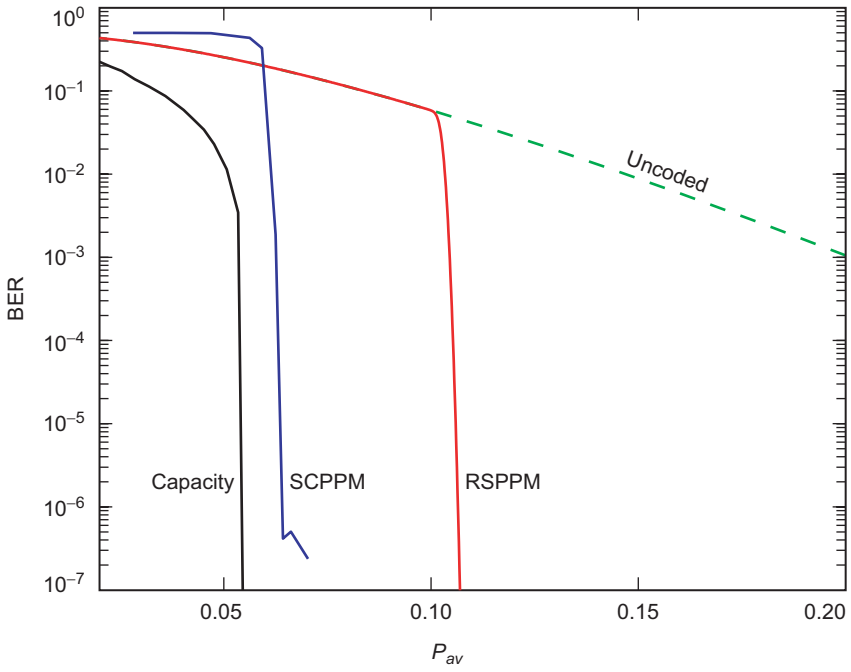


Fig. 4-35. Performance of SCPPM (3/5, 2, 64, 16410), RS (4095, 2457, 64), K_b = 1.0.

where $\mathcal{H}^{-1}$ is the inverse of the entropy function and C is the channel capacity. Unless otherwise noted, comparisons are made for bit error rates of 10^{-6}.

Figure 4-35 illustrates the bound in Eq. (4.8-6) for $C = 5.6$ Mbps as well as uncoded $M = 64$ performance, which, since it carries no coding redundancy, yields 9.4 Mbps. The SCPPM code operates 0.5 dB from capacity, the RS code operates 2.5–3 dB from capacity, and uncoded PPM operates 7.2 dB from capacity (at 5.6 Mbps). An appropriate comparison for uncoded 64-PPM is with capacity for 9.4 Mbps, from which uncoded performance is 4.7 dB. (It would be more efficient to achieve 9.4 Mbps with a rate 3/5 code mapped to 32-PPM).

These comparisons may be extended over a range of desired rates. Figure 4-36 illustrates achievable rates for $K_b = 1$ populated by points corresponding to the class of SCPPM codes, the class of RSPPM codes, and the uncoded PPM channel. The coded and uncoded channels are evaluated at a finite number of rates, which we connect in a line for illustration—this is justifiable by allowing time sharing. RSPPM points use the convention $n = M - 1$, although there is some degradation in using this convention for small M. For example, with $M = 64$, we illustrate a family of points where the RS code sym bol length is allowed to span multiple PPM symbols (as is done in Fig. 4-35).

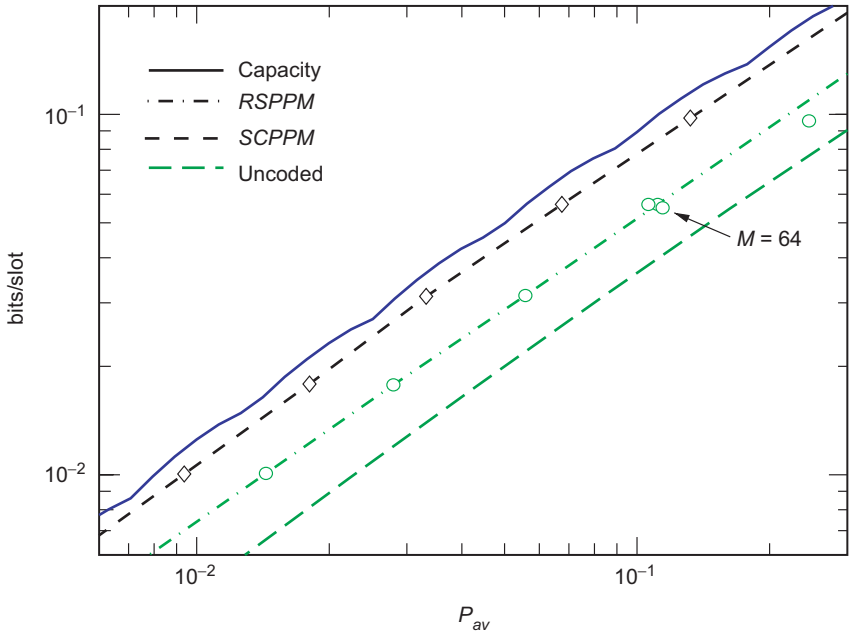


Fig. 4-36. Sample operating points, $K_b = 1$.

Points correspond to the average power at which the BER is 10^{-6}. We exclude iterative codes that exhibit error floors at BERs greater than 10^{-6}.

The class of SCPPM codes lies approximately 0.5 dB from capacity, while the class of RSPPM codes lies approximately 2.75 dB from capacity, and uncoded performance is 4.7 dB from capacity. These gaps will vary with K_b, but they provide a good approximation over a range of expected background noise levels.

References

[1] A. Schawlow and C. Townes, "Infrared and Optical Masers," *Phys. Rev.*, vol. 112, pp. 1940–1949, December 1958.

[2] T. H. Maiman, "Stimulated Optical Radiation in Ruby," *Nature*, vol. 187, p. 493, 1960.

[3] C. W. Helstrom, *Quantum Detection and Estimation Theory, Mathematics in Science and Engineering*, vol. 123, New York: Academic Press, 1976.

[4] R. J. McEliece and L. R. Welch, "Coding for Optical Channels with Photon-Counting," *The Deep Space Network Progress Report 42-52, May and June 1979*, Jet Propulsion Laboratory, Pasadena, California, pp. 61–66, August 15, 1979. http://ipnpr.jpl.nasa.gov/progress_report/

[5] C. W. Helstrom, J. W. S. Liu, and J. P. Gordon, "Quantum Mechanical Communications Theory," *Proceedings of the IEEE*, vol. 58, no. 10, pp. 1578–1598, 1970.

[6] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948.

[7] R. Gagliardi and S. Karp, *Optical Communications*, New York: J. Wiley and Sons, 1976.

[8] R. J. Glauber, "Coherent and Incoherent States of the Radiation Field," *The Physical Review*, vol. 131, pp. 2766–2788, September 1973.

[9] S. Barnett and P. Radmore, *Methods in Theoretical Quantum Optics*, New York: Oxford University Press, Inc., 1997.

[10] S. A. Butman, J. Katz, and J. R. Lesh, "Bandwidth Limitations on Noiseless Optical Channel Capacity," *IEEE Transactions on Communications*, vol. COM-30, pp. 1262–1264, May 1982.

[11] C.-C. Chen, "Figure of Merit for Direct-Detection Optical Channels," *The Telecommunications and Data Acquisition Progress Report 42-109, January–March 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 136–151, May 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[12] M. H. A. Davis, "Capacity and Cutoff Rate for Poisson-Type Channels," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 710–715, November 1980.

[13] C. N. Georghiades, "Modulation and Coding for Throughput-Efficient Optical Systems," *IEEE Transactions on Information Theory*, vol. 40, pp. 1313–1326, September 1994.

[14] J. Hamkins, "The Capacity of Avalanche Photodiode-Detected Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–19, August 15, 1999.
http://ipnpr.jpl.nasa.gov/progress_report/

[15] C. W. Helstrom, "Comments on the Capacity of the Photon Counting Channel," *IEEE Transactions on Information Theory*, vol. IT-28, p. 556, May 1982.

[16] V. K. Jain, G. Steudal, and C. Rapp, "Channel Capacity for Optical Space Communication Systems," *Journal of Optical Communications*, vol. 18, no. 2, pp. 57–63, 1997.

[17] Y. M. Kabanov, "The Capacity of a Channel of the Poisson Type," *Theory Probab. Appl.*, vol. 23, pp. 143–147, 1978.

[18] J. R. Lesh, "Capacity Limit of the Noiseless, Energy-Efficient Optical PPM Channel," *IEEE Transactions on Communications*, vol. COM-31, pp. 546–548, April 1983.

[19] R. G. Lipes, "Pulse-Position-Modulation Coding as Near-Optimum Utilization of Photon Counting Channel with Bandwidth and Power Constraints," *The Deep Space Network Progress Report 42-56, January and February 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 108–113, April 15, 1980.
http://ipnpr.jpl.nasa.gov/progress_report/

[20] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *IEEE Transactions on Communications*, vol. COM-29, pp. 1615–1621, November 1981.

[21] R. J. McEliece, "Practical Codes for Photon Communication," *IEEE Transactions on Information Theory*, vol. IT-27, pp. 393–398, July 1981.

[22] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The Capacity of the Photon Counting Channel," *IEEE Transactions on Information Theory*, vol. IT-27, pp. 61–77, January 1981.

[23] D. L. Snyder and I. B. Rhoades, "Some Implications of the Cutoff-Rate Criterion for Coded Direct-Detection Optical Communication Systems," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 327–338, May 1980

[24] H. H. Tan, "Capacity of a Multimode Direct Detection Optical Communication Channel," *The Telecommunications and Data Acquisition Progress Report 42-63, March and April 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 51–70, June 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[25] M. Takahsi, H. Yashima, I. Sasase, and S. Mori, "Capacity and Effects of Reed–Solomon Codes on Multi-Pulse PPM in Optical Communications," *International Conference on Communication*, vol. 345.4.1, pp. 1663–1667, April 1990.

[26] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part I," *IEEE Transactions on Information Theory*, vol. 34, pp. 1449–1461, November 1988.

[27] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part II," *IEEE Transactions on Information Theory*, vol. 34, pp. 1462–1471, November 1988.

[28] D. Zwillinger, "Differential PPM has a Higher Throughput than PPM for the Band-Limited and Average-Power-Limited Optical Channel," *IEEE Transactions on Information Theory*, vol. 34, pp. 1269–1273, September 1988.

[29] A. R. Calderbank and C. N. Georghiades, "Synchronizable Codes for the Optical OPPM Channel," *IEEE Transactions on Information Theory*, vol. 40, pp. 1097–1107, July 1994.

[30] D. Divsalar, R. M. Gagliardi, and J. H. Yuen, "PPM Performance for Reed–Solomon Decoding over an Optical–RF Relay Link," *IEEE Transactions on Communications*, vol. COM-32, pp. 302–305, March 1984.

[31] E. Forestieri, R. Gangopadhyay, and G. Prati, "Performance of Convolutional Codes in a Direct-Detection Optical PPM Channel," *IEEE Transactions on Communications*, vol. 37, no. 12, pp. 1303–1317, 1989.

[32] R. M. Gagliardi and S. Karp, *Optical Communication*, New York: John Wiley and Sons, Inc., 1995.

[33] J. Hamkins, "Performance of Binary Turbo-Coded 256-ary Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, AprilJune 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–15, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[34] M. Jeganathan and S. Mecherle, *A Technical Manual for FOCAS 2.0—Free-Space Optical Communications Analysis Software*, May 1998.

[35] C.-H. Lai and K. Kiasaleh, "Modified Viterbi Decoders for Joint Data Detection and Timing Recovery of Convolutionally Encoded PPM and OPPM Optical Signals," *IEEE Transactions on Communications*, vol. 45, pp. 90–94, January 1997.

[36] C.-H. Lai and K. Kiasaleh, "A New Performance Upper Bound for Convolutionally Encoded Direct-Detection Optical OPPM Communications Systems," *Proceedings of the International Conference on Communications (ICC'95)*, Seattle, Washington, pp. 1302–1306, June 1995.

[37] V. Vilnrotter, *Optical Receivers Using Rough Reflectors*, JPL Publication 25, Jet Propulsion Laboratory, Pasadena, California, May 15, 1985.

[38] V. A. Vilnrotter, "Spatial Acquisition of Optical Sources in the Presence of Intense Interference," *The Telecommunications and Data Acquisition Progress Report 42-58, May and June 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 91–96, August 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[39] Z. Xiaomai, X. Zhaofei, and H. Tiexin, "On Evaluating the Performance of Error-Correcting Codes in Lasercom PPM Systems with Direct Detection," *IEEE TENCON*, Beijing, China, pp. 489–491, 1993.

[40] R. J. McIntyre, "The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory," *IEEE Transactions on Electronic Devices*, vol. ED-19, no. 6, pp. 703–713, June 1972.

[41] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," *RCA Review*, vol. 35, pp. 234–278, June 1974.

[42] F. M. Davidson and X. Sun, "Gaussian Approximation Versus Nearly Exact Performance Analysis of Optical Communication Systems with PPM Signaling and APD Receivers," *IEEE Transactions on Communications*, vol. 36, pp. 1185–1192, November 1988.

[43] S. J. MacMullan and O. M. Collins, "The Capacity of Orthogonal and Bi-Orthogonal Codes on the Gaussian Channel," *IEEE Transactions on Information Theory*, vol. 44, pp. 1217–1232, May 1998.

[44] R. A. Cryan, A. J. Phillips, and J. M. Senior, "Optically Preamplified n-ary PPM Systems," *SPIE*, vol. 2024, pp. 72–80, 1993.

[45] T. T. Ha, G. E. Keiser, and R. L. Borchardt, "Analysis of Direct Detection Lightwave Systems with Optical Amplifiers," *Proceedings of MILCOM (Military Communications Conference)*, pp. 14.4-1–14.4-5, 1994.

[46] M. A. Herro and L. Hu, "A New Look at Coding for APD-Based Direct-Detection Optical Channels," *IEEE Transactions on Information Theory*, vol. 34, pp. 858–866, July 1988.

[47] L. W. Hughes, "A Simple Upper Bound on the Error Probability for Orthogonal Signals in White Noise," *IEEE Transactions on Communications*, vol. 40, p. 670, April 1992.

[48] G. S. Mecherle, *Maximized Data Rate Capability for Optical Communication Using Semiconductor Devices with Pulse Position Modulation*, Ph.D. Thesis, University of Southern California, Los Angeles, California, May 1986.

[49] J. Hamkins and M. Srinivasan, "Turbo Codes for APD-Detected PPM," *Proceedings of the Thirty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, pp. 29–38, September 1998.

[50] M. Srinivasan and V. Vilnrotter, "Performance of the Optimum Receiver for Pulse-Position Modulation Signals with Avalanche Photodiode Statistics," *The Telecommunications and Mission Operations Progress Report 42-133, January–March 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–10, May 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[51] S. Dolinar, D. Divsalar, J. Hamkins, and F. Pollara, "Capacity of Pulse-Position Modulation (PPM) on Gaussian and Webb Channels," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–31, August 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[52] M. Srinivasan and V. Vilnrotter, "Symbol-Error Probabilities for Pulse-Position Modulation Signaling with an Avalanche Photodiode Receiver and Gaussian Thermal Noise," *The Telecommunications and Mission Operations Progress Report 42-134, April–June 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, August 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[53] G. G. Ortiz, J. V. Sandusky, and A. Biswas, "Design of an Opto-Electronic Receiver for Deep-Space Optical Communications," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, August 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[54] M. Srinivasan, J. Hamkins, B. Madden-Woods, A. Biswas, and J. Beebe, "Laboratory Characterization of Silicon Avalanche Photodiodes (APDs) for Pulse-Position Modulation (PPM) Detection," *The InterPlanetary Network Progress Report 42-146, April–June 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, August 15, 2001. http://ipnpr.jpl.nasa.gov/progress_report/

[55] G. M. Lee and G. W. Schroeder, "Optical PPM with Multiple Positions per Pulsewidth," *IEEE Transactions on Communications*, vol. COM-25, pp. 360–364, March 1977.

[56] R. A. Horn and C. R. Johnson, *Matrix Analysis*, New Jersey: Cambridge University Press, 1988.

[57] C. E. Shannon, *The Mathematical Theory of Communication*, Urbana, Illinois: University of Illinois Press, 1963.

[58] K. A. S. Immink, *Codes for Mass Data Storage Systems*, The Netherlands: Shannon Foundation Publishers, 1999.

[59] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for Digital Recorders," *IEEE Transactions on Information Theory*, vol. 44, pp. 2260–2299, October 1998.

[60] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained Systems and Coding for Recording Channels," *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, eds., Chapter 20, Elsevier Science, 1998.

[61] P. A. Franaszek, "Sequence-State Coding for Digital Transmission," *The Bell System Technical Journal*, vol. 47, pp. 143–157, 1968.

[62] H. V. Poor, *An Introduction to Signal Detection and Estimation*, New York,: Springer-Verlag, 1988.

[63] V. Vilnrotter and E. Rodemich, "A Generalization of the Near-Optimum Binary Coherent State Receiver Concept," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 446–450, March 1984.

[64] R. M. Gagliardi, V. A. Vilnrotter, and S. J. Dolinar, *Optical Deep Space Communication via Relay Satellite*, JPL Publication 81-40, Jet Propulsion Laboratory, Pasadena, California, August 15, 1981.

[65] V. Vilnrotter, M. Simon, and M. Srinivasan, *Maximum Likelihood Detection of PPM Signals Governed by an Arbitrary Point Process Plus Additive Gaussian Noise*, JPL Publication 98-7, Jet Propulsion Laboratory, Pasadena, California, April 1998.

[66] R. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[67] J. Hamkins, "Accurate Computation of the Performance of M-ary Orthogonal Signaling on a Discrete Memoryless Channel," *IEEE Transactions on Communications*, 2003 (to appear).

[68] V. A. Vilnrotter, "An M-ary Coherent Optical Receiver for the Free-Space Channel," *The Telecommunications and Data Acquisition Progress Report 42-66, September and October 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 60–66, December 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[69] R. S. Kennedy, "A Near-Optimum Receiver for the Binary Coherent State Quantum Channel," *M.I.T. Research Laboratory of Electronics Quarterly Progress Report*, vol. 108, pp. 219–225, January 1973.

[70] S. Dolinar, "An Optimum Receiver for the Binary Coherent State Quantum Channel," *M.I.T. Research Laboratory of Electronics Quarterly Progress Report*, vol. 111, pp. 115–120, October 1973.

[71] S. J. Dolinar, Jr. "A Near-Optimum Receiver Structure for the Detection of M-ary Optical PPM Signals," *The Telecommunications and Data Acquisition Progress Report 42-72, October–December 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 30–42, February 15, 1983.
http://ipnpr.jpl.nasa.gov/progress_report/

[72] V. A. Vilnrotter, "A Binary Coherent Optical Receiver for the Free-Space Channel," *The Telecommunications and Data Acquisition Progress Report 42-61, November and December 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 27–38, February 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[73] V. Vilnrotter and C.-W. Lau, "Quantum Detection in the Presence of Noise," *Proceedings of SPIE, Free-Space Laser Communications Technologies XV*, vol. 4975, San Jose, California, January 2003.

[74] V. A. Vilnrotter and C.-W. Lau, "Quantum Detection of Binary and Ternary Signals in the Presence of Thermal Noise Fields," *The Interplanetary Network Progress Report 42-152, October–December 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, February 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[75] V. Vilnrotter and C.-W. Lau, "Quantum Detection Theory for the Free-Space Channel," *The InterPlanetary Network Progress Report 42-146, April–June 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–34, August 15, 2001.

[76] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[77] B. Moision and J. Hamkins, "Deep-Space Optical Communications Downlink Budget: Modulation and Coding," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–28, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[78] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-State Modulation Codes for Data Storage," *IEEE Journal Selected Areas Communications*, vol. 10, pp. 5–37, January 1992.

[79] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, New Jersey: Prentice Hall, 1983.

[80] S. Benedetto, D. Divsalar, G. Montorsi, and F.Pollara, "A Soft-Input Soft-Output Maximum A Posteriori (MAP) Module to Decode Parallel and Serial Concatenated Codes," *The Telecommunications and Data Acquisition Progress Report 42-127, July–September 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, November 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[81] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[82] K. Andrews, V. Stanton, S. Dolinar, V. Chen, J. Berner, and F. Pollara, "Turbo-Decoder Implementation for the Deep Space Network," *The Interplanetary Network Progress Report 42–148, October–December 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, February 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[83] G. Prati and R. Gagliardi, "Block Encoding and Decoding for the Optical PPM Channel," *IEEE Transactions on Information Theory*, vol. IT-28, January 1981.

[84] B. Moision and J. Hamkins, "Low Complexity Serially Concatenated Codes for the Deep-Space Optical Channel," International Symposium on Information Theory, Yokohama, Japan, June 2003.

[85] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative Turbo Decoder Analysis Based on Density Evolution," *IEEE Journal Selected Areas Communications*, vol. 19, pp. 891–907, May 2001.

[86] A. Ashikhmin, G. Kramer, and S. ten Brink, "Code Rate and the Area Under the Extrinsic Information Transfer Curves," *IEEE International Symposium on Information Theory*, p. 115, 2002.

[87] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding," *The Telecommunications and Data Acquisition Progress Report 42-126, April–June 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–26, August 15, 1996.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-126/126D.pdf

Notation

Systems Parameters

$R =$ Data rate, bits/second

$\rho =$ Photon efficiency, bits/photon

$n_s =$ Rate of photons incident on detector, photons/second)

$P_b =$ Probability of bit error (BER)

$P_s =$ Probability of symbol error (SER)

$P_c =$ Probability of correct detection

$X =$ Value (typically binary) transmitted during a time slot

$Y =$ Real value at output of detector

$m, \sigma^2 =$ Mean, variance of Y

$f_{Y|X}(y|0) =$ Conditional probability density (or mass) function of Y given $X = 0$

$f_{Y|X}(y|1) =$ Conditional probability density (or mass) function of Y given $X = 1$

$m_0, m_1 =$ Conditional mean of Y given $X = 0$ or 1

$\sigma_0^2, \sigma_1^2 =$ Conditional variance of Y given $X = 0$ or 1

$K_s =$ Average number of absorbed signal photons per pulse

$K_b =$ Average number of absorbed background photons per slot

$\beta =$ Slot SNR, $(m_1 - m_0)^2/\sigma^2$

$\gamma =$ Excess slot SNR, $(m_1 - m_0)^2/\sigma_1^2 - \sigma_0^2$

$\beta_b =$ Bit SNR, $\beta/(2R_c)$

Laser and Modulator Parameters

$\nu =$ Optical frequency, nanometers

$M =$ PPM order

$T_s =$ Width of the signal slot, seconds

Detector Parameters

$\eta =$ Quantum efficiency

$k_{eff} =$ Ionization ratio

$F =$ Excess noise factor, $k_{eff}G + (2 - 1/G)(1 - k_{eff})$

$I_b =$ Bulk leakage current, amperes

$I_s =$ Surface leakage current, amperes

$T =$ Noise temperature, kelvins

$R_L =$ Load resistance (transimpedance model), $5.75 \times 10^{12} \times T_s$ ohms

$B =$ Noise equivalent one-sided bandwidth, hertz

$G =$ Gain

Physical Constants

$\kappa =$ Boltzmann's constant, 1.38×10^{-23} joules/kelvin

$q =$ Electron charge, 1.6×10^{-19} coulombs

$h =$ Planck's constant, 6.624×10^{-34} joules/hertz

Chapter 5
Flight Transceiver

Hamid Hemmati, Gerardo G. Ortiz, William T. Roberts,
Malcolm W. Wright, and Shinhak Lee

5.1 Optomechanical Subsystem

Hamid Hemmati

The flight transceiver's major subsystems include opto-mechanical; laser transmitter; and acquisition, tracking and pointing. Flight qualification is a major aspect of a flight transceiver. Subchapters 5.1 through 5.4 provide a description of each of these areas.

5.1.1 Introduction

The flight transceiver terminal optical train is typically comprised of an afocal fore-optics accommodating transmit, receive, align (calibration), and beacon reference channels for acquisition, tracking and pointing. The optical system assembly typically consists of a front aperture, reflection or refraction type telescope, with or without a solar rejection filter, aft optics (including lenses, mirrors, beam-splitters, and filters), fine-pointing mirror(s), and array detector(s) (quadrant or larger area array). The optical system assembly also includes the mechanical support to provide a rigid, low-thermal-expansion structure for the optical system, baffling to reject stray light, and a thermal control assembly to control the temperature of the laser head module.

Figure 5-1 details the optical approach for the system that the Jet Propulsion Laboratory (JPL) has developed for communications to planets within our Solar System. This system is more complex than communication systems for Earth orbit due to requirements for better isolation of transmit and

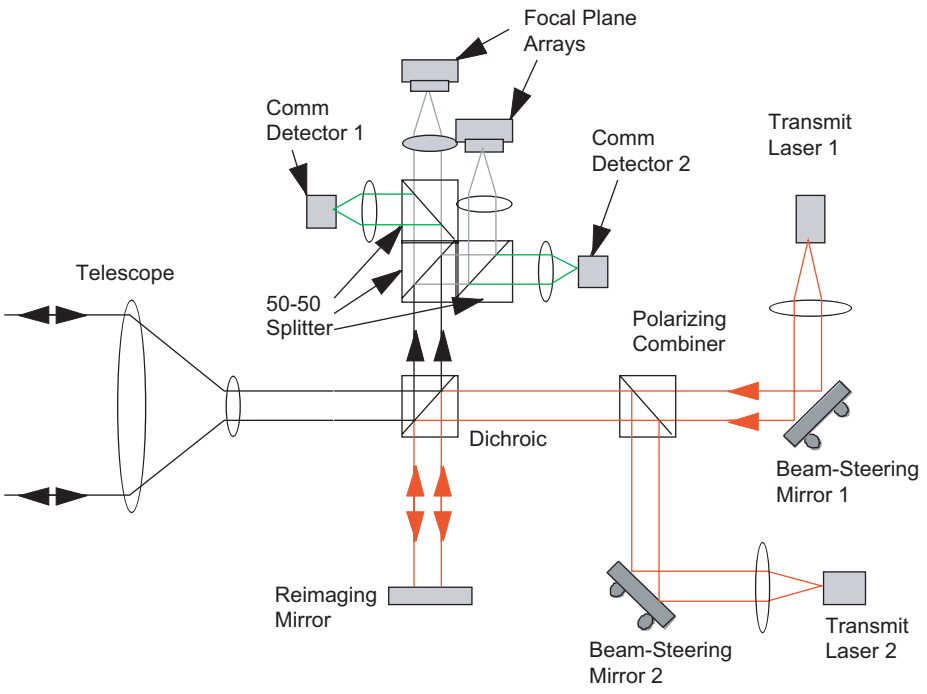


Fig. 5-1. Block diagram of the optical communications transceiver, including redundant channels.

receive channels; stringent demands on acquisition, tracking and pointing (imposed by the greater communication range); and the need to attenuate a larger amount of out-of-field radiation [1]. The out-of-field radiation is due primarily to operating conditions that periodically require the Sun to be near the field of view (within as little as 1 deg).

5.1.2 Optical Beam Paths

In addition to transmitting the optical data, the optical system receives the beacon signal, which is used for acquisition and tracking and to uplink command data from Earth or from another spacecraft. The beacon signal has a narrow spectral band, such as a laser signal from a cooperative target. Pointing may also be assisted by offset from wider band references, such as celestial reference signals from stars or from the Sun-illuminated Earth or Moon.

The transmit and receive channels may consist of separate or common apertures. Use of a single aperture, for both transmit and receive, minimizes coordinate transfer errors between the tracking and the transmit channels. It also minimizes size. In this case though, transmit and receive isolation may become challenging.

Fig. 5-2. A picture of JPL's Optical Communication Demonstrator optical assembly.

Figure 5-2 illustrates the Optical Communication Demonstrator (OCD) developed at JPL [2,3]. This optical system includes four channels and utilizes a single aperture for both transmitting and receiving signals. The telescope is a two-mirror system in a Cassegrain configuration. The telescope and the aft optics can be designed to provide a high degree of stray-light rejection. This may be achieved by the incorporation of field stop(s) and a Lyot stop(s). The field stop is located at the focus of the two-mirror telescope and limits out-of-field scattering into the rest of the optical system. The Lyot stop is located at a conjugate pupil plane near the imager lenses at a location behind the insertion point for the transmit laser. The Lyot stop eliminates diffracted spider and aperture radiation at the focal plane from bright out-of-field sources like the Sun. The location of the Lyot stop assures that it will not be backside-illuminated by the transmit laser.

Utilization of separate focal planes for the celestial and the transmit reference channel may be necessary. This is driven by a large difference in signal level between the two channels. Optimization through use of different sensor types for each channel may also be accomplished with this approach [4].

The transmit channel consists of an optical path extending from the output of the laser transmitter to the exit aperture of the optics. A design-driving requirement may be inclusion of a fine-pointing mirror at the rear pupil. The fine-pointing mirror module controls the downlink over the entire system field-of-view while maintaining good beam quality. A high wavefront quality on the order of 0.025 wave root mean square (RMS) or better is desired for this channel. The transmit optics in front of the laser are essentially a laser beam expander that matches the diameter of the laser beam to the rest of the optical train.

The receive channel's function is to accept light emerging from the fore-optics, and direct it to a circular detector. It is a light-collecting channel; therefore, a high wavefront quality is not important for the receive path as long

as there is not excessive spillover beyond the edges of the detector. In principle, the receiver could be either an imager or a pupil relay. Receiver photodetectors having small (<0.5 mm) active diameters pose many challenges for an efficient design that concentrates all received light onto the detector without overfilling it. As will be discussed later in this chapter an effectively larger active diameter may be formed by adding an immersed lens or a compound parabolic concentrator (CPC) to the photodetector. Narrow-band filtering is typically utilized in front of the concentrator to reduce the incident background radiation. This channel is expected to provide the field-of-view (FOV) required to cover the spacecraft deadband motion cycle. Also, it should provide the degree of optical isolation required to sufficiently minimize the effect of signal feedback from the transmit path.

The acquisition and tracking channel images part of the transmit laser signal onto a reference detector array. In the design of the OCD, the image formed by this optical path is purposely blurred while maintaining uniformity across the field. The receive optics may spread the tracking signal over two to three pixels. A good approach for blurring is to control the beam size with optics as opposed to introducing aberrations. However, the image must be void of coma and other non-symmetric aberration patterns. For OCD, a single array detector is used to receive both the beacon as well as the transmit reference signals. The instantaneous position of the downlink signal is measured in this channel. Over the large interplanetary distances expected, and particularly when Earth-image tracking is required, the dynamic range of the tracking signal can vary by 2 to 3 orders of magnitude. Readout of the beacon image position on the focal-plane-array detector relative to the reference location provided by the transmit signal and center of the array detector, provides the information for the fine pointing mirror to accurately point the transmit laser at the position of the beacon (e.g., the ground station). Due to the cross velocity between the spacecraft and the ground station, the transmit signal must be pointed ahead of (or behind) the apparent position of the ground station.

The reference (or align) channel is basically a simple lens that forms an image of a portion of the transmit light at the array detector without any high degree of image quality. Again, the image is purposely blurred to facilitate centroiding functions.

5.1.3 Optical Design Requirements, Design Drivers, and Challenges

The optical system typically consists of multiple channels including: (1) transmit channel accommodating a two-axis fine-pointing mirror and possibly a dedicated point-ahead mirror; (2) receive channel; (3) acquisition and tracking channel; (4) tracking reference channel, and (5) calibration channel. Some of the design drivers and design practices include:

Afocal fore-optics: A collimated beam behind the telescope provides design simplifications, for example, when accommodating a fine-pointing mirror within the terminal. Typically, for flight qualification reliability and onboard calibration difficulty reasons, a focusing mechanism for the telescope is not incorporated.

Short focal length primary mirror: The requirement for short telescope length limits the F/number of the telescope primary mirror to about F/1.0. A large secondary magnification is required to provide an adequately long overall focal length for the required FOV and spot size at the acquisition detector. This large magnification results in significant field curvature, which must be corrected by the auxiliary optics.

Field of view: The field-of-view (FOV) requirement of a fraction of a degree is relatively large for two-mirror telescope optical systems. Residual aberrations are curvature of field and off-axis astigmatism, which must be corrected by the auxiliary refractive optics. This makes the auxiliary optical design more complicated. The communication and acquisition FOVs are typically different, for example, 1 mrad (0.06 deg) for communications, and ±5 mrad for acquisition. Coincidence between the transmit laser and beacon beam within the communications FOV must be held to very tight tolerances.

Stability requirements: The optical system must provide, at all times, an adequate level of pointing stability between the transmit and the receive channels. The required stability is typically much less than 0.1 mrad. The distance between the telescope primary and secondary (or tertiary mirror, if applicable) should be held to very tight tolerances (typically on the order of a few microns) over lifetime of the mission.

Well-baffled telescope: Spurious (stray light) signals can cause significant radiometric and thermal problems for the laser communication terminal. Therefore, stray light rays that get past the baffles to anywhere inside the field-stop must be blocked off before they can go through the telescope aft optics. Typically, referenced to the internal telescope pupil, the maximum solar stray-light levels should be held to less than 1 μW/nm-sr.

Field-stop: A field-stop can effectively block the light from bright objects (like the Sun) near the edge of the FOV in the telescope. Due to diffuse scattering from the telescope mirror surfaces, ideally only a maximum of two mirror surfaces are allocated before the field stop. This means that astigmatism and field curvature will not be corrected in the telescope and must instead be corrected by the auxiliary optics.

Lyot stop: The Lyot stop can largely eliminate diffracted energy from bright out-of-field objects. The Lyot stop is designed so that no baffle, spider vane, or optical element edges ahead of the telescope focal plane can be seen at any of the detectors. The Lyot stop is a conjugate near-field point image of the entrance aperture of the optical system. Two conjugate intermediate images are required. One is at the fine-pointing mirror to assure that there will be no beam

walk at the primary mirror when the fine-pointing mirror is moved for fast pointing purposes. The other near field image is selectively masked with a Lyot stop at some point in the optical system so that it blocks out-of-field radiation that is diffracted into the FOV by the edges of the two telescope mirrors, the baffles, the spider vanes, etc. Providing a second near-field point for a Lyot stop will add size and mass due to the need for additional imaging and collimating optical elements.

Polarization state: For direct detection, polarization state of the beam typically is not of concern unless the design requires it. However, for coherent transmission systems, a well-defined polarization state (linear and circular) is a prerequisite.

Spectral band-pass: When a broad spectral band (non-laser) source beacon is used, it is often difficult to accommodate both the beacon and the transmit wavelengths without significant use of reflective optics. All-refractive designs are generally much more difficult to produce while maintaining a small transceiver volume.

Multiple redundant optical channels: Multiple and redundant optical channels may be formed using refractive optics (Fig. 5-1). However, beam-spread away from an intermediate point due to the FOV drives the size of these optical elements up. The larger acquisition FOV, along with the numerous beam-splitters required, makes the auxiliary optical path length long, increasing the overall volume.

Radiation environment: For certain missions, such as Jupiter and its moons, the radiation environments become a design consideration. This is an issue primarily for the refractive optics and some of the dielectric coatings within the system. Currently, only about a dozen radiation-hard optical glasses with suitable characteristics are commercially available.

5.1.4 Optical Design Drivers and Approaches

The optical design is driven by a need to minimize size and mass of the optical system. Reflective type telescopes provide the best performance versus size and mass for spacecraft use when the telescope aperture is greater than about 7-cm in diameter. Catadioptric, obscured reflective, unobscured reflective configurations, and variations of the Cassegrain telescope design, such as Richey-Chretien (RC), offer good matches to optical communication requirements. Off-axis telescopes are free of a secondary mirror obscuration at the cost of higher alignment tolerance. Off-axis telescopes are larger than on-axis designs of the same aperture. For RC-based or other on-axis Cassegrain transceivers, volume can be reduced and baffling of the telescope optics can be simpler than for most off-axis designs. Restriction on the overall size of the transceiver results in requirements for a relatively fast primary mirror for the telescope. Length of the acquisition and tracking receive channel optical system

can be kept short by using a secondary mirror having a relatively high telephoto ratio of about five to one. Use of all-spherical optics results in lower manufacturing cost compared to the hyperbolic primary and secondary mirrors of an RC telescope, but additional optics are needed to correct spherical aberrations.

An afocal characteristic (i.e., formation of a pupil behind the telescope) is desirable when a fine pointing mirror or Lyot stop is used. Fore-optic configurations with an afocal characteristic include: catadioptric, obscured reflective, unobscured reflective, and slightly off-axis reflective. For larger FOVs (on the order of 20 mrad), a two-mirror afocal system lacks the degrees of freedom to produce diffraction-limited wavefront error over the entire field at a convenient exit pupil location. Additional complexity (i.e., either a third mirror or a refractive collimating group) is needed. A Cassegrain-type design produces negative Petzval curvature (inward curving field). A concave mirror behind the Cassegrain focus can collimate the light. Its Petzval curvature is convex toward the collimator so it can match the Cassegrain Petzval. This is the on-axis version of the configuration which, when used in an unobscured manner is called the Three-mirror Anastigmat (TMA). The TMA and the above three-mirror Cassegrain, match most of the requirements set for the fore-optics. A partially (bent) off-axis primary mirror design in a three-mirror telescope may allow use of spherical mirrors. Folding of the optical path behind the telescope poses its own challenges. The requirements are minimal obscuration and the need to get the components out of each other's path. Good solutions that meet the requirements are available [3,4]. The lowest overall obscuration that can be achieved is on the order of 10 to 15 percent of the diameter.

5.1.5 Transmit–Receive–Isolation

Isolation is required to prevent system failure. The transmit powers are typically ten orders of magnitude larger than the receiver sensitivity levels. For a given transceiver that must point near the Sun, as much as 150 dB isolation of the receive channel from the transmit channel may be required. Specular backscatter from flat transmitting surfaces that are normal to the incident beam could cause actual ghost images of the transmit laser that are concentrated over just a few pixels of the detector array. Tilting of the flat transmitting surfaces alleviates the possibility of ghost images falling inside the detector array field-of-view. Backscattered energy from curved surfaces is often much fainter and less defined. Some of most likely sources of unwanted radiation falling on the various focal planes, and methods to mitigate them are as follows:

1) Scattered sunlight from optical surfaces due to surface roughness and contamination. This effect may be minimized by keeping the optical surfaces as clean as possible and by making them very smooth (on the order of 1 nm RMS) and relatively free of flaws. It is reasonable to expect that

the total integrated scatter due to surface roughness and contamination at each surface is 3×10^{-4} or less. This topic is described in more detail in later in this chapter under Stray Light Control (5.1.6).

2) Diffracted and scattered sunlight from mechanical surfaces such as baffles, aperture edges, and spider vanes. Some may include high order diffraction. Proper baffle design along with a field-stop and Lyot-stop will minimize these effects. The Lyot stop and the field-stop together will assure that no radiation from outside the FOV can be imaged or diffracted into the FOV of any detectors.

Different isolation schemes that can be implemented include:

Spatial isolation uses separate transmit and receive apertures to keep the outgoing transmit beam from back scattering off of common optical elements. This technique is only used for short range links with large beam divergences, which can allow for large transmit-to-receive beam misalignments associated with separate transmit and receive optical systems.

Spectral isolation uses separate transmit and receive wavelengths to isolate transmit and receive beams. This method allows common use of optical elements in the system (including the same aperture), and it can achieve isolation greater than 120 dB. Spectral isolation is applicable to long-range links where narrow beams and co-alignment are required.

Temporal isolation in which the receiver is effectively turned off to eliminate the chance of interference. Temporal isolation can only be used in systems where the data rates are favorable for these timing constraints. Temporal isolation is used in long-range links that require narrow beams and co-alignment, but where only a single wavelength is available.

Polarization isolation in which the transmit light and the receive light are linearly polarized, and then circularized (right hand or left hand) in different polarizations for isolation. This is accomplished with polarization-sensitive optical elements, and is good to about 6 to 8 orders of magnitude.

Aperture sharing in which the optical paths share the same aperture, but do not go through the same exact path. This scheme often requires larger diameter optics than what would otherwise be needed.

Coding utilizes codes with extreme depth of interleave.

Combined isolation where greater degree of isolation is achieved by combining any of the above techniques.

It is expected that the above approaches will result in less than 1×10^{-15} of the incident stray radiation on the telescope falling on any single detector pixel. This is an adequate level of isolation, and for a given optical system it needs to be verified by stray light analysis and actual scatter measurements.

Chopping of the beacon laser light used for acquisition and tracking (e.g. on the order of 5 to 10 kHz, followed by lock-in detection) is an effective means of isolating the incoming beacon light.

5.1.6 Stray-Light Control

Any undesired light reaching the focal spot or a focal plane of the transceiver constitutes stray light. Some of the pathways through which spurious photons may arrive at the focal spot include: scattering from mirror imperfections, contaminants and baffles, or diffraction from contaminants or edges of the secondary mirror obstruction and its supporting structure. Scattering from optical surfaces, a major contributor to stray light, has a strong wavelength dependence that varies as $1/\lambda^4$ [5,6,7,8]. When the transceiver communicates with Earth, the Sun is the primary contributor to stray light since it is one billion times brighter than the Sun-illuminated Earth. Therefore, even a miniscule fraction of the sunlight scattered from the front aperture optomechanical elements will amount to a significant scattered light level at the focus. At the small (<2 deg) Sun-Probe-Earth (SPE) angles typically encountered in deep-space missions, the spacecraft's tracking and pointing focal plane will see increased background noise. This can lead to an increase in pointing error and can cause difficulty in detecting the Earth image or uplink beacon signal. Thus, the required Sun angle has a major effect on the amount of stray light that may be expected within the terminal. Therefore, a design having adequate levels of stray-light rejection is essential to improve the tracking performance at low SPE angle. Here, we concentrate our discussion mainly on the primary mirror since it will likely be the dominating factor in the overall stray-light performance of most deep-space flight terminal optical systems.

The scattering probability density (that is, the probability that a photon with a given direction of polarization incident on a surface at a certain wavelength will be scattered into a particular direction and with a particular polarization) is related to bi-directional reflectance distribution function (BRDF). BRDF is the fraction of power scattered per unit solid angle and has units of sr^{-1} (1/steradian). BRDF may be predicted analytically by modeling of a given optical surface topology and represents the average scattering from the entire surface area. The integral of BRDF overall scattering angles is hemispherical reflectance. Clearly, low BRDF values are desired. Major contributors to BRDF are surface roughness, particulates, and reflective coatings materials and processes.

5.1.6.1 Operation at Small Sun Angles. In a mission to Mars, for example, it is periodically required that the flight terminal operate with the Sun near the FOV. At these times, the Sun-Probe (spacecraft)-Earth (SPE) angles are small ($\leq$2 deg). Under these conditions, it is crucial to prevent the sunlight from entering the telescope. Several preventive schemes are available. These include: (1) windows at the entrance aperture of the telescope and (2) narrow-bandpass filter with effective filtering of the out-of-band wavelengths. Baffles external to the telescope, such as a honeycomb shape structure whose internal surfaces are

coated with a highly absorbing black material do not work as well as expected. In the case of laser beacon tracking, the transmit and receive wavelengths are close to each other. Therefore, it is possible to use a narrow-bandpass filter as the front window of the telescope. In the case of star tracking with broadband wavelengths, the star-trackers utilize separate optical systems that are not co-aligned with the laser-communication transceiver. Therefore it is feasible to utilize a narrow-band filter, centered at the transmit laser wavelength, in conjunction with the flight telescope.

For stray light rejection, several telescope designs (e.g., a Gregorian telescope, with a Lyot stop) offer viable solutions. Given a well-defined set of optical and optomechanical designs, commercial software programs (such as ASAP, TRACEPRO, ZEMAX and GUERAP) can provide an estimate of the stray light magnitude (good to a factor of 2–10). A more precise estimate of the system performance requires laboratory measurements.

5.1.6.2 Surface Cleanliness Requirements. Contamination of an optical surface, by the particulates in the environment, greatly enhances surface scattering. Contamination may be avoided by special handling of the transceiver during integration, assembly with the spacecraft, and at launch [9]. A Class-100 clean room should be adequate for this purpose. Federal and military standards for surface cleanliness are defined in Federal Standard document 209D and 1246B, respectively [10,11]. With distribution and using the Mie scattering theory, one can determine the BRDF associated with a certain level of surface cleanliness level as defined in Military Standard 1246B. Surface contamination, in contrast with surface scattering, has very little wavelength dependence.

The statistical distribution of the size of dust particles in a clean room is reported in Federal Standard 209D. Using this distribution and the Mie scattering theory, it is possible to state the BRDF that would arise from a certain surface cleanliness level as defined in Military Standard 1246B [11]. Very low BRDF values can, in principal, be achieved if the surface is sufficiently clean. However, in any realistic environment, the surface cleanliness level will be an increasing function of time as the surface accumulates more particulate matter from the atmosphere. It is possible to relate, at least on an order-of-magnitude scale, the surface cleanliness level to the amount of time the surface spends in a clean room of a given class.

5.1.7 Transmission, Alignment, and Wavefront Quality Budgets

Any given design is optimized for maximum transmission through the telescope and to each of the focal planes. The design needs to be toleranced for decentering and misalignment of the secondary or tertiary mirrors relative to the primary mirror. These mirrors may be placed on a slow, thermally focusable

mount that once in a while focuses or otherwise adjusts the telescope in flight to remove significant misalignment error. The telescope wavefront quality affects the Strehl ratio that directly influences the telescope's antenna gain value.

5.1.8 Efficient Coupling of Lasers to Obscured Telescopes

A strong peak in the middle of the Gaussian distribution characterizes lasers with TEM_{00} spatial output beam quality. A TEM_{00} mode beam is typically used for free-space optical communications. Reflecting telescopes with on-axis secondary mirrors that obscure the primary mirror are also commonly used in laser-communication system. Significant central vignetting loss of the laser beam, as much as 50 percent or more may occur. For example, making certain assumptions for the optical system, a secondary mirror and baffle blocking of approximately 8.4 percent of the beam area, result in nearly 30 percent loss of the laser energy. While, an obscuration on the order of 25 percent may result in nearly 65 percent loss of the laser energy. Several different schemes have been devised for efficient coupling of lasers to obscured telescopes and are described below briefly [12]

5.1.8.1 Axicon Optical Element. Axicons are both afocal refractive and reflective optical elements with a flat front surface and conical rear surface, and can turn the incident beam inside out. The rays near the edge of the beam entering the axicon get located at the inside edge of the annular beam when exiting. Likewise, the rays at the center of the incident beam get located around the edge of the annular beam when exiting. The axicon must be designed specifically to match the resulting donut-shaped intensity re-distribution across the beam to that of the telescope obscuration, to avoid the coupling losses. Use of axicons for efficient coupling to telescopes was analyzed and experimentally evaluated in detail [13,14]. Difficulties in fabrication of the precise conical axicon devices, and the tight alignment requirements have so far limited the usefulness of these devices. Axicon insertion losses in reflectance and transmittance can be kept small by the proper choice of coatings. The far field pattern is Gaussian in appearance with normal diffraction point spread function (PSF). Diffraction rings can be observed, but are mostly suppressed.

5.1.8.2 Sub-Aperture Illumination. In this scheme, the transmit beam is coupled to the telescope for sub-aperture illumination, missing, and offset at the secondary mirror. This is the simplest arrangement that can be used to avoid the secondary mirror obstruction. In this approach, the far-field pattern is the same as the source. Divergence of the transmitted beam would be greater than if we were to use the full aperture.

5.1.8.3 Prism Beam Slicer. A prism device with afocal refractive and reflective optical elements behaves similarly to axicons. The difference between the axicon and the prism slicer is the use of multiple flat surfaces for the rear side of the slicer. Prism beam slicers, in effect, slice the incident beam into two or more pie-shaped beams that are then arranged in a circular pattern around the telescope aperture (Figs. 5-3 and 5-4). Each beam is sub-aperture in size and can pass through the Cassegrain telescope without any additional vignetting from the telescope secondary mirror and baffle. Because each beam is pie shaped, the far-field pattern for a single beam is not symmetrical (Fig. 5-5). However, the combined far-field pattern for the four beams will be nearly symmetrical.

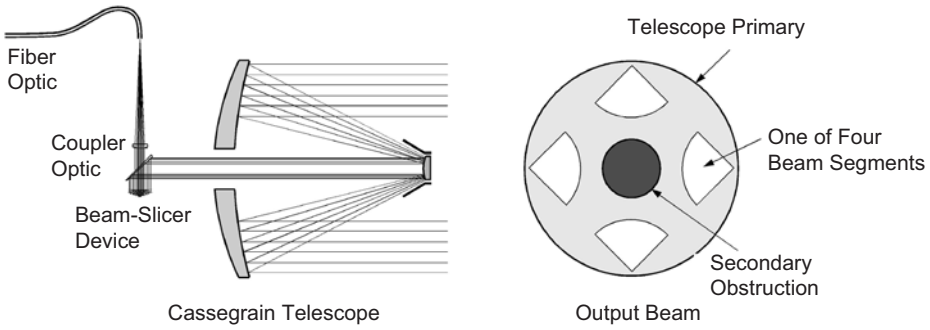


Fig. 5-3. Layout of the beam-splitter device.

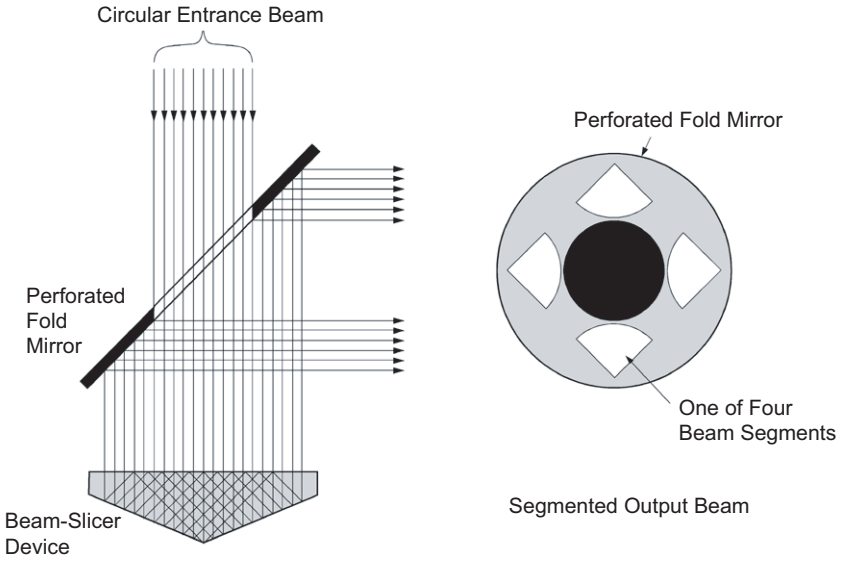


Fig. 5-4. Details of the prism beam-slicer function.

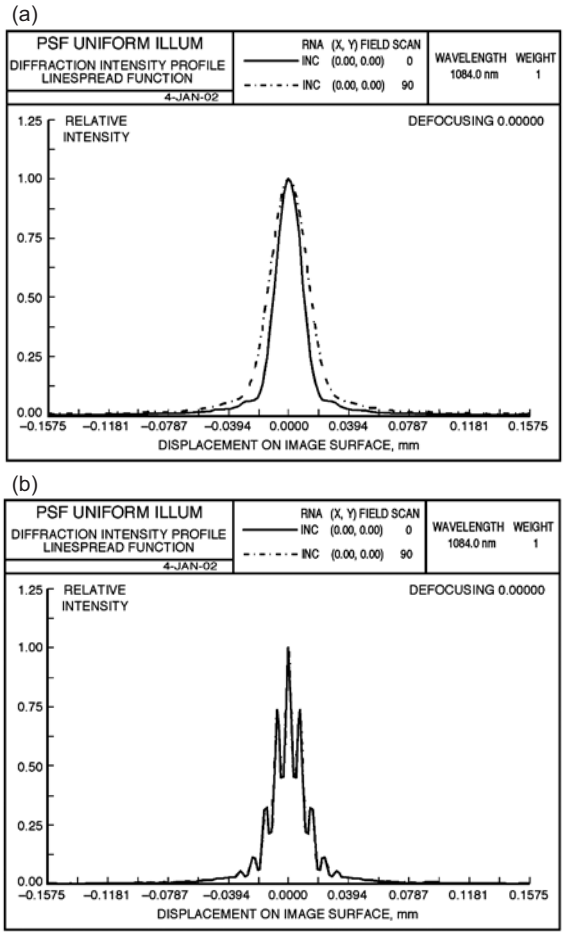


Fig. 5-5. Far-field pattern (a) for a single prism slicer beam and (b) for four prism slicer beams that are phased.

5.1.8.4 Beam Splitter/Combiner. The beam splitter/combiner is a variation of the sub-aperture approach, and one of its applications is beam combining. This configuration produces multiple beams equally spaced in a circle at the telescope aperture. The far-field pattern for each beam is the same as the source. When the multiple beams are combined in the far field, the beam pattern is the same as one of the sources. Because the source for each beam is a separate laser source, each beam is incoherent with the other beams. This approach also offers some degree of redundancy.

5.1.9 Structure, Materials, and Structural Analysis

On average, in an optical communication terminal, the optical bench and the structure that supports the optical components, together account for 25 to 50 percent of the total mass of the terminal. Minimizing this will clearly reduce the overall mass. The requirements for a generic optical communication telescope material are mirror surface figure of better than 1/12 of a wave (peak-to-peak), minimum weight, low scatter surface, high thermal stability, high fundamental resonance frequency, and ease of fabrication (low material fabrication cost). It is crucial to prevent any temperature gradients across the bench that holds the optics. A critical high structural stability requirement is that of the primary mirror and secondary (or even tertiary) mirror separation under temperature variations. It is possible to incorporate a slow (thermal) one-axis (piston for focus) or multi-dimensional actuator with the secondary or tertiary mirrors to actively align the telescope in flight.

Ultra-low expansion (ULE), titanium silicate glass, and Zerodur glass ceramic have excellent thermal properties, but they are heavy. Zerodur substrates with hollowed regions have been developed, but the cost is high. Silicon carbide (SiC) and SiC matrixes (e.g., standard metal oxide (SXA)) and some other composite materials to one degree or other satisfy all of the above requirements. Beryllium (Be) is a very lightweight telescope material, with a density that is comparable to aluminum, but with significantly higher thermal coefficient of expansion than the materials mentioned earlier. Be satisfies most of the above requirements, but it is difficult and expensive to fabricate due to the material safety hazards. Also, Be mirrors are not stable over temperature cycles. Primary and secondary mirrors made of this material should be of very high temperature stability and very high tolerance to ionizing radiation.

SiC is one of the best telescope (structure, primary and secondary mirrors) materials. Some of the outstanding features of SiC use in space are: (1) high specific stiffness and low mass; (2) very low thermal expansion coefficient (on the order of 1 part per million per kelvin (ppm/K)); (3) high thermal conductivity; (4) very high bending strength (400 megapascals (MPa)) and low built-in stress (<0.1 MPa); (5) capability for withstanding low and high temperatures without any loss of properties; (6) high resistance to fatigue; (7) very high immunity to radiation; and (8) it can be ground and polished without significant distortions. However, the surface quality achievable with large (>30 cm) primary mirrors is not yet ideal. Except for very low mass, current technologies are developed enough to satisfy the above requirements. Material characteristics of foremost importance to a laser communication terminal include the following.

Optical performance: Ease of fabricating a high quality surface figure in a spherical or aspheric shape and coating for high reflectance and low scattering and with lowest number of defects is of prime interest.

Structural integrity: Fundamental resonance frequency, dynamic response, specific stiffness, and fracture threshold are major characteristics that affect the structural integrity of the material. Most of these properties influence the terminal's weight [15]. For example, a certain weight is required to achieve a particular natural frequency or dynamic response to maintain integrity over a particular vibration frequency. The greater the specific stiffness for a material, the lower its weight is.

Thermal stability: Both soaked (e.g., relatively small gradient) and gradient temperature variations may be encountered. The main concern is temperature gradients that might change the surface figure of the optical system. Of major interest is avoiding a mismatch between the thermal expansion of the optics and the structure supporting the optics. For example, metering rods with a specific thermal coefficient of expansion are used to connect the secondary mirror to the primary mirror of the telescope in order to offset any bi-metallic bending effects that may arise. SiC is one of the most thermally stable materials, and it may be used for both optics and optical structure (Fig. 5-6).

Temporal stability: Inherent dimensional stability and micro-yield strength are major constituents of this property. Due to the temporal nature of this property, it is one of the more difficult parameters of a material to measure. In general, a material with low rate of creep and relaxation is desired.

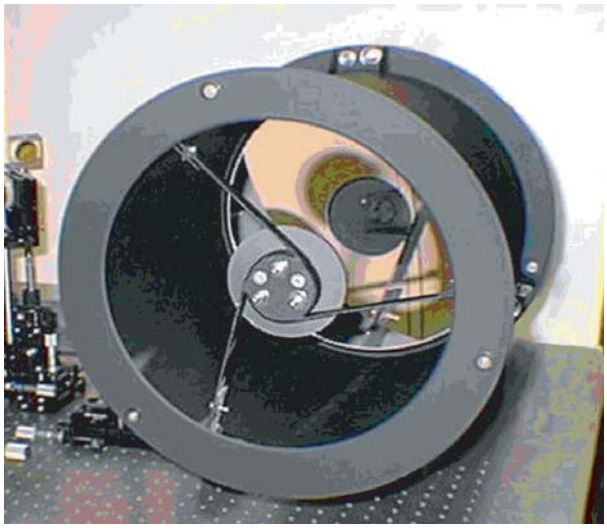


Fig. 5-6. An all (mirrors and structure) silicon carbide telescope with a 30-cm spherical mirror diameter built by SSG Inc. This telescope showed good performance over a ±50 deg C temperature range. This telescope weighs about 6 kg. A 4-kg version is feasible with additional work.

Packaging: To minimize the overall size of the transceiver, the aft-optics path may be folded such that the sensors and the laser head are mounted in one or (at the most) two planes behind the telescope. The telescope and optical structure material need to have very low thermal expansion, high thermal conductivity, and low weight [16,17,18]. Due to the thermal management difficulties and the heat generated by the laser, it is prudent to separate the laser from the optical head (telescope and aft optics) assembly. Fiber-optic-fed lasers work the best; however, with high peak power pulsed lasers, it is not always possible to use fiber optics.

5.1.10 Use of Fiber Optics

Free-space optical communication systems can greatly benefit from the wealth of development in fiber-optics technology [19] Use of fiber optics in the laser communication system can afford multiple degrees of freedom while reducing mechanical stability requirements, easing thermal management requirements, and potentially reducing system mass and volume. Fiber-optic systems may be applied to both non-coherent and coherent systems with polarized or unpolarized beams. In the transmitter subsystem, with the aid of fiber optics, the heat-generating laser transmitter may be located remotely from the heat-sensitive optical system. High-power continuous-wave lasers and pulsed-laser transmitters up to peak power levels that are below the nonlinear or damage threshold to the fiber may be applied. The received signal may also be coupled into a fiber or a fiber bundle to a remotely located senor. Conical scanning of the fiber's signal input end is another means of maximizing signal delivery and at the same time inferring tracking information.

5.1.11 Star-Tracker Optics for Acquisition and Tracking

With laser beacon tracking where the beacon emanates from the Earth, or with Earth-image tracking, stray light and Sun-interference becomes a major challenge at small Sun-probe-Earth angles. This is a particularly difficult challenge at outer-planetary ranges. For example at Pluto, the Earth and the Sun are always within a 2-deg cone angle. Precision star tracking is a viable alternative to Earth-emanated beacons, since stars of interest are far from the Sun. Addition of one to two precision star-trackers looking at orthogonal or at 180-deg angles relative to the optical axis of the flight terminal, may provide a beaconless alternative freeing the terminal from full-time availability of sufficiently strong signal from Earth and from the small Sun-angle. These star trackers need to be an integral part of the optical system and have built-in sensors such that any drift between the optical systems will be known and can be accounted for. This may be achieved by bringing a very low-power probe laser onto the star-tracker imager. This laser is completely collinear with respect to the transmit laser, and thus the situation is functionally as if the

transmit laser light were placed onto the star tracker. The dedicated star trackers will have a smaller aperture than the optical communication terminal (about one third or less), but with much higher FOV (on the order of degrees). In this case, the star-trackers may be used as the spacecraft star trackers as well, in order to reduce mass and power consumption. The star tracker may also be designed as part of the laser flight optical-communications terminal, where the front telescope design accounts for the required star-tracker FOV. As an example, Fig. 5-7 shows a backward-looking star tracker along with drift-detection sensors implemented with the flight transceiver.

Figure 5-8 shows the schematic of a catadioptric star-tracker optical system designed for inclusion with the optical communications terminal.

5.1.12 Thermal Management

On a deep-space mission, temperature external to the insulation that covers the transceiver might vary between 50 deg C to –200 deg C. Therefore, a controlled heater will be required for the terminal to control the optical assembly to, for example, ±5 deg C since optical systems may be designed to be athermal over this (soaked) temperature change. The temperature of the electronics box is typically maintained to within 10 ±40 deg C. To dissipate the heat generated by the laser transmitter and electronics, a dedicated radiator or the spacecraft radiator may be used. Dedicated radiators are more efficient and more convenient to implement. These radiators do not have to be accurately controlled in temperature. Some of the active elements within the terminal, for example the laser oscillator in an oscillator/amplifier or the pump diode lasers for a diode-pumped solid-state laser (doped crystal or doped fiber) may need precise temperature control.

5.1.13 Optical System Design Example

Table 5-1 outlines the specifications for an example design of an optical communication system with 10-cm front aperture diameter with the capability for precision pointing.

5.1.13.1 Afocal Fore-Optics. For the specified field of view, a two-mirror afocal telescope lacks the degrees of freedom to produce diffraction-limited wavefront error and a convenient exit pupil location. Either a third mirror or a refractive collimating group, is needed. A three-mirror Cassegrain provides good wavefront, intermediate image for stray anastigmat. In this case, the RMS wavefront error was calculated less than 0.001 waves at a 600-nm wavelength.

5.1.13.2 Receiver Channel. The function of the receiver is to accept light emerging from the fore-optics and direct it to a circular detector. It is a light-

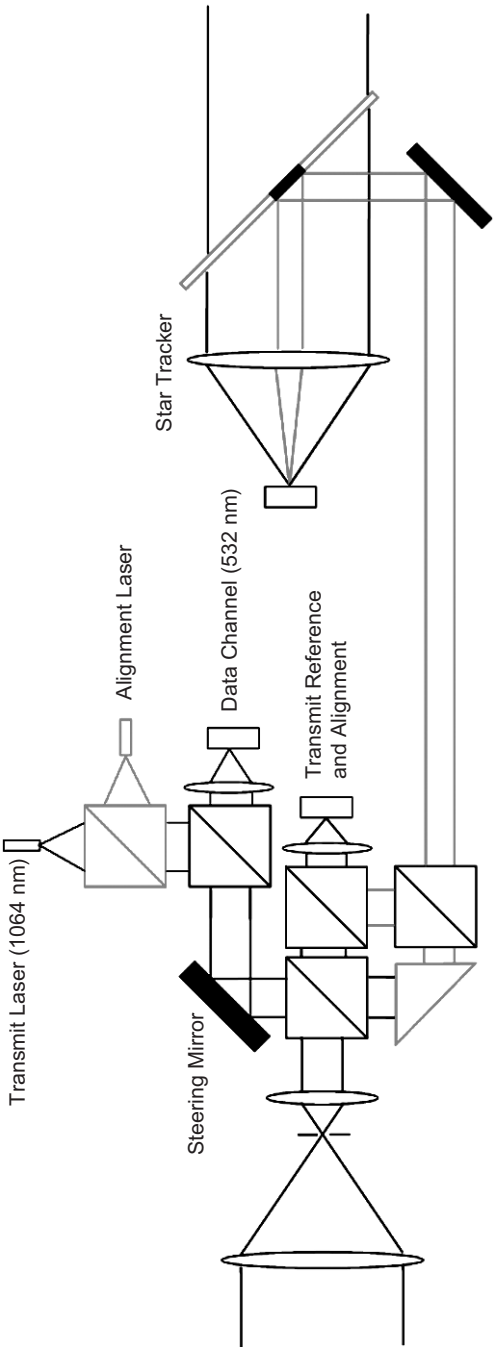


Fig. 5-7. Block diagram of the optical communication terminal, including a dedicated star tracker and a redundant transmit subsystem.

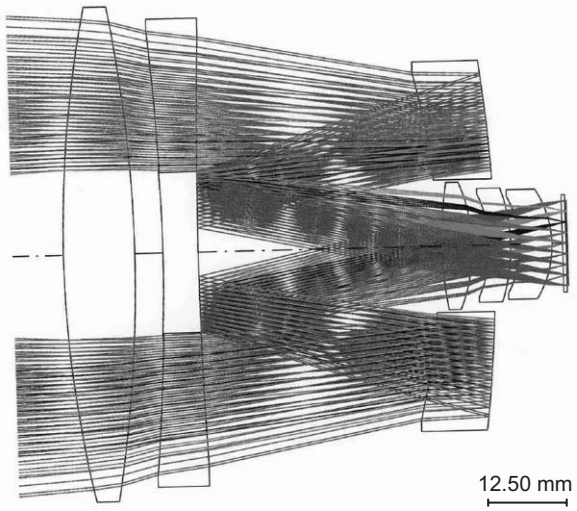


Fig. 5-8. The designed catadioptric star-tracker optical system.

collecting channel, and image quality is not important as long as there is not excessive spillover beyond the edges of the detector.

The receiver could be either an imager or a pupil relay. A pupil relay is preferred due to the need for inclusion of the scan mirror. The Lagrange invariant severely limits the achievable spot size at the focus. An alternative is to relax the specification on detector diameter to a larger value, e.g., 1.0 mm. To illustrate the difficulty of a design in air, Fig. 5-9 shows a receiver with a 0.5-mm detector in air. Due to high ray angles, the field coverage in the sky is only 8×8 mrad, versus 10×10 mrad, which is the goal. Pupil aberrations cause the beam footprint from off-axis field angles to overfill the circular detector (Fig. 5-10(a)). A higher-performance design uses an immersed detector, in contact with the final, plano-convex element. In that case, the beam prints across the field to fill the detector with little (a few percent) mismatch (Fig. 5-10(b)). To enhance collection efficiency with a pupil relay receiver channel, one may either use an immersion lens with a 0.5-mm diameter detector, or use a lens in air with a larger detector diameter, such as 1 mm.

It is possible to collect the light from the afocal section and concentrate it onto the detector using a compound parabolic concentrator (CPC). A CPC may be of a solid catadioptric design or be a hollow reflector. Each would provide different properties. Index matching cement between the CPC and the detector will be required to obtain the desired performance. For a solid catadioptric CPC, its length gets progressively shorter as the refractive index increases. However, it is useful to note that the semi-diagonal field leaving the afocal

Table 5-1. Requirements, specifications, and goals for a preliminary optical design of a laser communication transceiver.

Parameter	Requirements, Specifications, and Goals
Configuration	Space-qualified laser communication transceiver. Common afocal plus four channels: Transmit 1064 nm, receive 532 nm, alignment laser 980 nm (transmit), and stellar reference 550–900 nm (receive). Scan mirror for pointing transmitted laser beams. Minimal size, mass, and power.
General system specifications	
Entrance pupil diameter	100 mm
Vignetting (of area), maximum	
Secondary obscuration	10 percent
Off axis	10 percent
Thermal environment	
Operating	20 ±10 deg C
Survival	–40 to +45 deg C
Instrument package envelope	<5 × 5 × 6 in. (13 × 13 × 15 cm)
Size and mass	Minimize
Lyot stop	
In system	Required in transmit and align channels
Location	Near scan mirror before align/transmit optics
Radiation-hard glasses	Not needed
Scan mirror	
Function	Scan align and transmit lasers over the FOV in sky in reduced-beam space inside align/transmit channels
Location	
Scan range (in sky)	±5 mrad
Clear aperture	15 × 22 mm elliptical
Accessible internal focus	Needed for all channels
Cemented refractive elements	Avoid
Afocal fore-optics	
Purpose	Perform beam diameter reduction
Configuration	3-mirror centrally obscured; re-imaging to control pupil
Aperture stop	On secondary to minimize obscuration
Field of view in sky	±5 mrad (±0.29 deg)
Afocal magnification	Implied by scan mirror size specifications
Wavefront quality	Maximize (waves RMS, 600 nm)
Transmit channel	
Purpose	Expand and collimate laser beam
Transmit laser	
Profile	Gaussian TEM_{00}
Diameter	0.6 ±0.1 mm ($1/e^2$ assumed)
Divergence	1.5 ±0.3 mrad (reference)
Strehl ratio	0.90–0.91
Wavelength	1064 nm
Clear aperture over sizing	$2 \times 1/e^2$
Field of view	On axis

Table 5-1. Requirements, specifications, and goals for a preliminary optical design of a laser communication transceiver. (cont'd)

Parameter	Requirements, Specifications, and Goals
Diameter of beam emerging from transmit channel, before afocal fore-optics	15-mm ($2 \times 1/e^2$ diameter)
Focal length of transmit optics	afocal
Transmit optics Strehl ratio	≥0.95
Receive channel	
Function	Receives monochromatic radiation & relays pupil onto detector
Wavelength	532 nm
Receive channel FOV, total	10 × 10 mrad in sky
Detector	Single detector, ≥0.5 mm diameter
Image quality	Image the pupil onto the detector, with <5 percent area overfilling
Stellar reference channel	
Purpose	Receive and focus broad band radiation
Wavelength	Different star temperatures and uniform CCD response assumed
Focal length of stellar reference channel	Assumed this refers to full channel including afocal fore-optics
F-number	Derivable from 100-mm input beam diameter and 800-mm focal length
Detector	1024 × 1024 × 12 × μm Note, active image area is 8 × 8 mm, per specifications on effective focal length and sky FOV
Field of view in sky, total	10 × 10 mrad
Image quality	PSF spread over 30 μm to facilitate centroiding
Image centroid shift between stars at 2800 & 10,000 K	<±0.1 pixel (±1.2 μm) need <±0.02 pixel (±0.6 μm) goal
Telecentricity error	Up to 1.1 deg
Centroid shift with stellar temperature (10,000, 5800, & 2800 K).	Centroid shift with respect to 600-nm chief ray: 0.78 μm (2800 K) quantum efficiency (Q.E.) 0.61 μm (5800 K) Q.E
Alignment channel	
Purpose	Transmits laser beam
Wavelength	980 nm
Laser Type Waist, divergence	 Diode laser Model as point source, F/5 emerging from fiber
Field of view	On axis
Focal length of alignment channel	500 mm, derivable from F/5 and 100-mm entrance pupil diameter
Wavefront quality	Purposely blurred (e.g., spherical aberration), 0.14 waves RMS (42 μm 80 percent encircled energy diameter)

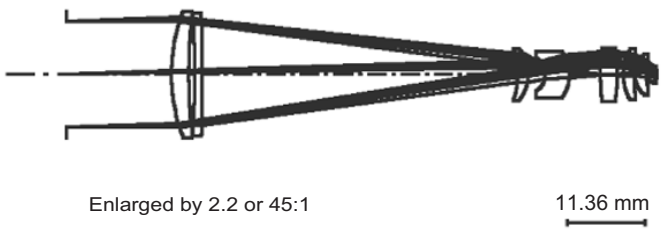


Fig. 5-9. Receiver with a 0.5-mm diameter detector in air (non-immersed).

fore-optics has a semi-diameter of 7.5 mm and a semi-diagonal field of view of 2.73 deg. The Brightness Theorem of radiative transfer demands that the quantity $n^2 dA \mathrm{Sin}^2\theta$ be conserved throughout the system. In this equation, n is the index of refraction, dA is an infinitesimal element of area along a light beam, and $\mathrm{Sin}^2\theta$ is related to the solid angle. The square root of this expression may be regarded as a generalization of the Lagrange invariant, the requirement for brightness to be conserved throughout the system. Measured at the exit pupil of the fore-optics, this quantity has a value of 0.357 mm. Since the detector has a semi-diameter of 0.25 mm and the maximum value of $\mathrm{Sin}\theta$ unity, it follows that a concentrator must have an index of at lease 0.357/0.250 = 1.43 in order to concentrate all the light from the exit pupil to the detector. The CPC gets progressively shorter as the refractive index increases. However, since it must also be cemented to the detector with index matching cement, lets limit ourselves to $n = 1.6$. The CPC disadvantages (1.8 times longer than the immersed pupil relay and is more difficult to fabricate) probably outweigh the advantages of athermal, alignment-free design. Figure 5-11(a) shows a solid glass CPC operating over the conditions described above, and Fig. 5-11(b) shows the associated footprint data. Use of the pupil imager with an immersed receiver will probably result in the most compact configuration.

5.1.13.3 Stellar Reference Channel. The stellar reference channel forms an image that is purposely blurred but uniform across the field. The 80-percent encircled energy diameters of the stellar reference optics alone (not including the afocal fore-optics) range from 28–30 μm.

5.1.13.4 Align and Transmit Channels. The optical elements of the alignment and receiver channels are common up to a beamsplitter. A design-driving requirement is that these channels must contain a scan mirror located at a pupil. In the concept discussed here, the pupil formed by the three-mirror afocal fore-optics is relayed to a remote location by a refractive 1 afocal relay (Fig. 5-12). The beam feeds both the alignment and receiver channels.

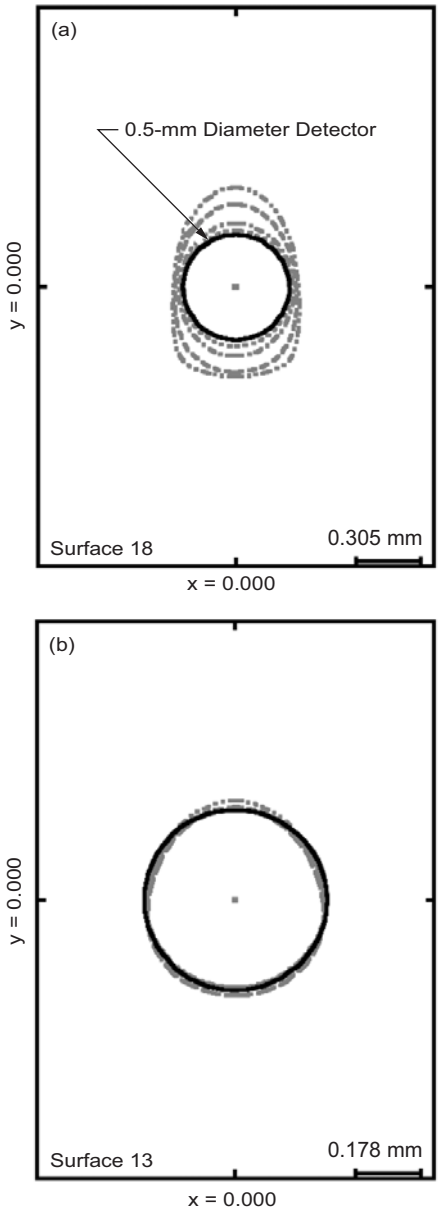


Fig. 5-10. Overfill of detector aperture due to off-axes field angles: (a) assuming a detector in air and (b) corrected for by use of an immersed detector.

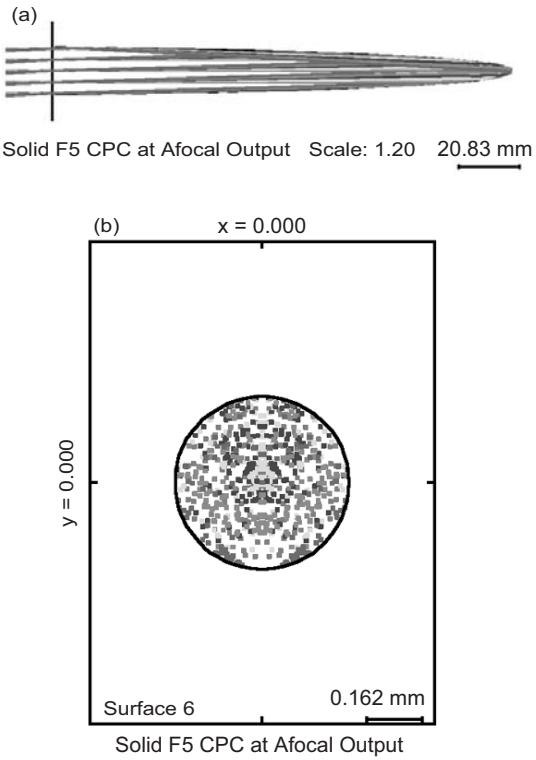


Fig. 5-11. The CPC: (a) a solid-glass CPC and (b) its associated beam footprint.

The transmit optics are essentially a laser beam expander, bringing a 1.2-mm laser beam (0.6 mm $1/e^2$ diameter with 2× over sizing) to the 15-mm expanded beam diameter. A three-element configuration achieves the beam expansion with sub-0.001 wave RMS wavefront error.

The align channel is a simple lens (e.g., singlet) that forms an image at a detector without a high degree of image quality, as per the specification. The image is purposely blurred with spherical aberration to facilitate centroiding. The design has an 80-percent encircled energy diameter of approximately 47 μm. The bending of the lens can be adjusted to create the desired level of spherical aberration. As mentioned earlier, another key section in these two channels is an afocal relay, to relay the exit pupil of the reflective fore-optics to the scan mirror location. The design shown below is a 1× relay, relaying one 15-mm diameter pupil to the other with about 0.025 waves RMS wavefront error at either 980 nm or 1064 nm (separately focused).

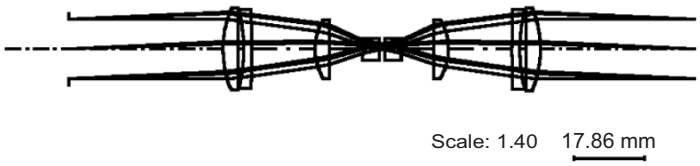


Fig. 5-12. A refractive 1x afocal relay design.

5.1.13.5 Folded Layouts. The system's optical path must be folded to minimize the envelope diameter and length (Fig. 5-13). There are different options for folding, and what is discussed is a plausible arrangement, but other arrangements could be also used. To fold the aft optics and to put the exit pupil in an accessible location, several fold methods may be applied, including, fold mirror at the Cassegrain focus, or a fold in collimated space. In principle, both of these fold concepts could have the same obscuration produced by the fold mirror or the opening in the fold mirror, i.e., about 12 percent in area. The first concept would need a way of holding the small fold mirror (e.g., by a mounting bar that would introduce as much as 15-percent additional obscuration, depending on the width of the mounting bars). The second concept is preferable in that there is better access to the internal focus for a field stop.

Figure 5-14 is a three-dimensional view of the folded aft optics. The goal is to fold the optics into one or at most two planes behind or around the primary mirror.

5.1.13.6 Tolerance Sensitivity Analysis. Tolerance sensitivity analysis must be performed on the high-performance imaging channels, namely the afocal fore-optics, the stellar reference channel, and the transmit optics. Sensitivity analysis is done with a basic defocus compensator, and in some cases with an additional compensator that counteracts asymmetric aberrations. In all cases, the performance criterion is RMS wavefront error. The following design improvements may be implemented: (1) perform tolerance analysis using more specific performance parameters as the criterion (e.g., diffracted 80-percent encircled energy diameter for the stellar reference channel) to achieve higher Strehl ration; (2) adjust the tolerances to produce a higher Strehl ratio; and (3) simulate the beneficial effects of re-spacing the elements using measured data for radius, thickness, and refractive index. This is a simple operation that can make a meaningful improvement to the as-built system performance. Table 5-2 summarizes the tolerance levels used in the sensitivity analysis.

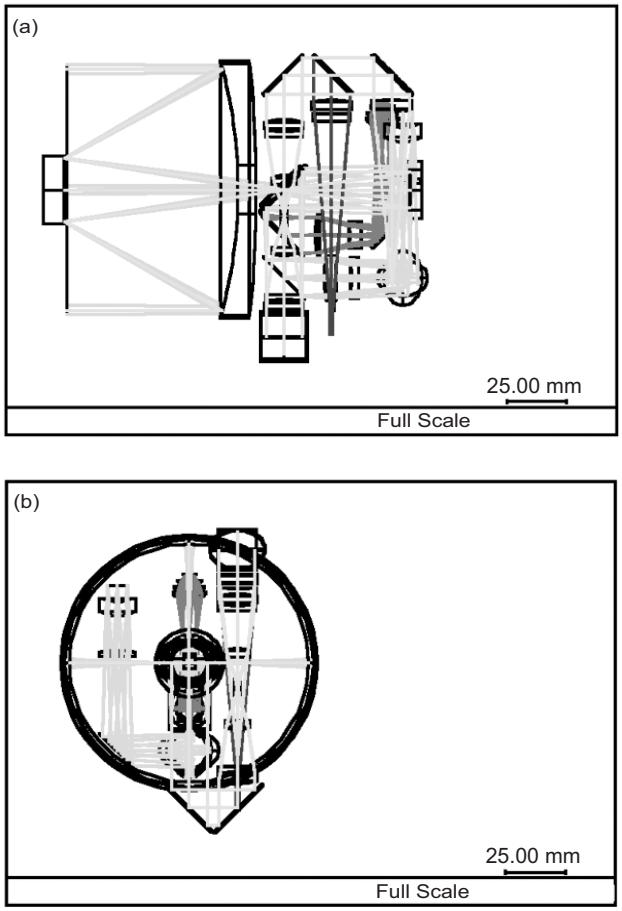


Fig. 5-13. Views of folded channels for an afocal design: (a) side view and (b) end view.

5.1.13.6.1 Afocal Fore-Optics. For the selected set of tolerance levels, the most sensitive tolerances in the fore-optics are the surface figure of the mirrors, and the tilt and decentration of the primary and secondary mirrors. A focus compensator is needed; axial adjustment of the secondary and tertiary give equivalent levels of focus compensation. De-centration of the secondary mirror is an effective compensator for asymmetric aberrations caused by mirror tilts and decentrations. Decentration of the tertiary is much less effective. The as-built RMS wavefront-error (WFE) is about 0.10 waves at 600 nm (a Strehl ratio 0.67 at 600 nm or 0.87 at 1000 nm) with the secondary decentration compensator, and 0.31 waves without it.

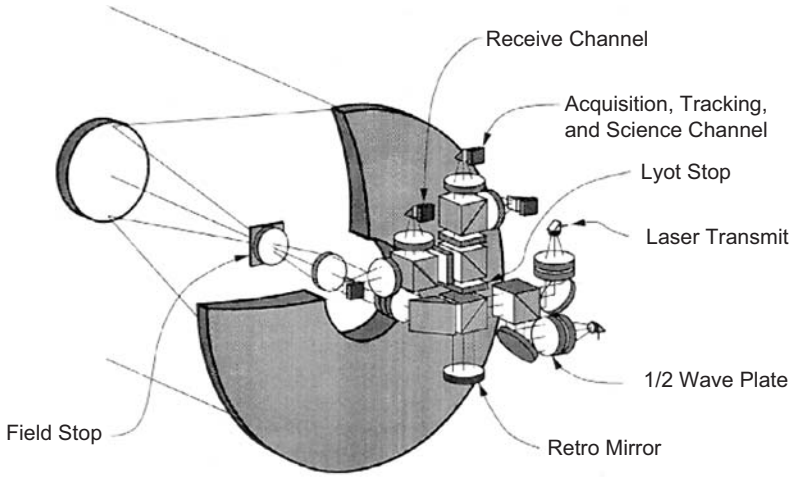


Fig. 5-14. Three-dimensional view of folded aft optics.

Table 5-2. Summary of the tolerance levels used in the sensitivity analysis.

Parameter	Tolerance value (±)
Radius	0.5 percent of design radius
Power/irregularity	1.0 / 0.25 fringe
Thickness or airspace	0.025 mm
Refractive index	0.001
Element wedge TIR*	0.01 mm
Element tilt	0.0003 radian
Element de-centration	0.025 mm

*Total internal reflection

5.1.13.6.2 Stellar Reference. The tolerance sensitivity analysis for the stellar reference channel is done with the aspheric phase plate in place, to enable us to quantify the as-built difference from ideal performance. The as-built RMS WFE is in the 0.06–0.08 range. A focus shift of the detector was assumed for the focus compensator. Performance limiting tolerances are irregularity, radius, and element wedge. Measuring each radius and re-spacing the fabricated design can reduce the effects of radius error.

5.1.13.6.3 Transmit Optics. With set of tolerance levels listed above, the as-built RMS WFE is about 0.025 waves. The focus compensator was the airspace between the front and rear groups. The surface irregularity tolerances limit the performance.

5.1.13.6.4 1× Afocal Relay. The as-built RMS wavefront error with the above tolerance set ranges from 0.09 to 0.17 waves across the field. This assumes that the final doublet will be axially adjusted to maintain best focus. Adding a decenter compensator only modestly improves performance to the 0.09 to 0.14 range. The performance is limited by radius errors, element decentrations, and element wedges. Of all the optical subsystems in the overall system, this is one of the inherently most sensitive elements because of the amount of 'work' being done in a relatively short path. Lengthening the path would likely reduce sensitivity, at the expense of a larger envelope. It is possible to lower the as-built wavefront error by desensitizing the afocal relay to tolerances, as well as simulating the effects of performing a re-space using measured data.

5.1.13.7 Thermal Soak Sensitivity Analysis. A thermal soak sensitivity analysis is made for the same three high-performance sections as for the tolerance sensitivity analysis, namely, the afocal fore-optics, stellar reference, transmit optics, and the 1× afocal pupil relay between the fore-optics and the transmit optics. The thermal soak was +10 deg C, relative to the design temperature of 20 deg C.

5.1.13.7.1 Afocal Fore-Optics. The afocal fore-optics section is potentially highly sensitive to the thermal soak. The performance is driven by the spacer material coefficient of thermal expansion (CTE). The substrate material is assumed to be fused silica (CTE 5.2×10^{-7}); not surprisingly, best performance is when the spacer CTE (invar nickel-iron alloy) nearly matches the substrate CTE, making the system nearly athermal, as if the entire assembly was made out of the same material. The trends of RMS WFE for different spacer and substrate assumptions are shown in Table 5-3.

We make the assumption of fused silica substrate because the wavelength, as low as 500 nm, may require grind-and-polish fabrication. However, if diamond turning can be done, then both the substrates and spacers can be aluminum, and the fore-optics will be athermal up to the homogeneity of the temperature and CTE within the parts.

5.1.13.7.2 Stellar Reference Channel. The stellar reference optics undergoes negligible focus shift for a 10-deg C thermal soak (i.e., 4 μm, well within its quarter-wave depth of focus).

5.1.13.7.3 Transmit Channel. The transmit optics (three-element beam expander) is more sensitive to a 10-deg C thermal soak. The performance is driven by the expansion coefficient of the material maintaining the 75-mm distance between front and rear groups. Table 5-4 shows the RMS WFE at elevated temperatures for different assumptions on the CTE of the spacer. It may also be possible to use a dual-metal spacer to passively athermalize the channel.

5.1.13.7.4 Afocal Relay. The afocal pupil relay optics, which relays the exit pupil of the fore-optics onto the entrance pupil of the transmit optics, was separately given a thermal soak analysis. Assuming aluminum spacers, the RMS WFE at elevated temperature is 0.13 waves, up from about 0.02 for the nominal design, averaged across the field. This is driven by the expansion coefficient of the spacer materials, as seen by Table 5-5.

5.1.13.8 Solid Model of System. Figures 5-15 and 5-16 show solid model renderings of the system, as constructed in LightTools. The first figure shows a side view of the full system. The different channels are color-coded and are labeled. The details of the fold arrangement in which an alignment beam is fed into the star tracker can be refined if needed.

Table 5-3. Trends of RMS WFE for different spacer and substrate assumptions.

CTE of Spacer Material	RMS WFE, Waves at 600 nm for ΔT = 10 deg C (not refocused)
236×10^{-7} (aluminum)	1.12
99×10^{-7} (stainless steel 416)	0.45
5.6×10^{-7} (invar 35)	0.002

Table 5-4. RMS WFE at elevated temperatures for different assumptions of the CTE of the spacer.

CTE of Material for 75-mm Airspace	RMS WFE, Waves at 1064 nm for ΔT = 10 deg C (not refocused)
236×10^{-7} (aluminum)	0.025
99×10^{-7} (stainless steel 416)	0.015
5.6×10^{-7} (invar 35)	0.008

Table 5-5. RMS WFE as a function of the spacer material used.

CTE of Material for Spacers	RMS WFE, Waves at 1064 nm for ΔT = 10 deg C (not refocused), Average Across Field
236×10^{-7} (aluminum)	0.128
99×10^{-7} (stainless steel 416)	0.076
5.6×10^{-7} (invar 35)	0.042

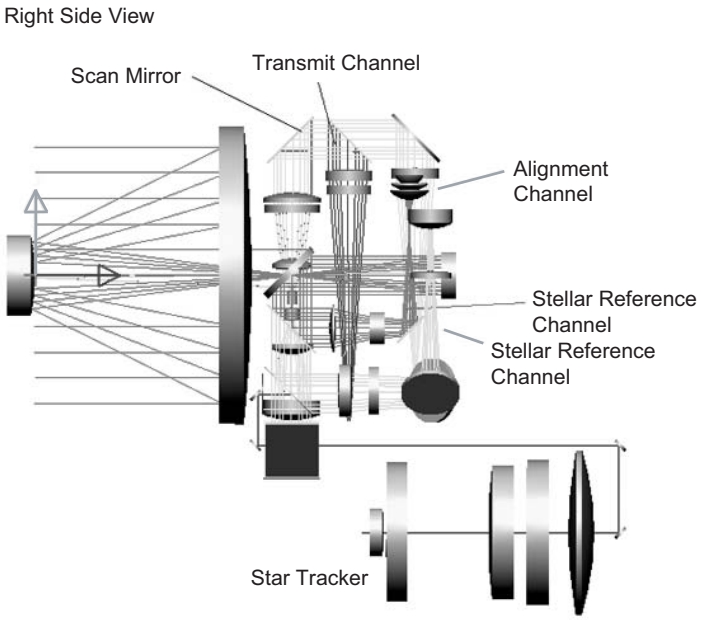


Fig. 5-15. A top view perspective of the layout.

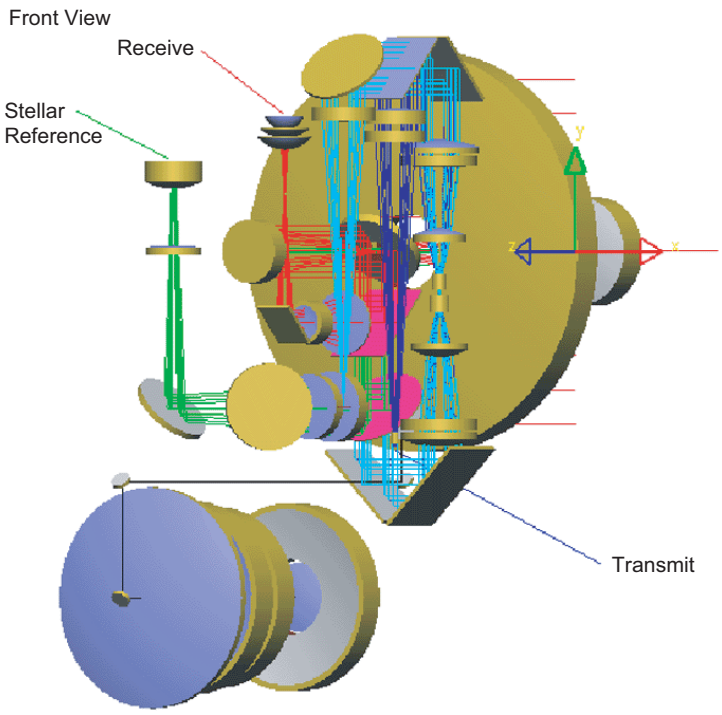


Fig. 5-16. Another perspective view shows the transmit laser, the receive channel detector, and the stellar reference channel detector.

5.2. Laser Transmitter

Hamid Hemmati

5.2.1. Introduction

Modulated laser beams carry the information that is transmitted from the laser-communication terminal. Generally, amplitude modulation is implemented for direct-detection, while amplitude or phase modulation is used for coherent communications. Applicability to onboard spacecraft use necessitates a compact, lightweight, and efficient laser transmitter. The large distances over which to communicate require a laser that is capable of high peak powers in a single-spatial mode beam. Sources for direct and coherent detection vary significantly and are discussed separately.

For deep-space direct-detection communications, pulsed lasers need to provide multiple watts of average power as well as kilowatts of peak power in sub-microsecond timeframes. Modulated continues wave (CW) lasers are inherently average power sources only and are more suited to the closer

distances of near-Earth environments. A modulation technique known as the pulse position modulation (PPM) enhances the signal-to-noise ratio (SNR) at the receiver. The benefits and details of PPM are described in Chapter 4 on Modulation Techniques. The PPM technique requires narrow pulses of moderately high (a few kilowatts) peak power that can be produced with varying repetition rates from the laser. Table 5-6 summarizes relevant parameters for a downlink from Mars for differing PPM orders. Assumptions are: range of 2.7 astronomical units (AU); data rate of 10 Mbps, a 5-m diameter ground receiver; a 30-cm diameter flight transmitter aperture with secondary obscuration, day-time reception (Sun angle of 3 deg); high quantum efficiency (≥35 percent) photon-counting indium gallium arsenide phosphite (InGaAsP) avalanche photo-diode (APD) for 1550-nm reception and silicon (Si) APD for 1064-nm reception; 0.1-MHz noise for InGaAs and 10-MHz noise for InGaAs detector; 2 dB of pointing budget allocation; 2 dB of atmospheric losses; 30 dB of transmitter modulation extinction ratio; and equivalent transmitter and receiver losses for both wavelengths.

Table 5-6 shows that pulsed lasers with moderate average power and high peak power are significantly more efficient for deep-space missions. Due to lack of peak power, the PPM alphabet implementation is limited to a maximum of about M = 4 for modulated CW sources. A well-behaved pulsed laser with adequate average power, or a low power oscillator amplified to the required power levels can satisfy laser power requirements, as shown in this table. Examples of such oscillators are semiconductor laser pumped solid-state lasers that are pulsed through several well-developed schemes (e.g., Q-switching and cavity-dumping), or pulsed oscillators amplified in a waveguide (fiber or bulk crystal). On the contrary, for near-Earth applications (spacecraft in low Earth Orbit [LEO], medium Earth orbit [MEO], or geosynchronous Earth orbit [GEO]) where just average power and significantly higher data-rates (on the order of Gbps) are required, modulated CW sources are the laser transmitter of choice.

Table 5-6. Comparison of the required input DC powers to obtain a 10-Mbps link for various laser modulation choices.

Link Characteristics	Parameter Set 1	Parameter Set 2
PPM format	4	256
Link margin (dB)	3	3
Channel capacity (Mb/s)	10	10
Required laser power (W)	70	10
Required electrical power for the laser* (W)	466.7	66.7

*assuming 15 percent overall efficiency

5.2.2 Requirements and Challenges

Some of the driving requirements for a given laser transmitter include average output power in conjunction with peak power per pulse, output beam quality, pulse-repetition-frequency (PRF), pulse-width and pulse generation time delay, pulse jitter, beam-pointing accuracy, overall efficiency, pulse extinction ratio, mass, volume, effect of ionizing radiation, thermal management requirements, and lifetime of active components. Descriptions of the critical parameters influencing the selection and design of the laser as well as some of the salient features of laser transmitters for space follow.

Pulse-repetition-frequency (PRF): The selected laser pulsing mechanism (e.g., Q-switching, cavity-dumping, or amplitude modulation of a seed laser followed by amplification) determines the laser's PRF. Q-switched lasers utilizing acousto-optic or electro-optic modulators are limited in PRF to less than 200 kHz. Cavity-dumped lasers are limited to PRFs on the order of tens of megahertz. Master-oscillator, power amplified lasers that start with a low power pulsed oscillator and are followed by several stages of amplification can potentially lead to repetition rates on the order of tens of gigahertz. In this case, the oscillator may be a diode laser that is either directly modulated or operated with an external modulator. The amplifier needs to have adequate saturated gain to sustain the repetition rate in conjunction with the hundreds to thousands of watts of peak power desired for deep-space communications. Figure 5-17 shows the behavior of laser's critical parameters as a function of the PRF.

Average output power: The laser should provide sufficient average power and the corresponding peak power to support a communication link with a sufficiently positive margin. For most telecommunication applications within the Solar System, today's diode-pumped solid-state lasers can provide power levels that are about an order of magnitude higher than those typically required. However, a communication transmitter also requires the ability to handle variable data-rates, provide nearly constant average power over different data-rates, and provide overall pulse-to-pulse power stability [20]. Reference [20] describes pulse widths, energies, and build-up time as a function of the initial inversion ratio. Efficient and short-pulse lasers operate with an inversion ratio of about 3 to 4 times threshold.

Peak power: Peak power is driven by energy per pulse and pulse-width. The PPM order utilized in the link and the required link margin that is based on a given bit error rate (BER) determine the required peak power. Solid-state lasers have demonstrated peak powers greater than 1 MW at low repetition rate but are generally limited to less than 100 kW at multi-kilohertz repetition rates. However, for a typical deep-space terminal, the anticipated maximum peak power is expected to be on the order of a few kilowatts due to spacecraft power limitations and heat dissipation issues. Fiber amplifiers have shown the capability to handle multi-kilowatt peak power in narrow (1-ns level) pulses.

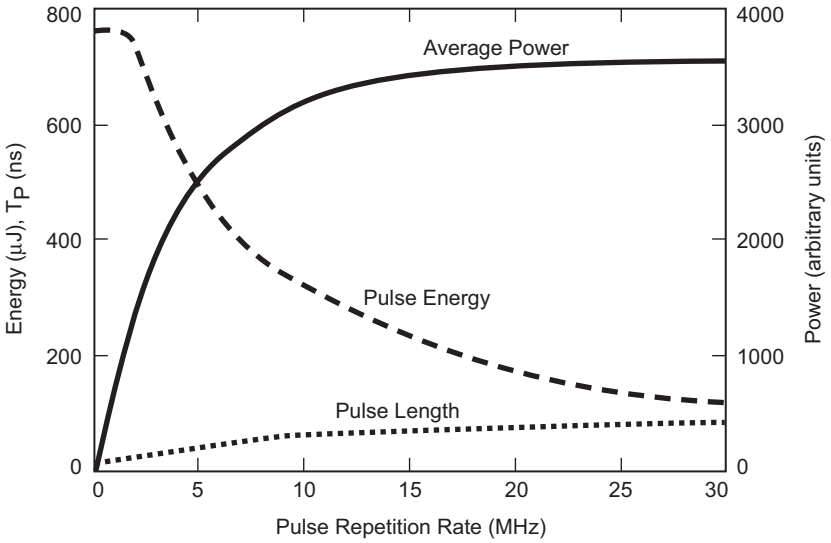


Fig. 5-17. A representative example of laser average output and peak power, and pulse width as a function of pulse-repetition frequency.

Pulse-width: Shorter pulse widths over the entire PRF range are desired since less background light is integrated in a narrow temporal slot. In optical communication systems that use the PPM scheme, the timing of the pulse carries the data in contrast to an on-off keyed modulation scheme used in near-Earth links where the pulse threshold is the determinant. The communication bit rate (R_b) is related to the PPM order (M) and the slot width (T_s) as: $R_b = (\log_2 M)/MT_s$. Therefore, the required pulse-width of the laser is inversely proportional to the magnitude of the PPM order. For example M = 8 may require a laser with a 2-ns pulse-width, while M = 256 requires only a 0.2-ns pulse width. Each of the above requirements dictates a specific type of laser that may employ a very different architecture relative to other lasers. A pulsed laser amplifier may be suitable for the low-order PPM in the above example. Whereas, a pulsed bulk crystal laser or amplified laser may be needed to satisfy the high order PPM requirements.

Pulse generation time delay: PPM requires accurate positioning of the pulses in the time domain. Laser pulse jitter results in positioning errors or the need to increase the temporal detection slot width to compensate for the jitter. In the latter case, more background light will be integrated during the increased slot width, raising the detector noise threshold at the receiver. Thus, to establish a low bit error rate communication link, the timing delay between the pulse trigger and the actual laser pulse emission must be nearly constant for all pulses. To avoid detection losses and to minimize the probability of error, the

laser transmitter's pulse jitter should be minimized to a fraction of the slot detection width. Otherwise, the effective pulse-width (temporal slot-width) becomes large, and pulse position uncertainties rise. Pulse jitter may be caused by three kinds of delays: electrical, switching, and/or build-up. Different schemes have been devised to mitigate pulse jitter effects.

Pulse extinction ratio: This is the ratio of the laser power in the on-mode to that of the off-mode. Laser emission, if not shut off completely, degrades the modulation extinction ratio, and this results in lower link margin. Solid-state lasers may have a modulation extinction ratio of 10^{-4} or better, while some directly modulated semiconductor lasers may have a relatively poor modulation extinction ratio of about 10^{-1}. Fiber lasers and amplifiers generally have an extinction ratio on the order of 10^{-3}, but, along with direct diode lasers, they are susceptible to amplified spontaneous emission if the amplifier gain is not completely saturated.

Output beam quality: To effectively transmit the beam from a spacecraft terminal with minimal losses, the laser's output beam should contain a single spatial mode or at least have no null in the center of the far-field pattern. Beams of high spatial quality may be generated through proper laser-resonator design through the use of single-mode fibers. A critical measurable parameter in this regard is the M^2 factor where an $M^2 = 1$ corresponds to a diffraction limited beam shape. An M^2 of 1.2 or lower will minimize transmission and coupling losses to an acceptable level and is generally achievable at the required output power levels. Depending on the optical design, beam ellipticity or divergence may be of concern as well. Feedback isolation of the laser from back-reflected beams is also required to avoid undesired oscillations either within the laser itself or the transmitter optical train.

Beam-pointing stability: Any given laser is subject to angular and positional uncertainties (jitter) in beam pointing stability. Resonator's spatial mode hopping and mechanical, thermal, or electro-optic effects within the laser may all contribute to jitter. Depending on the deep-space mission, in particular the spacecraft range and platform stability, the transmit laser beam will have to be pointed at the receiver with an accuracy on the order of one micro-radian or better. This requirement necessitates that pointing stability of the laser itself be maintained to a tolerance that is better than the pointing requirements for the mission. Judicious optomechanical and laser-resonator design should result in meeting these requirements. Fiber coupling the laser or using a single-mode fiber laser not only improves the beam quality but also allows more stable beam pointing by eliminating higher order modes.

Overall efficiency: The highest possible overall efficiency is desired to minimize the electrical power demand from the spacecraft. Power consumption drivers are the pump diode lasers, the thermal management of the diode lasers, and the pulsing mechanism. Improving laser efficiency poses many challenges. In pulsed lasers, laser emission efficiency is dominated by the product of

stimulated emission cross section and laser upper-state lifetime. As discussed later in this chapter, quantum defect mode matching, scatter, and other losses affect the overall efficiency as well.

Mass and volume: Clearly, because of enormous deep space mission launch costs, mass and volume of all subsystems should be minimized. The trend towards smaller and lighter-weight spacecraft necessitates the use of optomechanical designs for the laser resonator to minimize its dimensions and mass while maintaining thermal stability and radiation hardness. Diode-pumped solid-state lasers (including fiber laser) are inherently compact systems.

Lifetime of active components: The lifetime of the laser's active components (diode laser(s), modulators/pulsers and their drivers) should well exceed the expected operational lifetime of the mission. It is prudent to use redundant (block redundant or pump-laser redundant) lasers to increase the laser's lifetime. Data from diode laser manufacturers, for diodes with a few watts of continuous output power at 810 nm or 980 nm used as the optical pump sources, indicate lifetimes exceeding 50,000 hours (nearly 6 years). The higher the pump power, the lower the expected lifetime for the diode. Redundancy of the active elements, or block redundancy of the laser, is an effective means of extending the flight terminal's lifetime. Inclusion of linear arrays or grids of pump diodes also allows for de-rating the power level of an individual laser along with adding redundancy.

Radiation: Missions to the Earth's Van Allen belts, Jupiter, and Europa will encounter very challenging radiation environments. Most other mission destinations, for example, Mars and Pluto, have much more benign radiation environments and are not considered as challenging in terms of radiation hardness of components. Care should be taken to use radiation-tolerant diode lasers and optics (e.g., laser crystals, cavity mirrors, and intra-cavity pulsing devices). Typical diode lasers used as either pump or seed sources are based on GaAs material systems, which are fairly robust with respect to low-level radiation, induced defects compared to silicon-based electronics. Shielding can be an effective method of reducing the radiation tolerance requirements for laser components. Often, only a limited amount of data on the specific state-of-the-art component that is baselined is available in the existing literature. Therefore, additional testing and shielding is typically required. Section 5.4 provides more detailed explanation of radiation effects on lasers.

Thermal control and management: Current laser transmitters are only about 10 percent efficient. Therefore, about 90 percent of the input electrical power is converted to heat and optical losses. This heat has to be dissipated without affecting the optical alignment integrity of the terminal. The laser subsystem may be directly coupled to the terminal or may be located remotely, with its output beam piped in via optical fibers. The former is more efficient but can impart heat into the rest of the terminal, the latter is less efficient due to fiber-coupling losses, but affords much greater flexibility for thermal

management of the pump diodes, which are the primary source of heat generation in the optical terminal.

5.2.3 Candidate Laser Transmitter Sources

The above requirements on suitability for spacecraft use and the need for modulation to high data-rates, limit the pool of practical options, among those available now, to a class of lasers known as diode-pumped solid-state lasers. Included in this class are: (1) pulsed bulk crystal solid-state lasers, in the 914 nm to 1080 nm (e.g., neodymium: yttrium aluminum garnet [Nd:YAG] at 1064 nm), 965 nm to 1550 nm master oscillator power amplifiers (MOPAs); and (2) continuous-wave (CW) and pulsed fiber-amplifiers (such as erbium-doped fiber amplifiers (EDFAs) at 1550 nm and ytterbium-doped fiber amplifiers at 1064 nm). Lasers with wavelengths in the 2000-nm to 5000-nm region are also of interest due to their superior atmospheric transmission and reduced optical surface accuracy requirements for the transmitter telescope.

From an efficiency and technical maturity point-of-view, rare-earth-doped solid-state lasers and fiber amplifiers are the leading candidates for deep-space laser communications since they provide a combination of both high peak power and moderate average power. These lasers operate in the 1000-nm to over 2000-nm wavelength range. An alternate method is to use the second harmonic of this wavelength generated through nonlinear conversion. Table 5-7 compares the merits of five viable laser wavelengths. Assumptions are: a Mars mission of the range of 2.5 AU, 30-cm flight aperture diameter; a 5-m ground aperture diameter, hazy sky with 5-km visibility, cirrus clouds, and 70-deg zenith angle, PPM order of 128 and bit error rate of 1E–6. With today's detector technology and wavelength conversion efficiency, the 1064-nm laser appears to be the leading candidate.

The key requirements for a laser transmitter on a deep-space spacecraft include: (1) high electrical-to-optical efficiency and reasonable power consumption; (2) output power (average and peak); (3) excellent beam quality; (4) variable repetition rate; (5) reliability; and (6) low weight and small size.

The link parameters with significant wavelength dependence include laser output power; laser efficiency; atmospheric propagation; detector quantum efficiency and availability; background light (noise) at the receiver, and transmit/receive isolation.

Among the many types of lasers currently known, primarily diode-based MOPAs and diode-pumped solid-state (DPSS) lasers come close to satisfying all of the above requirements simultaneously. High power semiconductor lasers and bulk-crystal solid-state lasers, or doped fiber lasers and amplifiers that are amplitude modulated are useful in multi-gigabit links for near-Earth laser communications. However, these sources lack significant peak power and are more suitable for lower-order PPMs.

Table 5-7. Merits of five deep-space communication link wavelengths.

Wavelength (nm)	532	775	1064	1550	3100
Spacecraft's laser transmitter power (W)	14	10	20	20	14
Detector's detection efficiency (%)	90	50	35	35	30
Detector noise (megacount/s)	0.3	0.1	0.1	10	0.3
Transmit and receive optics losses. both ends (dB)	–7	–7	–7	–7	–7
Background light (W/cm^2 sr m) at 70-deg zenith angle (ZA)	0.186	0.144	0.0842	0.0325	7.00E-04
Atmosphere transmission at 70-deg ZA	0.3	0.47	0.55	0.6	0.3
Pointing loss (dB)	–2	–2	–2	–2	–2
Required peak power (W)	179	128	256	256	179
Channel capacity (megabits/second)	10	10	10	10	10
Code rate	0.05	0.05	0.05	0.05	0.05
Atmospheric coherence length (r_0, cm)	4.36	6.84	10	15.7	36.1
Seeing (μrad)	12.2	11.3	10.6	9.87	8.58
Link margin (dB)	–2.7	0.5	2.4	1.8	–0.43

5.2.3.1 Pulsed Laser Transmitters. Depending on the required average power, pulse width and data rate, either a single oscillator or an oscillator/amplifier will be required.

Single oscillators are typically limited in power to several watts. At higher power levels, control of the pulse width, laser power, pulse jitter, and pulse-to-pulse repeatability become more difficult. A MOPA scheme affords the amplification of a well-behaved oscillator through a suitable and efficient amplification medium. The oscillator and amplifier can then be individually tailored for high speed and high power, respectively.

The type of pulsed oscillator used will depend on the required data-rate. Examples are:

1) Amplitude modulated diode lasers provide modulation rates up to several GHz. However, peak power, pulse-to-pulse stability, and modulation extinction ratio of these sources are limited. For these lasers, peak power is typically just a few time above the average output power. Therefore, additional amplification stages would be required utilizing either a fiber, waveguide or bulk solid-state design [21]. These are detailed below.

2) Q-switched DPSS lasers can be pulsed to sub-megahertz levels, with the maximum data rate being limited by the available acousto-optic or electro-optic Q-switchers. These can provide pulses of one to a few ns with high modulation extinction ratio and high peak power. In a Q-switched laser, the population inversion is built-up by preventing feedback, and then it is switched to laser emission for a short period of time by improving the finesse or "Q" of the cavity to extract highly energetic pulses. The energy is stored in the gain material. This results in high peak power (on the order of 10 kW or more) and high average output power (watts) of the laser [22,23,24].
3) Cavity dumped DPSS lasers offer a high extinction ratio, high peak power, a short (a few nanoseconds) pulse width and pulse rates up to about 10 MHz. In cavity dumping, the intra-cavity field within the resonator stores the energy that is built-up by preventing laser emission; this field is then switched and dumped out of the cavity, by, for example, using polarization effects along with an electro-optical material. These lasers also provide moderate peak power and moderate average power [25]. Figure 5-18 shows a schematic of a version of a cavity-dumped laser.
4) Mode-locked lasers are capable of high peak power and modulation rates exceeding 1 GHz. However, the very short pulses involved add to data demodulation and encoding complexities [26].

For a uniform Gaussian or square wave pulse, the peak power (P_p) is related to energy per pulse (E) and pulse-width (pw) as $P_p = E / pw$. Also, the laser's average output power is related to energy per pulse and pulse-repetition frequency (PRF) as $P_a = E_p \times$ PRF. As shown in Fig. 5-19, maintaining a nominal value for average output power and pulse-width will result in steep reduction of the peak-power per pulse as the lasers' PRF increases.

Different amplifier media are available to boost the average power of the pulsed oscillator, up to 30 dB, while maintaining the output beam quality and the pulse-width, pulse-to-pulse jitter, and modulation extinction ratio of the oscillator. With all amplifiers, care has to be taken to avoid spontaneous lasing,

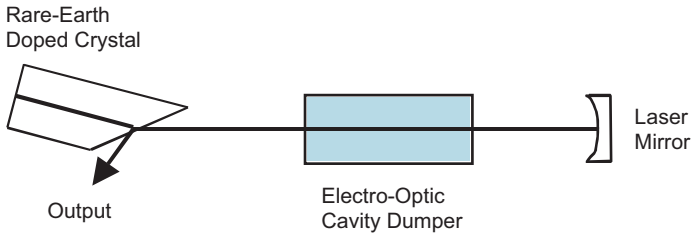


Fig. 5-18. Schematic of a high-power cavity-dumped oscillator.

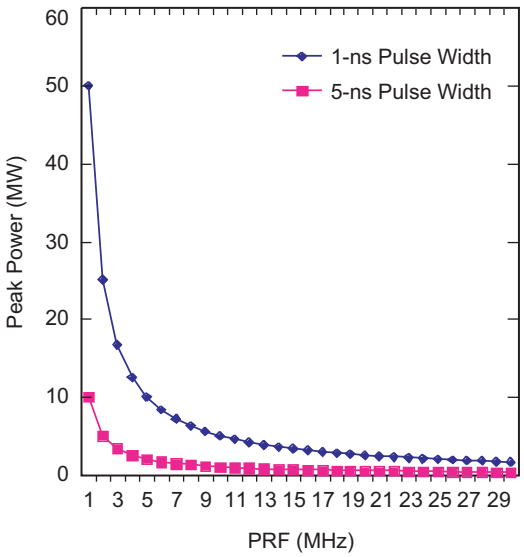


Fig. 5-19. Peak power as a function of pulse repetition frequency (PRF) for the case of fixed average power and pulse width (1 ns and 5 ns).

bulk damage, and parasitic nonlinear losses (such as Brillion scattering in fibers) [27]. Also, isolators may be required between the oscillator and the amplifier to avoid amplified spontaneous emission (ASE) and feedback to the oscillator. The amplification may be implemented in multiple stages. Examples of amplifier-based laser transmitters include: fiber waveguide, bulk crystal, and bulk-crystal waveguide. These are addressed separately below.

5.2.3.2 Fiber-Waveguide Amplifiers. Fiber-waveguide amplifiers use rare-earth-doped optical fibers for the amplification medium with optical pumping (excitation) provided by compact and efficient semiconductor diode lasers. Depending on the amplification architecture, fiber amplifiers (or lasers) can generate broadband or narrow linewidth output beams. These types of amplifiers, shown schematically in Fig. 5-20, provide a long amplification path (obviating the need for multi-passing), but have a lower threshold for damage than the bulk crystal counterparts. Single-mode fiber systems are limited in their peak power capability, but broadening of the fiber core while maintaining single-mode propagation is possible through a large mode area fiber design. Tens of kilowatts of peak power can then be propagated without being limited by fiber nonlinearities. Advantages of the fiber waveguide are ease of use, efficient coupling to fibers, and a relatively low noise floor. The challenges are nonlinear effects, excited state absorption, and cooperative upconversion. The

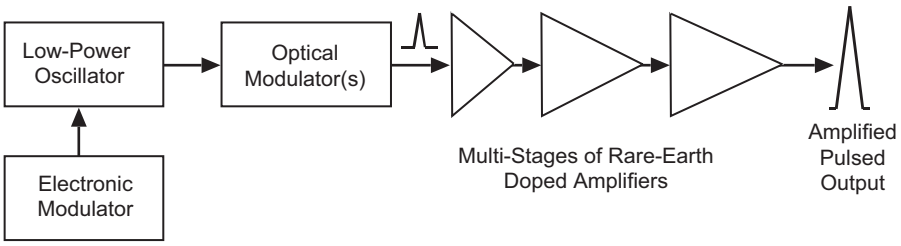


Fig. 5-20. Schematic representation of a pulsed fiber oscillator amplifier. As shown in the figure, the oscillator may be amplified directly or externally. The number of amplifying stages will depend on the output power characteristic requirements.

nonlinear effects include: stimulated Brillion scattering (SBS), stimulated Raman scattering, self-phase modulation, cross-phase modulation, and four-wave mixing. SBS is interaction of photons with acoustic phonons, resulting in lower SNR due to both signal reduction and introduction of additional noise [27].

In a fiber amplifier, a well-behaved oscillator beam is injected into a rare-earth-doped fiber amplifier that operates at the same wavelength as the oscillator. Common dopants for the fiber are: Nd, Yb, Er, or Yb/Er ions (co-doped) [28,29,30]. To generate amplification, the fiber oscillator is pumped with diode laser(s) operating at the absorption wavelength of the dopant. To generate high peak powers at the output of the amplifier, the oscillator is operated in the pulsed mode. The signal modulation (for communication) is imposed on the master oscillator by modulating and then amplifying the oscillator in one or more stages) [31,32,33].

A master-oscillator-power fiber-amplifier (MOPFA) source offers the following advantages:

1) Minimal requirement on structural integrity compared with the stringent resonator mirror alignment required for bulk crystal solid-state lasers.
2) Significant potential for higher overall efficiency relative to DPSS lasers, due to near 100-percent absorption of the pump beam in the medium (fiber), smaller quantum defect, and higher extraction efficiency.
3) Wider temperature tolerance for the pump laser wavelength shift (about 20 nm compared with about 1 nm for bulk crystal lasers).
4) Lower demand on tight control of the pump diode-laser temperature, resulting in significantly lower power consumption and higher overall efficiency.
5) Modulation extinction ratios on the order of 40 dB or more are feasible owing to a high degree of control on the master oscillator's performance. The fiber amplifiers do not alter input pulse characteristics from the master oscillator (MO) in a major way.

6) Shorter pulse-width and potentially higher pulse repetition frequency (PRF) due to simple control of the MO.
7) It is feasible to obtain narrow (sub-angstrom, <0.02 μm) linewidth with the use of Bragg gratings.

The peak power desired at a given data-rate drives the MOPA design due to possible nonlinear effects or damage to the fiber at higher peak powers.

Achievable gain and noise figure (NF) for fiber amplifier operation around 1000 nm is shown in Fig. 5-21 [34]. Selection of the transmitter wavelength depends on the specific mission requirements, efficiency, technological maturity, and compactness of the source at the time of selection. There is currently no general preference of one wavelength over the other since each offers certain advantages and suffers from some disadvantages with the benefits often outweighing the drawbacks.

5.2.3.3 Bulk-Crystal Amplifiers. Bulk-solid-state amplifiers are not peak power limited and also provide improved modulation extinction ratio. An example is an Nd:YAG amplifier that is optically pumped. Until the saturation limit is reached, generally the higher the number of passes through the amplifier, the higher the gain. A few different versions of these amplifiers have been reported [35].

Several different diode-pumped lasers have shown relatively high efficiency (~10 percent). The laser active elements (crystals) include: Nd:YAG, neodymium: yttrium vanadate (Nd:YVO_4), neodymium: yttrium lanthanum fluoride (Nd:YLF), and ytterbium:glass (Yb:glass), thulium: yttrium aluminum garnet (Tm:YAG), Tm, holmium yttrium lanthanum fluoride (Ho:YLF). In these crystals, the Nd, Yb, or Tm ions are trapped in a host crystal. The ions absorb light at a short wavelength and emit at a longer wavelength.

The fundamental wavelength of the Nd lasers is at about 1000 nm (for example, 1064 nm for Nd:YAG and Nd:VO_4). With a small nonlinear frequency-doubling crystal, it is possible to generate the second harmonic wavelength of these lasers at about 532 nm. Harmonic conversion efficiency is typically 30 to 50 percent, depending on the laser's peak power.

Tm and Ho ions emit laser light near 2000 nm. In terms of atmospheric propagation and background noise, this and longer wavelength ranges are superior to the 1000-nm wavelength range. However, these lasers are less efficient than 1000-nm lasers and operate best at low temperatures due to the three-level structure of the ions. Also, detectors at 2000 nm are significantly noisier than those for 1000 nm. Moreover, because of the very long upper-state lifetime of the ion (10 ms for Ho in YLF compared with 0.24 ms for Nd in YAG), the highest energy per pulse (or peak power) occurs at low repetition rates (near 200 Hz), making them unsuitable for significantly higher data rates. Lasers in the 3800- to 4200-nm range provide some of the best combinations of

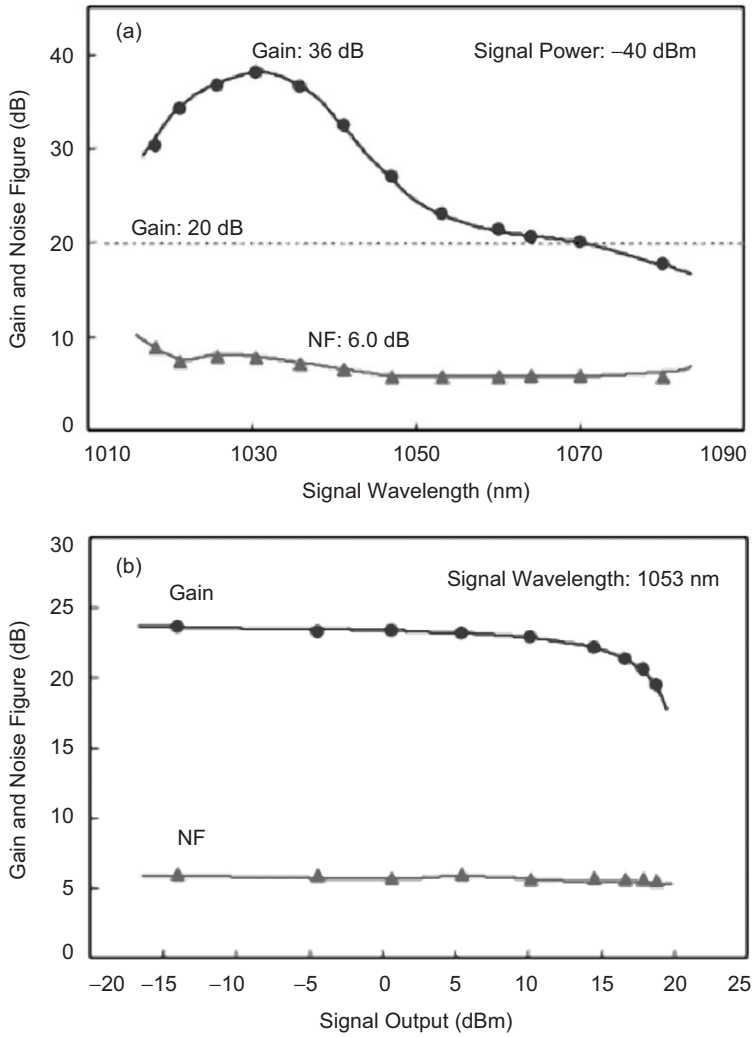


Fig. 5-21. Gain and noise figure (NF) characteristics of a ytterbium-doped fiber amplifier (YDFA) from 2003 catalog of Mitsubishi Cable America, Inc.: (a) signal wavelength and (b) signal output.

atmospheric transmission and lowest background light characteristics. However, efficient, compact, high power lasers are not readily available yet. Also, as discussed in Chapter 2, the space losses term increases quadratically with wavelength.

Energy per pulse (E_{pulse}) is given by equation: $E_{pulse} = P_{cw}T_s(1-e^{-1/T_s f})$, where P_{cw} is the continuous-wave power, T_s is the upper-

state lifetime of the active laser material, and f is the pulse-repetition frequency. Figure 5-22 shows a plot of the E_{pulse} as a function of PRF. For both Q-switched lasers in this example, the initial state inversion density saturates as the pumping time $(1/f)$ begins to be long compared to the respective upper-state lifetimes [36]. Assuming a CW power of 1 W, the saturation value for the pulse energy is (T_s X 1W); hence, for low pulse rates, the Ho:YLF laser pulses are 50 times larger than those for Nd:YAG. For high pulse rates, the pulse energies become asymptotically equal. For high pulse rates $f >> 1/T_s$, $E_{pulse} \simeq P_{cw}/f$, the laser's CW power is effectively collected over the pump time $1/f$ and emitted as a short pulse. For low pulse rates, the pulse energy saturates as

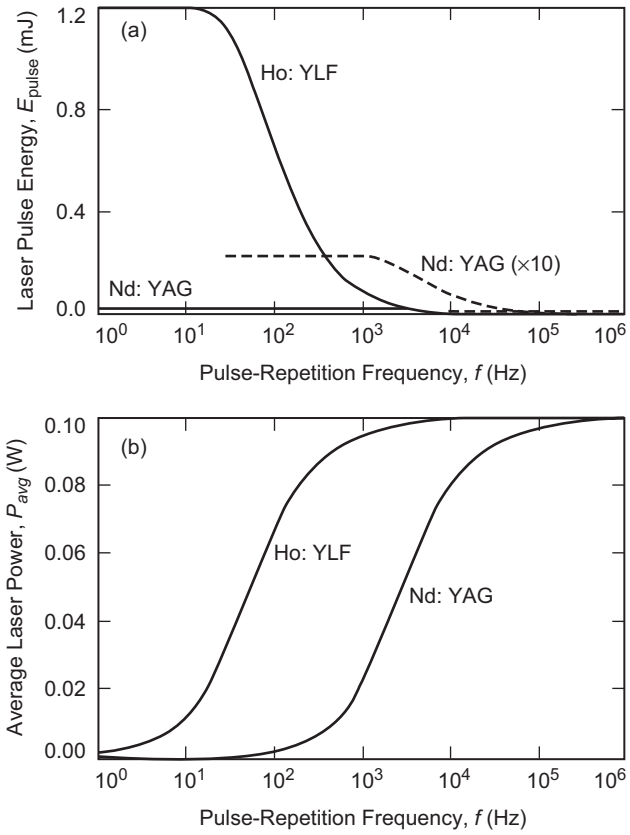


Fig. 5-22. Plots of the laser energy per pulse and average laser power as a function of pulse repetition frequency for Nd: YAG and Ho: YLF lasers: (a) output energy per pulse versus pulse-repetition frequency assuming continuous-wave output power of 1 W and (b) average output power versus pulse repetition frequency.

the pumping time becomes long compared to the "storage time" T_s. In the latter case, $E_{pulse} \simeq P_{cw} T_s$.

5.2.3.3.1 Bulk Crystal Waveguide Amplifiers. Multi-mode dielectric waveguides may be fabricated by bonding dissimilar crystals to an active amplifying medium and utilizing the self-imaging effect in the crystal. Circular, rectangular, tapered, and hollow waveguide geometries are possible with this scheme. This concept incorporates the waveguiding advantages of fiber-based designs with those of a bulk-solid-state crystal gain medium [37].

5.2.3.3.2 Pulsed-Diode Lasers. Pulsed-diode lasers or low (average power) pulsed DPSS laser (oscillators), amplified in fiber amplifiers, can generate the required peak power (kilowatt level), and moderate average power (watt level). Common fiber-amplifier media include: Nd:glass, Yb:glass, Er:glass, or a multiple doping of these ions in glass. Dual cladding, where a larger second clad region surrounds the core, allows the efficient coupling of the pump light from broad-area diode lasers though multi-pass absorption. The corresponding output wavelength varies in the range of 1030 to 1550 nm. Two factors may limit the usefulness of these types of MOPAs: (1) nonlinear effects in the fiber can cause a broadening of the spectral width of the laser (to about 2 nm); (2) the (glass) fibers darken when exposed to greater than about 0.1 Mrad of radiation. Glass is also inhomogeneously broadened, requiring broadband oscillator sources to get any efficiency.

5.2.3.4 Semiconductor Optical Amplifiers. Semiconductor optical amplifiers (SOAs) can provide linear amplification below their oscillating threshold. In an SOA, the population inversion of the atoms is reached by electrical pumping of the active semiconductor region and employing anti-reflection coating on the typical semiconductor laser mirrors. Benefits of SOAs are broadband (over 40 nm) amplification, high (~30 dB) net gain, compactness, and ease of integration with opto-electronic amplifiers. Their drawbacks are higher noise floor (compared with fiber amplifiers), highly nonlinear gain, and a low damage threshold for high-power pulses [38]. Various schemes have been developed to improve SOA performance, such as tapered gain regions and distributed feedback designs.

Depending on the power levels required, all of the above MOPA schemes may require a pre-amplifier prior to the primary amplification in order to obtain adequate saturated peak power levels. A pre-amplified DPSS laser is excited (optically pumped) with semiconductor diode laser(s). The diode laser itself is electrically more efficient than diode-pumped lasers and can be directly current modulated, but lacks significant peak power. Amplified diode lasers with moderate continuous-wave (CW) power and low peak output powers (in the watt class) are now available. Even the highest available semiconductor laser

powers do not support a link with adequate margin at deep space ranges greater than a fraction of an AU. High-power diodes also typically suffer from poor beam quality, and direct current modulation of the diode driver at high modulation rates results in poor efficiency. This is again due to the fact that for these lasers, the maximum achievable peak power is only a few times higher than its average power.

5.2.4 Lasers for Coherent Communications

Coherent communications require a frequency stable laser for the transmitter and at the receiver, and compatibility with the modulation schemes unique to optical communications. Critical aspects of coherent communications include frequency stability and laser linewidth. Diode-pumped solid-state (DPSS) lasers and amplified-frequency stable lasers can provide excellent frequency stabilities. DPSS lasers combine excellent spatial and longitudinal mode quality with high power and inherent redundancy. Mode stability requirements make these lasers more complex than those for direct-detection. Similar to the lasers discussed earlier, lasers for coherent detection are currently fairly inefficient, and lasers for coherent communication typically require an external modulator capable of handling high powers. Linear, ring, discrete-element, and monolithic resonator, as well as oscillator amplifier configurations of single frequency lasers have significantly matured. They each offer certain advantages and disadvantages, and their selection will be driven by the mission requirements. Methods to obtain single-mode operation include: intracavity etalon, ring resonators, mode-twisting techniques, and use of short cavities. Injection seeding is another scheme for enforcing single-mode operation in a high power laser. A variety of laser transmitter sources have been developed for coherent free-space communications [39,40,41].

5.2.5 Laser Modulators

Generally, two classes of amplitude modulators exist for laser transmitters. One class includes intra-cavity pulsers, such as Q-switchers or cavity-dumpers, that are used to generate pulses from solid-state lasers. The other class includes extra-cavity modulators, such as $LiNbO_3$ modulators, used in conjunction with the output of a semiconductor laser in a MOPA system. The advantage of the latter is that much higher repetition rates can be achieved. The modulators may operate in acousto-optic, electro-optic, or magneto-optic mode. Due to the excessive radio-frequency (RF) power needed for acousto-optic modulators, electro-optic modulators can be made more efficient than acousto-optic modulators by nearly an order of magnitude. However, driving most modulators simultaneously at high voltages (on the order of several hundred volts) and high modulation rates (above half a gigahertz) becomes challenging.

The pulse-width modulation method is a technique to optimize and control the output energy of the laser below the damage threshold of the crystal or fiber active medium. In this scheme, the pulse width is adjusted to maintain a desired output energy from the laser system. An output energy monitor operates as the feedback senor in a closed loop to pulse the pump diode lasers as the laser PRF and environmental conditions vary [42].

For coherent communications, modulation options include: an externally modulated laser; phase modulation; polarization modulation; and amplitude modulation. Every modulation scheme can be combined with homodyne or heterodyne detection. Phase-sensitive detection schemes are typically the most sensitive method, followed by polarization, and amplitude modulation. An electro-optic modulator may be integrated with the laser amplifier, or a bulk modulator may be utilized with the high-power laser. There is no clear advantage in terms of overall efficiency. Phase modulators are more efficient than polarization modulators, which are in turn more efficient than amplitude modulators.

5.2.6 Efficiency

Due to strict budgets for electrical power and mass on deep space missions, it is extremely important to maximize the laser transmitter efficiency. Higher overall efficiency translates directly into lowered mass for the flight terminal and reduced launch costs. The efficiency of a solid-state laser is determined by three key parameters: the pump semiconductor-laser diode electrical-to-optical conversion efficiency, η_D; the coupling or transfer efficiency of the pump light into the active medium, η_T; and the optical-to-optical conversion efficiency of the active-gain media, η_O These parameters can be further broken down to give the overall efficiency as [43–46]: $\eta = \eta_D \eta_T \eta_O = \eta_D \eta_T \eta_{abs} \eta_S \eta_Q \eta_B \eta_{ST} \eta_{ASE} \eta_E \eta_R$, where η_T is the optical efficiency of coupling the pump light, η_{abs} is the absorption efficiency of the gain media, η_S is the stokes efficiency or ratio of the output pump photon energy to input photon energy, η_Q is the quantum efficiency or fraction of pump photons reaching the upper laser level, η_B is the spatial beam overlap of the resonator modes with the upper state inversion, η_{ST} is the storage or depletion efficiency, η_{ASE} represents the loss due to amplified spontaneous emission which is the reciprocal of the depopulation rate of the upper laser level, η_E is the fraction of absorbed energy extracted and η_R is the resonator loss including reflective and scattering losses. Sometimes the efficiencies are grouped as the transfer efficiency, η_T, upper-state lifetime efficiency $\eta_U = \eta_S \eta_Q$ and extraction efficiency under Q-switched operation, $\eta_{eq} = \eta_{ST} \eta_{ASE} \eta_E$. Moreover, the wall-plug efficiency of a flight laser transmitter takes all the possible power requirements into account. These

include the thermal control of the laser components and electronics, auxiliary control electronics for monitor photodiodes, thermisters etc, power consumption for the Q-switcher or cavity-dumped and the DC-to-DC power conversion efficiency of all the drive electronics.

Table 5-8 summarizes wall-plug efficiency of some pulsed laser transmitters, for an example, with a few Watt of average output power. The optimized design assumes 50-percent (66 percent for fiber based) diode pump laser efficiency, where the theoretical analysis assumes 75-percent efficiency. The heat sink temperature determines the actual efficiency range.

5.2.7 Laser Timing Jitter Control

Several sources can contribute to the timing delay (or pulse jitter) of high repetition rate Q-switched and cavity-dumped diode-pumped solid-state lasers. These sources, which introduce either a fixed delay or in some other way effect the pulse build-up time in the laser cavity, include: electronics and switch-related jitter; longitudinal-mode build-up time jitter; and stored energy and build-up time jitter [47].

The first two sources are expected to contribute minimally or can be minimized; whereas, the last source makes up the bulk of the contributions to pulse jitter. The effect of each of the above contributors to jitter and the techniques proposed to alleviate them are described below. The pulse width of a pulsed laser optimized for high-efficiency energy extraction is typically about 1/3 of the build-up time, so maintaining the timing jitter to well under 10 percent of a pulse width requires that the laser energy and gain be controlled to within about 3 percent.

5.2.7.1 Jitter Control Options. Several schemes have been developed and tested for jitter control of diode-pumped solid-state lasers. These approaches, when implemented, should reduce the laser output pulse jitter to less than about 20 percent of the laser pulse width. Among them are:

Table 5-8. Comparison of overall efficiencies of solid-state lasers.

Pulsed Laser	Current Best Results (percent)	Estimated Optimal Efficiency (percent)
Nd:YAG or Nd:YVO4… 1064 nm		
With thermo-electric cooler	~10	22
With active loop heat pipe	~15	25
Yb-doped fiber amplifier, 1060 nm	~20	>30
Er-doped fiber amplifier, 1550 nm	~15	>30

5.2.7.1.1 Electronics and Switch-Related Jitter. Jitter from electrical and switching delays can be mitigated in a number of ways. Electrical delays are expected to be nearly constant from pulse to pulse and should contribute negligibly to timing jitter, or could be subtracted electronically. Jitter caused by electrical delays is constant and amounts to only about 10 percent of the pulse-width. Therefore, it may be accounted for and corrected.

5.2.7.1.2 Longitudinal-Mode Build-up Time Jitter. Random build-up of the laser power from noise will result in randomness of the modal composition of the laser's longitudinal mode, and that could result in timing jitter. Empirically, this (rms) timing jitter is about 10 percent of the pulse width (measured at full width at half maximum—FWHM) [48]. Injection seeding of the laser with an external source is a common method to alleviate this source of jitter.

5.2.7.1.3 Stored Energy and Build-up Time Jitter. The build-up time is affected by the repetition rate. Non-uniformity in the pumping time (such as, inter-pulse timing variations due to pulse repetition rate changes) and variations of the pump power cause variations in stored laser energy. Variations of the stored energy effect pulse build-up time, which in turn affect the output pulse timing and shows up as pulse jitter.

Possible techniques to mitigate the jitter due to stored energy include:

1) Time-variable (pulse) pumping. When the laser is continuously pumped for a non-uniform duration prior to each pulse, and the pulse is extracted with variable pulse-to-pulse timing (due to PPM), the stored energy is bound to vary. However, continuous pumping is not required, and the pump diodes may be pulsed (e.g., turned on and off in a controlled manner). If the pump diode is on for a period of time corresponding to the dead time (the inter-pulse period at maximum pulse repetition frequency) of a PPM pulse, the laser gain will always reach the same value. Subsequently, the pump diode power is lowered to a value that is just enough to sustain the gain. By this method, pulse jitter is reduced significantly since each pulse will have the same gain and build-up time.
2) Negative amplitude feedback with a constant offset provides stability and minimizes the timing jitter. In this case, varying the pulse-width of the output energy pulse while maintaining a current amplitude set point controls the laser output energy. By using a pulse-width-modulation-based control system, the current applied to the pump diodes is regulated at an efficient set point below the damage level, and the pulse-width is adjusted to maintain a desired output optical energy of the laser system.

5.2.7.1.4 Injection Seeding. One process that may eliminate electronics related pulse jitter is injection seeding. Synchronizing the RF oscillator with the pulse

trigger may also eliminate jitter caused by switching delays. In cavity dumping, for example, a mode-locked cavity-dumped laser minimizes timing jitter by taking advantage of the precision timing of its mode-locking effect.

5.2.8 Redundancy

Critical active components of the laser include the pump diode lasers, the driver electronics, and the control electronics. Critical passive components are the active laser element, the optical surfaces, and coatings on the optics. Aging, facet damage, and radiation effects are some of the potential failure mechanisms. Operation below the maximum rates safe current limits and operation at low temperature enhance the diode laser lifetime. Of all of the above-mentioned components, the diode laser lifetime is typically of highest concern. Multiple redundancy of the pump diodes, or block redundancy of the entire laser, is often a prudent approach to minimizing risk. The control loop and driver electronics are low-power, low-voltage devices and can be designed for an adequate level of redundancy. For passive elements, block redundancy may be applied.

5.2.9 Thermal Management

The thermal management of the laser transmitter is driven mainly by the need for thermal control of the pump lasers. The FWHM absorption bandwidth of most active bulk crystal laser mediums is narrow (approximately 1.5 nm for Nd:YAG and 9 nm for Nd: YVO_4). The pump laser diode wavelength varies with temperature on the approximate order of 0.3 nm/deg C. To pump rare-earth doped-crystals at the peak of absorption, therefore, the diode laser's temperature has to be controlled to within ±0.3 deg C and ±3 deg C for Nd;YAG and Nd:YVO_4 respectively.

For doped-fiber-based transmitters, the temperature-control requirement is much relaxed to approximately ±20 deg C, depending on the pump absorption band utilized. Yb doped glass lasers can be either pumped at around 980 nm or at the broader but weaker absorption peak at 915 nm. Active temperature control of the diode pump laser will thus be required. Options are thermo-electric coolers (TECs) and active-loop or passive-loop heat-pipes. TECs are typically inefficient. Conversely, heat pipes and radiators are more efficient but introduce additional mass.

Remote pumping of the laser (or amplifier) can significantly reduce the thermal management difficulties. In this case, the pump lasers are fiber-coupled and mounted in an area (e.g., a radiator) where their generated heat may be removed conveniently. The only drawback is the fiber-coupling loss that will be encountered. Remote location of the pump lasers (source of heat) will also simplify the optomechanical design for the laser-communication terminal.

A detailed finite element analysis (FEM) for the flight laser transmitter will help assess weaknesses in the optomechanical structure designed for the laser that might adversely affect the performance of the laser.

5.3 Deep-Space Acquisition, Tracking, and Pointing

Gerardo G. Ortiz and Shinhak Lee

5.3.1 Unique Challenges of Deep-Space Optical Beam Pointing

For optical communication links, mispointing of the transmit beam results in a variation of the downlink signal power. Because of the diffraction-limited transmit beamwidths used, the received signal power is extremely sensitive to the transmitter pointing error. A large transmitter off-point can lead to intolerable signal fades on the ground and significantly degraded system performance. This makes the spatial acquisition, tracking, and pointing (ATP) function critical to laser communication systems. This problem is compounded by the fact that the platform jitter present in the spacecraft due to dead-band cycle and random platform jitter are generally much larger than the transmit beamwidth. As a result, an ATP control subsystem is required to reduce the signal loss due to the pointing error. Such a subsystem must be capable of first acquiring a reference beacon source for absolute attitude determination in the presence of large attitude uncertainty. Then, it must accurately point the communication transmit beam to the Earth receiver in the presence of spacecraft orbital motion and microvibrations. To keep mispoint losses low (<2 dB), the required total pointing accuracy of the transmit signal is generally less than 40 percent of the diffraction-limited beamwidth, generally on the order of a microradian.

The two key issues for the ATP system are determining pointing knowledge of the Earth receiving station relative to the spacecraft (S/C), and then aiming the downlink beam to the receiving station. The pointing knowledge can be derived from acquiring and tracking on either an uplink ground based laser beacon or on passive celestial sources, such as the Earth, the Moon or stars. Due to its passive, non-cooperative nature, tracking on celestial sources has come to be known as beaconless tracking. Uplink beacon tracking can support pointing at short range and during opposition when the Earth image alone does not provide sufficient signal power for tracking. Uplink beacon tracking is an attractive alternative, although ground-based beacon uplink cannot provide the power required for high-rate pointing without needing additional inertial-sensors. Furthermore, at low Sun-Earth-spacecraft angles when the Earth image is brightest, the Earth background can cause a shift in the measured beacon centroid and interfere with beacon tracking. Finally, by requiring a clear path for uplink in addition to clear downlink path, a beacon-based system has a lower overall link availability.

For Mars- and Jupiter-range missions, the current baseline pointing and tracking approach is to perform Earth image tracking with occasional calibration using a laser beacon, Earth-Moon images, or Earth-star images. At high phase angles when the Earth image does not provide sufficient brightness for high rate tracking, inertial sensors (accelerometers) measurements are required to propagate pointing knowledge at a higher rate in between celestial reference updates. Control of a steering mirror is maintained by closed-loop control of a portion of the downlink reflected to a second detector.

Earth image tracking is desirable because of its high brightness (over most of the orbital period) and angular proximity of Earth intensity centroid to the receiver location. Among the challenges of Earth image tracking are the unknown Earth albedo variation due to cloud coverage and the solar stray light. The baseline design answers the albedo variation problem by performing periodic imaging of the Earth with other celestial references such as the Moon or nearby stars. These sources have well defined intensity patterns that allow accurate measurements of their position, but they require long integration times. The position of the Earth can then be calibrated using the measured celestial references' position and the known Earth ephemeris to determine the correction offset.

Since the Sun–spacecraft–Earth (SPE) angle becomes small during the mission, stray light control is important. At the low SPE angle, the subsystem is intended to operate 2 deg from the Sun; consequently, both optical surface quality and cleanliness need to be controlled to ensure low scattering of incident sunlight. Studies have shown that the required surface quality for the mirror can be achieved [49]. In addition, the optical design incorporates both a field stop in the telescope and a Lyot stop in the post secondary optics to control out of field scattered sunlight.

The innovation and uniqueness of JPL's ATP System has been in the development and integration of advanced components and subsystems, which improve random and system noise and dynamic range. Secondly, system-level improvements have been made in ATP algorithms and architectures to achieve ATP pointing accuracy to the sub-microradian level. Lastly, a unified ATP architecture has been developed that enables precise pointing throughout the Solar System.

5.3.1.1 State-of-the-Art ATP Performance. The objective of JPL's ATP work has been to develop and validate a complete set of acquisition, tracking and pointing (ATP) technologies with <1 μrad pointing accuracy for laser communications throughout the solar system. A comparison with the state-of-the-art systems is presented in Table 5-9. Ultra low pointing accuracies have been achieved for large systems such as the Hubble Telescope with a pointing accuracy of 35 nanoradians. Accurate relative pointing systems, such as Starlight, have also demonstrated low relative pointing control. But for the

Table 5-9. Comparison of some leading-edge precision optical pointing systems.

Organization	ESA (SILEX)	NASDA (OICETS)	Boeing (LADAR)	NASA (Hubble S/C)	NASA (Starlight S/C)
Accuracy (1σ)	0.22 μrad	0.86 μrad	10 μrad	35 nrad	0.5 μrad
Range	LEO-GEO	LEO-GEO	500 km (Ballistic target from airplane)	Observation	Up to 1 km separation
Applicability to deep space	Range-limited	Range-limited	Range-limited	Large, Expensive	Relative pointing only

Legend: ESA = European Space Agency
LADAR = laser radar
LEO = low Earth orbit
NASDA = National Space Development Agency (Japan)
OICETS = Optical Inter-Orbit Communications Engineering Test Satellite
SILEX = Semiconductor Intersatellite Link Experiment

optical communications system, the technologies need to demonstrate preciseness in absolute pointing accuracy.

5.3.2 Link Overview and System Requirements

The main function of the ATP system is to accurately and precisely point the downlink communication beam towards the receiver antenna. This is done by acquiring and tracking an external reference source and using that information to point the downlink to the receiver target. The performance of the ATP system is specified by its pointing accuracy.

5.3.2.1 Pointing Requirement. The communications link equation determines the relation between the mean received signal power and the transmitted power. Any transmitted power that is allocated for pointing of the downlink beam is power not used for communications. This means the data rate and pointing loss are inversely proportional, and any gain in pointing accuracy is a direct benefit to the data rate (i.e., a 3-dB decrease in pointing loss equals a 3-dB gain in data rate). Therefore, it is highly desirable to keep the loss allocated for imperfect pointing of the narrow laser beam very low.

Any mispointing of the laser beam that causes the far-field irradiance profile to be located off-axis from the receiver will result in a signal loss. This is called a pointing loss. Furthermore, keeping the narrow-angular-width laser beam pointed in the presence of spacecraft attitude and vibration disturbances becomes a formidable challenge. Therefore, in determining a link budget, some losses are allocated to pointing.

Another factor that impacts the quality of the link is the probability of burst errors. Due to the random nature of the tracking sensor noise and the control

system error, the achieved pointing accuracy has a statistical behavior and displays a probability of fade errors. The overall link designer must consider the level of fades that it can tolerate in order to provide the desired data rate and data volume. Therefore the data rate depends not only on the pointing loss allocation but also on the pointing-induced fade (PIF) probability allocation.

For high data rate links, the gain of the transmitter has to be high. This in turn, pushes the aperture to be large, typically in the range of 30 to 50 cm. Also, the lasers currently being considered and developed for deep-space communications have a wavelength in the range of 500 to 2000 nm. Therefore, the diffraction-limited beam width (at FWHM) is in the range of 1 to 7 μrad. Projects currently developing practical deep-space applications are considering implementation of lasers with transmitter 1064-nm wavelengths and 30-cm apertures [50,51]. Therefore, the expected diffraction-limited transmit communication laser FWHM beamwidth is 3.65 μrad.

The pointing requirement (also known as mispoint angle) depends on the allocated mispoint loss and the required probability induced fade (probability of burst errors). The pointing accuracy required of these systems depends on the power link budget allocation and the allowed PIF probability. The recent pointing designs for a deep-space optical communications link have set this total pointing loss allocation at 2 dB, which translates to a total mispoint angle of $0.42 \times \lambda / D$, see Fig. 5 23. For a 1064-nm system with telescope aperture of D = 30 cm, the total mispoint allocation is 1.5 μrad.

As an example of possible high data rates links and their pointing requirements, a 40-Mbps rate has been shown to be feasible for a Mars-to-Earth

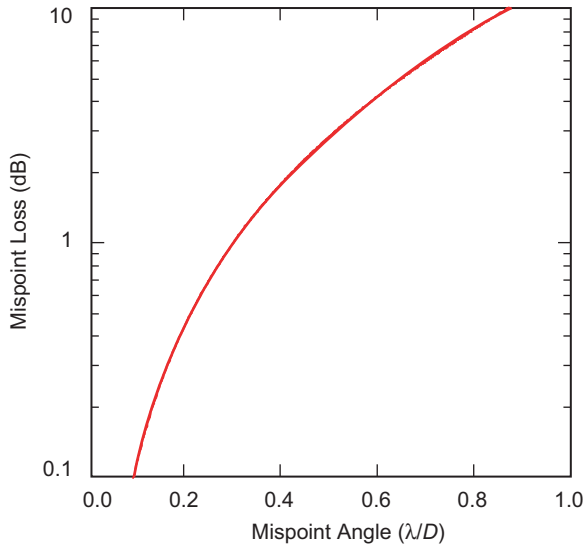


Fig. 5-23. Mispoint loss (dB) vs. mispoint angle (λ/*D*).

link [50] using free-space optical beams. This link was simulated at greatest range (2.7 AU), with a transmitter aperture of 30 cm, a downlink wavelength of 1064 nm, and an average laser power of 5 W. The design of this link allocates to the ATP System a pointing loss of 2 dB with a PIF probability of 0.12 percent. Another example is a design for a Europa Orbiter-to-Earth link with a data rate of 400 kbps at 6.4 AU [49]. This link had a transmitter aperture of 30 cm, a downlink wavelength of 1064 nm, and an average laser power of 3 W. This design allocated to the ATP System a pointing loss of 2 dB with a pointing induced fade (PIF) probability of 1 percent.

Transmitter pointing errors can result in undesirable signal fades at the receiver. These fades decrease the signal power level, which in turn cause a significant degradation of the coded link performance. Therefore, the probability of fades (aka. pointing induce fade probability) that can be tolerated by the link also needs to be specified as part of the ATP pointing requirement. For a particular jitter and bias error of the system, the resulting PIF will depend on the allowed mispoint angle (allowed mispoint loss). As an example, in Fig. 5-24, this relationship is plotted for a jitter and bias error of 0.5 μrad each (bias plus 3 times jitter for a total of 2 μrad mispoint angle). As can be seen, with this mispoint angle of 2 μrad, the probability of a PIF is about 0.3 percent.

Because of the statistics of the error distribution, the PIF depends on the particular distributions of the jitter and bias error, even though the total

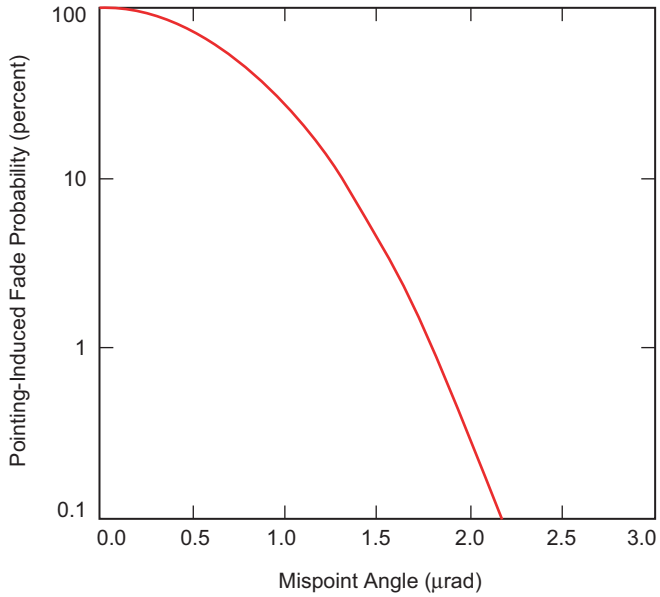


Fig. 5-24. PIF vs. mispoint angle (μrad). Jitter and bias are set at 0.5 μrad.

mispoint angle may be constant. Figure 5-25, shows this dependence by plotting the PIF as a function of allocated loss for cases of different jitter and bias distributions, but keeping the total mispoint angle constant. In this example, the total mispoint angle of bias plus three times σ equals 30 percent of the transmit beamwidth, but with different proportions between bias and jitter. As an example, for a 3-dB mispoint loss, the PIF varies from 0.7 to 0.3 percent for a {bias, jitter} allocation of {0.3, 0.9} µrad to {2.1, 0.3} µrad, respectively. To a first order, the PIF and loss are the same as a function of total mispoint angle, except for extreme cases where either bias or jitter are near zero.

The total mispoint angle includes jitter and bias errors. When designing an ATP system, allocations are made to these two categories depending on the components considered, the platform micro-vibration spectrum, and the particular link scenario (i.e., range, dead-banding, and noise background sources). These allocations have to be taken into consideration while simultaneously meeting the PIF requirement. In summary, to meet the requirements of mispoint loss with a certain PIF, care must be taken to design the jitter and bias errors judiciously. Of course, this value is limited by what the system components can perform, but it sets a preference upon which to set the requirements for jitter and bias errors.

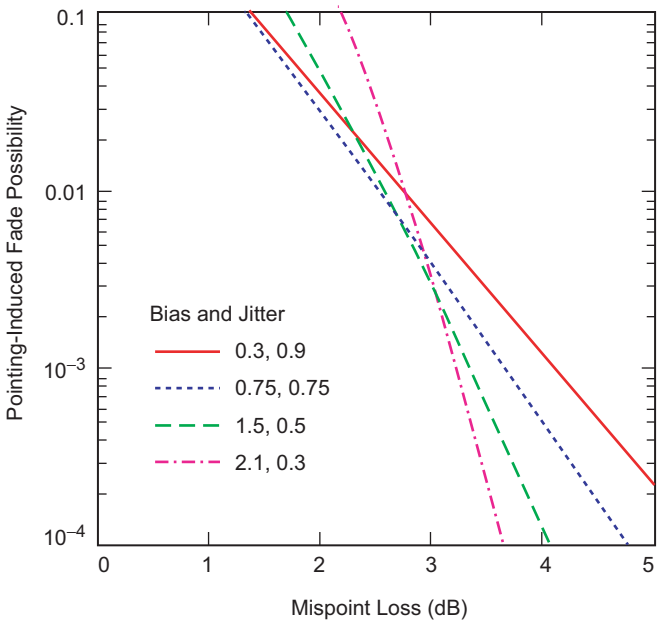


Fig. 5-25. Pointing-induced fade (PIF) probability vs. total mispoint loss (dB) showing dependence on varying jitter and bias error distributions.

5.3.2.2 Pointing-Error Budget Allocations. The pointing accuracy can be book kept in a pointing-error budget. The main sources of error are the pointing knowledge jitter, residual tracking jitter, knowledge bias, and misalignment bias due to thermal/mechanical effects. These relate to the main functions of the subsystem. The pointing requirement, as derived in the previous section, is allocated to these terms based on the estimated performance of the sensors, algorithms, mission parameters and environmental factors (e.g., spacecraft microvibrations and background signal). The total mispoint angle is equal to the bias term plus three times the jitter term (1 sigma). A reasonable allocation of the total mispoint angle partitions divides it into four parts and distributes it into one part for bias and three parts for jitter, as shown in Table 5-10 for a 30-cm and 50-cm flight terminal aperture with a 1064-nm downlink wavelength. This allocation can be later refined once estimates of the separate components are obtained. But, this initial apportioning serves as a guide to set the initial requirements that the subsystems would need to meet in order to support high rate deep-space optical communication links.

The jitter and bias error can be further decomposed into its major contributors. This is shown also in Table 5-10 with the major sources of jitter being the pointing knowledge jitter error (knowing the position of the receiver) and the residual tracking error, which is the amount of vibration not compensated by the tracking loop. The major sources of the bias error are the bias in knowledge of the receiver position and the bias caused by mechanical and thermal effects.

5.3.3 ATP System

5.3.3.1 Pointing Knowledge Reference Sources. Historically the pointing systems developed for optical communications have been based on using a ground laser as the reference beacon source. For deep-space links and

Table 5-10. ATP System Requirements (2-dB mispoint loss).

Total Mispoint Angle (assuming 1064-nm downlink wavelength)	1534 nrad (*D* = 30 cm)	921 nrad (*D* = 50 cm)
Total pointing jitter error (1 sigma)	383	231
Pointing knowledge jitter	271	163
Residual tracking error	271	163
Total pointing bias error	383	231
Pointing knowledge bias	271	163
Mechanical, thermal	271	163

particularly for planetary missions the pointing system called out for in these designs requires that an extremely powerful ground laser be used as the beacon source [49,50,51]. Some of the main impacts of these high power lasers on the system are to make ground operation difficult, to require reliable high power lasers, and to limit the range of the communication links.

Due to the required laser power, ground laser beacon concepts limit the range of the communication links. As is shown in Fig. 5-26, with current high power lasers, the range limit of laser-beacon-based pointing systems is limited to less than 0.4 AU. This range can be extended to about 3 AU, by introducing inertial sensors on the ATP subsystem to measure the high frequency vibrations, which then allows the laser beacon tracking camera on the flight terminal to integrate for longer exposure times and thereby improve its centroiding accuracy. Further studies have shown that the range can be extended to Jupiter with ground beacon lasers of the order of a few kilowatts [52].

The ground laser issues can be mitigated by utilizing pointing systems that rely on natural sources for the beacon. Since the pointing system still requires an absolute reference source, one can use passive celestial sources as the reference, such as stars or the Earth. This ATP approach has been called ‘beaconless’ due to its inherent nature of not using an active ground laser beacon as the reference source. With beaconless ATP concepts, the range of communications can be extended to cover the entire Solar System.

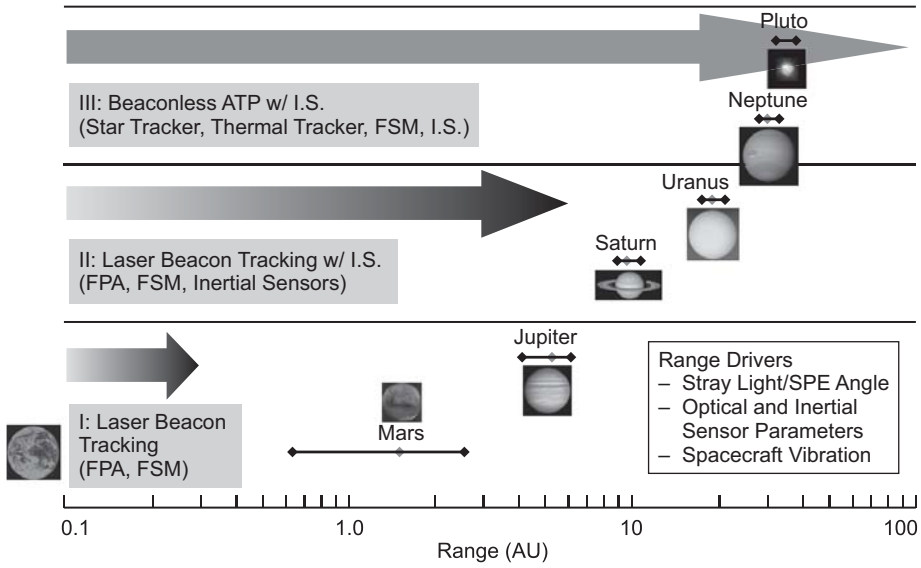


Fig. 5-26. Range capability based on ATP tracking architecture (I.S. = inertial sensor, FPA = focal plane array, FSM = fast steering mirror, SPE = Sun–probe–Earth-angle).

The technical advantages of celestial sources and corresponding sensors in the visible (also known as star tracker) and infrared (IR tracker) regions is to increase the signal to noise ratio (SNR) at high phase angles for the acquisition, tracking and pointing system. This high SNR then allows the flight terminal to accurately point the communications beam from anywhere in the Solar System to the Earth receiver without the need for a high power laser ground beacon. Both the star tracker and IR tracker concepts provide the reference information needed to precisely point the downlink beam. This information is provided at a high enough resolution to enable sub-microradian pointing systems. And with high bandwidth sensors (e.g., gyros, angular displacement sensors) integrated in the loop, the information is provided with sufficient bandwidth to compensate for spacecraft vibrations.

5.3.3.1.1 Optical References. To achieve the desired pointing performance, the orientation of the telescope with respect to the Earth must be determined. This requires a high-accuracy tracking mode to measure an absolute attitude (pointing knowledge) reference target (also known as beacon). The target can be an uplink beacon laser from Earth, the Sun-illuminated Earth-visible signal, the thermal emission from Earth, or other celestial sources, such as the Moon or bright stars. Optical references are used to provide absolute line of sight (LOS) pointing knowledge. From a celestial reference for which the J2000 location is also known in telescope coordinates, and given a S/C-to-J2000 attitude estimate (primarily for twist about the boresight), the full telescope-to-J2000 coordinate transformation can be computed.

The optical reference target is used to determine the LOS of the optical system. This measurement is corrected for distortion, jitter, etc. The (estimated) J2000 location of the Earth centroid and the measurement are then used in the attitude calculation, which in turn is used to estimate where the receiving station will be when the downlink signal reaches Earth. Except for the visible illuminated Earth, all these sources have a predictable light distribution where the mapping from a centroid measurement to a J2000 location is only limited by straylight, noise, S/C jitter, and modeling error. In the case of the visible Earth image, it is additionally limited by the ability to compensate for albedo variations that are a function of weather.

Key considerations for the selection of the optical absolute attitude reference source include the following.

1) Expected signal level and track rate: how bright and how high a track rate can be achieved?
2) Signal availability coverage: When is the source available?
3) Stray light considerations: How significant is the stray light contribution during usage?

4) Target feature location knowledge: How well do we know the location of what we are measuring (e.g., the brightness centroid of the Earth shifts due to albedo variations, contributing to error in the knowledge of the location we are measuring) contrasted to an error in the measurement process itself?
5) Derived point-ahead accuracy: How well can we determine the pointing for downlink?
6) Expected signal wave-band and detector responsivity.
7) ACS requirements, attitude knowledge required from the S/C: Assumptions are that the S/C gives attitude knowledge better than 1 mrad about the telescope boresight, allowing a single target to be used for the tracking function. Needed are 1 mrad in twist for point-ahead, which is generally available, and 160 μrad in twist for Moon-Earth tracking.
8) Field of view (FOV) considerations: For optimal performance in Earth tracking, the FOV should be as small as possible while still containing the Earth during acquisition and dead-band motion. For example, the combination of spacecraft dead-band and pointing uncertainty appear to require about 5 mrad minimum FOV diameter for a Jupiter range.

5.3.3.1.2 Summary of Possible Pointing Targets. Obtaining an accurate celestial reference is a critical step in pointing the optical downlink. Table 5-11 summarizes five tracking approaches. The Earth image provides a bright reference that is close to the receiving location both in the visible band and the long-wavelength infrared (LWIR) band. In the visible tracking approach, the Earth albedo variation is calibrated with occasional laser beacon, Earth–Moon, or Earth–star image tracking. At high phase angle when the Earth image is dim, uplink beacon tracking can be used to provide the accurate reference.

Except for the star tracking option, the S/C attitude is required for downlink pointing (boresight twist is needed for the downlink pointing). For star tracking, multiple stars are expected to be in the FOV, and the point-ahead angle is determined from ephemeris and star measurements. All options have outage at superior conjunction. The star tracking option will probably require a larger FOV to guarantee coverage, to possibly as large as 2 deg.

5.3.3.2 Pointing System Architecture. Over the past decade, JPL has adapted the design of JPL's optical ATP architecture to encompass all deep-space ranges within the Solar System. The driving factor behind this development is the lack of a high-intensity reference source in deep space. The high-intensity reference is a critical source of information for overcoming the two largest pointing errors: estimation of the receiver location and S/C vibrations. Current laser beacons do not have sufficient power to reach deep space. Alternative reference sources such as stars or the Earth have their own strengths and

Table 5-11. Comparison of various tracking approaches.

Approach	Requires	Notes	Sun Geometry Limited	Inertial Sensors Required
Laser beacon tracking	Requires uplink signal	Only applicable at close (<1 AU distances) without inertial sensors	Yes	Not near to Earth
Visible Earth-only tracking	Albedo variations cause center of brightness shift. Calibrate/live with offset error	At close distances, edge tracking can provide updates, or defocus downlink	Yes	During high phase angles
	At 0 phase and 1 AU, Earth has a magnitude of –3.8	Signal varies with phase angle/distance ~40× worse at Pluto than at Jupiter		
Visible Earth–Moon tracking	Moon has predictable Albedo, and can help determine albedo offset	Degraded if Earth–Moon has a large separation or is too close	Yes	During high phase angles
	Error of Moon measurement induces pointing error bias	Signal varies with phase angle/distance ~40× worse at Pluto than at Jupiter		During long Moon exposures
	Moon 40 times dimmer than the Earth	Moon requires 40× more exposure time		
Star tracking	Requires stars to be in FOV Requires inertial sensors	Low signal	No	Yes
	Pointing based on J2000 coordinates/ attitude	Track signal not a function of distance		
	May require offset pointing for stray-light rejection	10–20 Hz for 10^{th} mag stars		
LWIR Earth tracking	Requires cooled sensors (QWIP, HgCdTe)	Very low phase variation	Yes	Yes
	Edge detection to reduce bias error			

HgCdTe = mercury cadmium telluride
QWIP = quantum well infrared photodetector

weaknesses [53,54]. In response to these needs, a unified pointing control architecture for the system has been utilized.

A combination of a low intensity reference source and measurements of S/C vibration are used to provide equivalent pointing as a high intensity reference source. The resulting constraint from the reference source and the addition of S/C vibration measurements led to a new architecture of the

pointing system for deep-space missions. Figure 5-27 shows the diagram and the information flow among the key elements of the pointing system. A typical operating scenario is as follows: the pointing offset is computed from the telescope attitudes and the receiver location. The computed pointing offset is used to command the high-bandwidth steering mirror to direct the downlink laser beam. The telescope attitudes are estimated from the S/C vibration measurements. The receiver location is estimated from the centroids of a reference source seen on the focal plane array (FPA) and measurements of S/C vibrations. The role of the S/C vibration measurements is to compensate for the smearing and jitter of the beacon during the long exposure of the FPA due to the low intensity of a reference. This compensation is done through the enhanced centroid measurement processing which makes use of the jitter and motion during image exposure. The S/C vibration (high-frequency vibration) may need to be dampened to meet the stringent pointing-error budget.

In order to meet sub-microradian pointing requirements, the key pointing system elements should perform with high precision over a broad bandwidth. These elements are the inertial sensors, the FPA, and the steering mirror. The accuracy of the inertial sensors depends on the frequency response over the range of the vibration spectrum, electronic random noise from both the sensor and the sampling device, and any error from the algorithm that performs

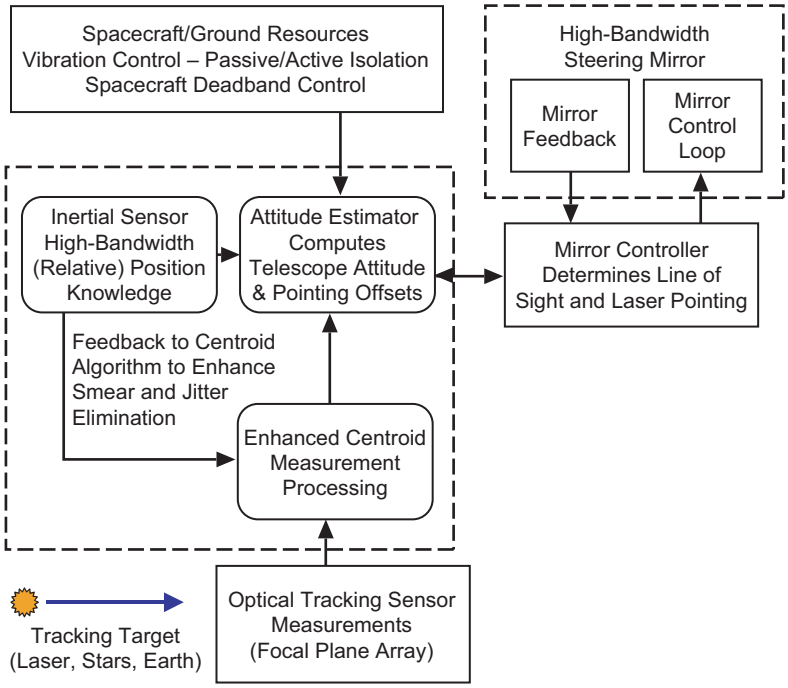


Fig. 5-27. Unified deep-space ATP architecture.

filtering and/or integration. The challenges have been in developing the integration and calibration algorithms for the initial velocity estimation, compensation for acceleration bias and scale factor bias. The main role of the FPA is to collect photons from the low-intensity reference and transfer the high-SNR signal to the sampling device, which will be used by the enhanced centroid algorithm to produce an accurate estimate of the reference position on FPA. The critical parameters of the FPA are low read noise and sub-window read capability at relatively high speed. The challenge on the steering mirror control loop is the rejection of S/C vibration on the line of sight (LOS) of the downlink beam. This requires a high-bandwidth closed-loop control, which can be achieved with the proper design of a mirror driver (controller).

5.3.3.3 Design Considerations. Because of the large trade space available to the system designer, one can easily choose to optimize the design in one aspect and ignore the other problems. An example of this is the flight-ground trade-off. One can require a larger aperture and higher power on the ground and simplify the flight system design. The optical communication technology, which is sensitive to background radiation and pointing loss, will require some adjustment in the operational methodology and mission planning process, both requiring mission inputs. In this section, major system drivers and acquisition and tracking/pointing requirements will be discussed. External parameters or constraints affecting the system design can be defined as system drivers. These parameters influence the system design at various stages with different impacts. Major system drivers include S/C attitude uncertainty, S/C vibration, stray light, and link/mission parameters (such as SPE angle, range, aperture size, wavelength, and FOV, mispoint loss allocation, and pointing induced fade probability). Smaller SPE angles give more straylight. Large range and smaller aperture size require more laser power. Wavelength directly affects transmitting/receiving efficiency among many other impacts. FOV influences acquisition /search time and tracking accuracy. Each of these system drivers is discussed in terms of its impact to the ATP system.

5.3.3.3.1 Pointing Error Sources. The overall pointing error of the subsystem includes a random contribution (which varies with a short time constant and can potentially vary from frame to frame) and a quasi-static error term, which is slowly varying.

The sources of the static pointing error include algorithm error, the error in estimating the Earth–receiver position, the ephemeris error, error in computing the point-ahead angle, and alignment errors. For example, for Earth tracking, the largest static error source is allocated to the error in estimating the geometric center of Earth using the image centroid. This error is due (primarily) to the uncertainty in image intensity distribution, and it will require periodic Earth–Moon or Earth–star calibration to achieve the allocated pointing

accuracy. The boresight alignment error and errors due to thermal-mechanical distortion are the next largest sources of static errors. These sources can be controlled with careful opto-mechanical design and with careful alignment of the optics.

The two major contributors of random pointing error are the sensor noise and control error. The sensor noises include the noise introduced by the random photon noise (shot noise) and errors introduced by the pixel non-uniformity and spatial quantization. The control loop noise includes the uncompensated platform jitter (vibrations) and noise introduced by the control loop electronics.

5.3.3.3.2 Spacecraft Attitude Control Uncertainty. The S/C attitude control uncertainty impacts the design of the FOV of the acquisition detector such that the beacon needs to be always in the FOV of the acquisition detector. Therefore, the FOV of acquisition detector should be larger than the twice the attitude control uncertainty (since this covers only one side of the 3-sigma value of the attitude control uncertainty). If the acquisition detector is also used as the tracking detector, the tracking error will increase as the FOV increases due to the reduced per pixel resolution. Therefore, there will be a trade-off between the required tracking accuracy and the acquisition FOV.

To define a typical range of S/C attitude control uncertainty, 34 spacecraft were surveyed on the JPL mission and spacecraft library website [55]. These were grouped into two categories: 27 Earth-orbiting spacecraft and 7 deep-space spacecraft. A histogram of the spacecraft attitude control uncertainty is shown in Fig. 5-28. Both types of spacecraft have been built with a wide range of attitude control capability. The control uncertainty ranged from less than 1 to

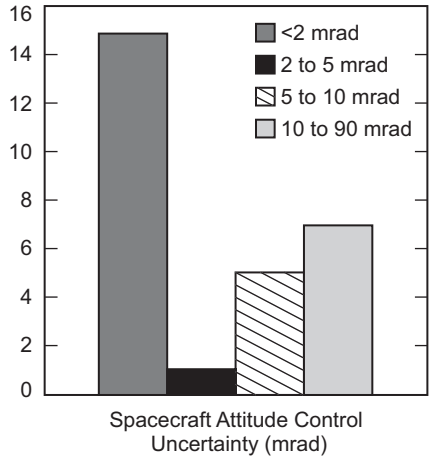


Fig. 5-28. Histogram of spacecraft attitude control uncertainty (mrad) from a survey of 27 Earth-orbiting and 7 deep-space craft.

90 mrad. In both types of spacecraft, the majority fall within +3 mrad in attitude control. This implies that an acquisition detector with an FOV of 6 mrad will cover most missions.

5.3.3.3.3 S/C Vibrations. Spacecraft platform vibrations cause jitter of the downlink beam, which increases the mispointing loss. Compensation for S/C vibration is necessary for precise pointing and generally requires high-speed updates of the downlink steering mirror on the order of few kilohertz. Without proper compensation, these vibrations can result in a mispoint of the beam on the order of more than 10 µrad depending on the vibration power spectral density (PSD).

Several S/C vibration power spectral densities (PSDs) are shown in Fig. 5-29 for the following spacecraft: Space Shuttle, Landsat, Bosch, ESA's communications satellite (Olympus), High Resolution Dynamics Limb Sounder (HRDLS), Relay Mirror Experiment (RME). As an example of mispoint jitter, the total RMS jitter for the Olympus S/C without any compensation is about 16 µrad. To reduce this jitter effect a compensating control loop is designed to stabilize the outgoing optical beam. The design of the control loop depends highly on the frequency content of the vibrations as shown on the PSD plots. The magnitude and frequency content of the spacecraft vibrations drives the need for the control system to include varying levels of passive isolation, active isolation, and active compensation to reduce the effective residual jitter impacting the downlink beam.

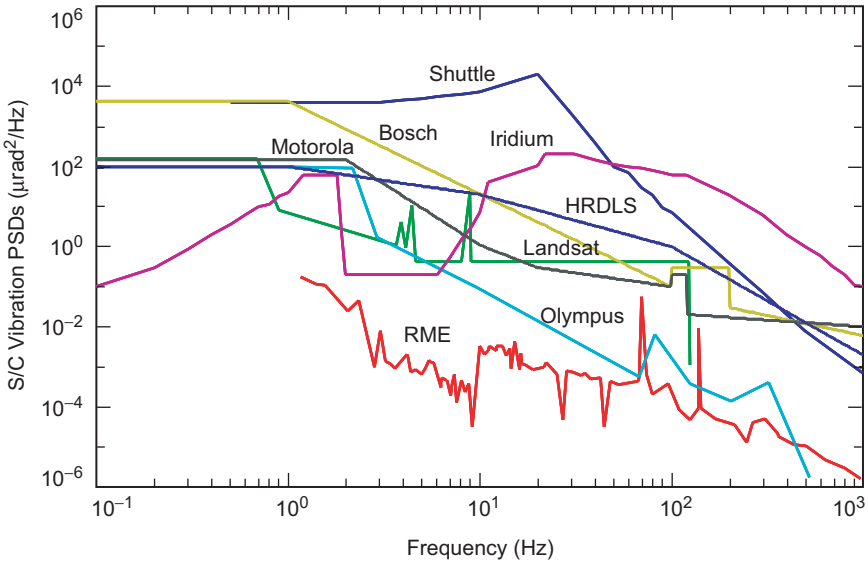


Fig. 5-29. Sample of spacecraft vibration power spectral density characteristics.

5.3.3.3.4 Detector NEA and Bias. For the image detector (or any array detector), there are several components of measurement error. The detector noise equivalent angle (NEA) is usually used to summarize the effects of errors caused by photon statistics, read noise, or other sources, which are temporally random. There are also several bias terms that are functions of the detector pixelization or optics design that may appear to be random. First, there is the high-frequency spatial bias (the S-curve, or centroiding bias) that is approximately periodic on the pixel pitch and is determined primarily by the image distribution—the point-spread function relative to the pixel response function. Low spatial frequency (LSF) bias (or macro distortion) is due to optical distortion, detector shape (flat versus spherical), focal length changes, and chromatic effects that will cause slowly varying offsets as a function of position in the FOV. The LSF bias terms are generally expected to be calibrated and included as part of the focal length calibration. The key elements of the image detector to consider are quantum efficiency, read noise, dark current detector pixel non-uniformity, full well size, and analog-to-digital (A/D) converter resolution.

- Quantum Efficiency. The ability to convert photons to electrons. This is really a simplification (like an average value over a spectral range) of the spectral response, which determines how well the detector converts photons as a function of wavelength.
- Read Noise. We use this term very loosely to mean the noise contributed in the course of reading a pixel.
- Dark Current Effects. Are separately allocated, since they are dependent on exposure time and temperature.
- Detector Pixel Non-Uniformity. Refers to the variation of the average pixel response over the detector.
- Sub-Pixel Response Non-Uniformity. Gives the variation of response within a pixel.
- Full Well. The maximum integrated signal that can be stored in a pixel and measured. Usually the full well is determined by the point where the light transfer curves become non-linear.
- A/D Resolution and Preferred States. The number of electrons per A/D least-significant bit (LSB) determines the resolution. Any preferred states in the A/D converter are modeled as an additional noise source and as reduced resolution.

The following equation for NEA is an approximation, and is based on the center of mass centroid calculations. For a well-known, static image distribution, a reduction in NEA by $\sqrt{2}$ can likely be achieved. The benefit of the center of brightness algorithm is that jitter or small angular rates do not affect the measurement as much as a “fitting” or shape dependent algorithm.

The derived equation separates the centroid noise due to the signal itself, and expresses the noise variance as the sum of the image centroid noise and the per pixel noise, such as read noise or background (e.g., stray light) noise. Errors in the background and A/D conversion are included with the R_{ij} defined below. All quantities are expressed in electrons.

Δt = the exposure time (seconds)

S_{ij} = Signal in the *i*; *j* pixel

R_{ij} = Non-signal noise contribution in the *i*; *j* pixel. For each pixel, R_{ij} is assumed to have the same variance, namely Var(R_{ij}) = Var(R_F)+ $\Delta t R_T$ with integration time Δt. This emphasizes the time-dependent behavior for stray light and dark current.

R_F = Fixed per pixel noise (1σ), such as read noise. A typical value is 5 e to 10 e for the low noise FPA such as scientific quality charge-coupled devices (CCDs).

R_T = Per pixel background signal rate (including stray light and dark current).When separating external (stray light) from detector generated (dark current), they are denoted R_{ET} and R_{DT}, respectively. The noise variance terms are $\Delta t^2 R_{ET}$ and $\Delta t^2 R_{DT}$.

X_C = Centroid computed from signal + noise

I_S = Total star signal intensity, electrons/second.

S = Total image signal, $S = \Delta t I_S$

N = Truncated half width of centroiding area. See usage in equations below.

N_P = Number of pixels involved in the centroiding area, $N_P = (2N+1)^2$

$$NEA^2 \approx \mathrm{Var}(X_C) = \left(\frac{S + Np\left(\mathrm{Var}(R_F) + \Delta t R_T\right)}{S^2} \right) \frac{N(N+1)}{3} \qquad (5.3\text{-}1)$$

Note that with non-uniform pixel response, the error due to pixel non-uniformity can become a major contributor of the NEA.

5.3.3.3.5 Digital Quantization. To convert electron counts in a pixel to digital values, an A/D converter, typically 8 to 12 bits, is used. The A/D output is referred to as DN (digital numbers, or data numbers). Because the number of electrons can be quite large (50,000 to 250,000 electrons) compared to the resolution of the A/D converter, full knowledge of the electron count is lost due to the truncation error, and so is some knowledge of the image. This error is treated as being a random distribution; for the case where the A/D has no

preferred states, it would be described by a uniform distribution. Since we estimate the mathematical centroid of the image from the DN, the error due to this conversion is needed to evaluate the effect of the noise sources. For analysis, the noise contribution due to quantization is a uniform distribution:

$$\begin{aligned}\text{A/D Variance} &= 1\sigma^2 \text{ noise/pixel} \\ &= \frac{1}{12} \times (\text{number of electrons/resolution})\end{aligned} \tag{5.3-2}$$

At best, the resolution is the number of electrons per DN for a perfect A/D converter. Using the value of the number of electrons per 1.5 DN is probably a better estimate of what is actually achievable.

5.3.3.3.6 Interfering Image Sources (Interlopers). If there are interfering targets in the same measurement area (e.g., centroiding window), the composite measurement will be shifted. Assume a target with signal BE and centroid location (x_E, y_E), and a close, interfering object with centroid (x_I, y_I) and brightness B_I. Then the combined system has centroid

$$\left(x_I, y_I\right) + \frac{B_I}{\left(B_I + B_E\right)}\left(x_E - x_I, y_E - y_I\right) \tag{5.3-3}$$

Consider an example where the Earth is 20 μrad in diameter, with the *y* detector direction aligned with the Earth pole, and a star near the pole and 2.5 magnitudes dimmer (i.e., 2.5 magnitudes if the Earth is fully lit and 10.0 magnitude if the Earth is at a 160-deg phase angle). The system centroid shifts by

$$0.09 \times \left(x_E - x_I, 10\ \mu\text{rad}\right)$$

The error is nearly 1 μrad in the *y* axis alone. The 1-μrad error will result in pointing loss. The probability of a star being near, but not behind the Earth is small, but there will be times where the image is degraded in this way. Once a mission is planned, there should be an evaluation of the stars that are angularly close to the Earth (seen from the S/C) to predict/plan degraded pointing or devise workarounds.

5.3.3.3.7 Pixel-to-Pixel Non-Uniformity. Pixel non-uniformity is a property of the individual pixel response, and hence, it does not change during short periods of time. (It can change with radiation damage.) The effects on centroid error appear as a slowly changing bias while an image moves across a pixel. There are two effects to consider. The first is when there is high background signal, such as high stray light. This can be treated in an RMS sense; if the

(uncalibrated) RMS non-uniformity value is σ_U, and the background rate signal is R_T electrons per pixel, then the RMS spatial noise variance is $(\sigma_U \Delta t \times R_T)^2$ per pixel for an integration time of Δt. This can be treated for simplicity in the analysis as though it were a read noise term. Note that the error uncertainty contribution grows with the signal (linearly with time), not with the square root of the signal as do noise contributions, so that this contribution becomes much worse as the background grows.

The second effect is when the scene background is low, and only the image signal is of importance. Considering the worst case when only one column is not uniform while the other columns are uniform for an $M \times M$ centroid window where ρ is the pixel responsivity, $\rho = 1 \pm$ the pixel non-uniformity. The corresponding centroid error due to the pixel non-uniformity is

$$C_x = \frac{M(M-1)(\rho-1)}{2(M(M-1)+M\rho)} \tag{5.3-4}$$

If the responsivity, ρ, is a uniformly distributed random variable (with half the range corresponding to the non-uniformity value), then the RMS error will be $C_x/\sqrt{12}$. If the responsivity, ρ, is a Gaussian random variable, then C_x is the RMS centroid error due to the pixel non-uniformity. The magnitude of this error is therefore proportional to the size of the centroid window and the distribution and magnitude of the pixel non-uniformity.

5.3.3.3.8 Platform Jitter Considerations. Platform jitter micro-vibrations (due to the amplification or transmission by the S/C or terminal or optics structure) will degrade downlink pointing. Such platform jitter can be induced by reaction wheels, thruster rings, external torques, or other moving parts on the S/C, such as the steering mirror for downlink control, or other instrument steering mirrors. The key factors in minimizing platform jitter are

- Passive or active isolation of the optical communications terminal (OCT) from the S/C.
- Using common mode design for measurement and control whenever possible.
- High rate measurement loop using inertial sensors to measure the change of motion of the telescope pointing.
- High rate measurement loop to determine the direction of the pointing steering mirror in telescope coordinates.

5.3.3.3.9 Point-Ahead Angle. Here we assume that the Earth is moving at a constant angular rate about the Sun, which gives a velocity of about 30 km/s. When the reference measurements are based on the Earth/Moon, then twice the

one-way light time is used to calculate the difference between where the Earth position will be when the downlink signal reaches the station, relative to the OCT-measurement-calculated location. When only stars are used as references, then there is no such effect. The computation of the Earth location at the downlink signal arrival time is now relative to the absolute J2000 frame, and may be worse or better, depending on the knowledge of the J2000 Earth location in the star measurement-based frame.

At a distance D from the Earth, the round-trip light time is $2D/c$, where c is the speed of light in kilometers per second $= 2.998 \times 10^5$ km/s. The worst-case point-ahead angle (at inferior conjunction, 180-deg phase angle) is independent of the distance between the S/C and the Earth, and it is given by

$$\begin{aligned} \text{Point–ahead angle} &= \frac{2D}{c} \times \frac{\text{projected velocity}}{\text{D}} \\ &= \frac{2 \times 30}{(2.998 \times 10^5)} = 200\ \mu\text{rad} \end{aligned} \tag{5.3-5}$$

As an example of the effect, with a twist error of 1-mrad accuracy about the boresight, the induced point-ahead error is 0.2 μrad radial. During a full Earth orbit, this error will vary between + 0.2 μrad due to the changing point-ahead, with the largest at 0 and 180-deg phase angle.

The worst-case error in the receiving station location due to rotation is given by

$$\begin{aligned} \text{Earth rotation displacement} &= \cos(\text{lat}) \times 3.6 \times 10^{-5} \\ &\quad \times (\text{angular diameter})/\text{second} \end{aligned} \tag{5.3-6}$$

where cos(lat) is the cosine of the receiving station latitude.

For a receiving station located on the equator, and at a distance of 4 AU, Earth is about 20 μrad in diameter. For this case, the receiving station moves up to 0.1 μrad relative to the S/C in 140 s. For beacon tracking, the same considerations are required when considering the location of the uplink beacon (there possibly may be common-mode cancellation).

5.3.3.3.10 Solar Conjunction Availability. Limited solar-conjunction availability is imposed by the Sun-spacecraft-Earth geometry as well as by the Sun-Earth-spacecraft geometry, which contributes stray light. In addition to the effect of increased background noise at the Earth receiver, at small Sun-Probe-Earth (SPE) angles, the spacecraft's pointing and tracking detector, if co-boresighted, may experience an increase in background noise due to the photon noise in the straylight from the Sun, leading to an increase in pointing error and, at worst case, inability to detect the Earth image or uplink beacon signal.

5.3.3.3.11 Weather-Related Availability of Ground Laser Beacon. The limit imposed by weather-related availability of the ground laser beacon is caused primarily by the limited clear weather probability from the ground stations. Furthermore, it also imposes restrictions on ground station hand-offs and flight-ground coordination if a cooperative beacon tracking scheme is used. Atmospheric availability due to cloud coverage is a significant issue for optical communication systems.

5.3.3.3.12 Stray Light. Stray light is defined as any unwanted photon reaching the terminus of an optical system. Such photons may, among other methods, arrive at the terminus through scattering from mirror imperfections or contaminants, diffraction by the secondary mirror or its supporting structure, or through scattering from the baffles. The dominant source of stray light will be sunlight that is approximately a billion times brighter than the Earth. Even a miniscule fraction of sunlight scattered from the primary mirror or baffling will be significant.

Stray light (or any locally uniform background signal, including dark current) gives two separate types of noise. The first is the shot or photon noise, which can be treated as simply an additional photon noise source. The second is due to the non-uniform background—pattern noise—that contributes to the image error as a bias term. The bias term does not decrease with longer exposure time as does the second term, which is basically the contribution due to photon statistics. For high stray light rates, such as at a phase angle of 160 deg and say with a 5×5 centroiding region, the 1σ NEA (in pixels) is $> 10\sigma_U R_T / I_S$. When the stray light rate equals the signal rate, the non-uniformity uncertainty σ_U must be kept smaller than 0.15 percent to hold the stray-light contribution to less that 0.1 μrad, about 0.14 pixels. Since this is not a priori achievable in the detector design, this places a strong requirement to calibrate and/or measure the background in high stray-light environments. This can be done by maintaining either a calibrated pixel response map or by calibrating in real-time the local background in each pixel near the spot as the image (and stray light) moves. In summary, background calibration is required for high stray-light conditions, while for high phase angles the calibrated pixel non-uniformity is required to be held to less than 0.15 percent.

5.3.3.3.13 Radiation Considerations. There are two basic types of effects to consider. The first can be considered as a single-event upset (SEU), where a particle hits the detector or associated electronics. This produces false images, and it can cause degradation of measurement accuracy. The second, the fluorescence and Cerenkov effect, is primarily caused by high-energy electron flux passing through glass elements and resulting energy deposition, and it increases with the mean path length in the glass as well as the flux rate. The main effect expected here is increased background; this may be significant

depending on the amount and type of glass in the system. SEU radiation presents a different problem—namely, electrons or other high-energy particles depositing corrupting signal into pixels can degrade images.

5.3.3.3.14 Pointing Error Trade Sample. The magnitude and sources of errors are quite varied. These errors must be judiciously traded and managed to meet the overall pointing requirements. As a sample, a laser beacon tracking scenario is analyzed. There are various pointing error sources that can be classified into three groups (Fig. 5-30). The RSS (root- sum-squared) value of the total dynamic pointing error was allocated to meet the sub-microradian pointing requirement. The error allocation has been done using the Acquisition Tracking Link Analysis Software [56] (ATLAS) simulation tool. The simulation results are based on the projected pointing system performance such as FPA read noise, closed control loop update rates, and inertial sensor (accelerometer) noise. As indicated in Fig. 5-30, the largest error comes from S/C vibrations, which are determined by the specific S/C vibration and the disturbance rejection of the tracking control loop. The second largest error source, inertial sensor noise, is mainly determined by the given noise specifications (Honeywell QA-3000 accelerometer specifications used in this example) [57]. The centroid errors on transmit laser (NEA, pixel-to-pixel non-uniformity, spatial quantization) are relatively smaller than those of the beacon since the transmit laser power on FPA can be easily controlled to meet the requirements.

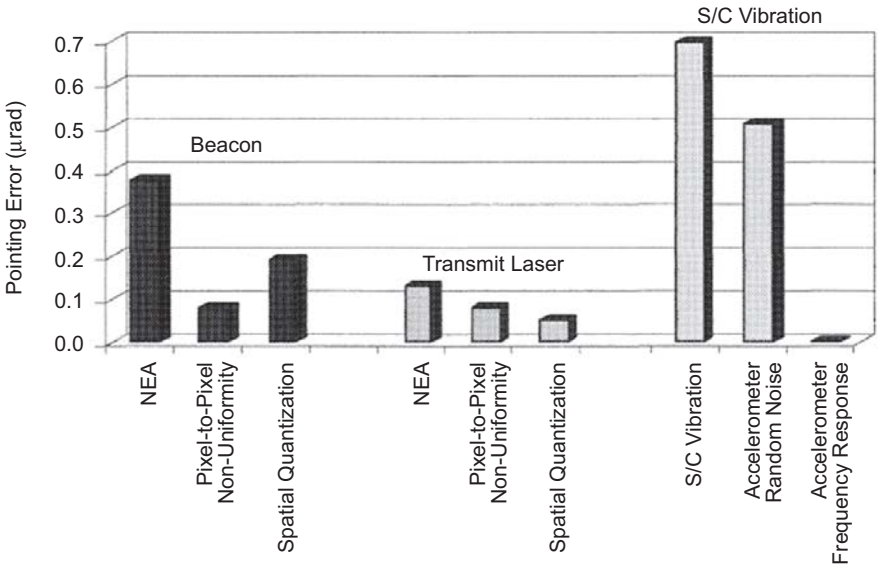


Fig. 5-30. Pointing error allocations to various error sources using simulation results.

5.3.4 Cooperative Beacon (Ground Laser) Tracking

Uplink beacon tracking is an attractive option from the implementation point of view. A laser signal presents a well-defined point spread function, and the very narrow spectral bandwidth allows easier rejection of the background-scattered sunlight. The problem with uplink beacon tracking is the amount of available power. For example, at Jupiter (6 AU range) using a 500-W uplink laser operating at 532 nm, for example, the number of photons received at the tracking detector is approximately 2.3×10^5 photons/s. Even with aggressive assumptions on the uplink power and beamwidth, the amount of available uplink power is not expected to exceed 5×10^5 photons/s. Compared with the required signal power of approximately 10,000 *detected* photons/frame, it can be seen that uplink beacon tracking alone cannot provide the tracking bandwidth required to control the pointing error [54].

Signal availability has two key limitations. The ground laser beacon transmitting station must be in direct line of sight with the spacecraft, and the weather conditions must be relatively clear to permit transmission of the optical beam.

Stray light can be reduced, using a narrowband filter. It appears that a 0.2-nm to 1.0-nm wide filter can be used, reducing the stray-light contribution. A narrowband filter is needed to filter out both the stray light from the sun and light from the Earth background. The Earth at zero-phase angle, when seen from Jupiter, has a similar intensity and spectral distribution as that of a 0th magnitude black body. Using the blackbody signal model, the Earth at 0-deg phase angle and 5-AU distance generates about 6×10^6 photons/nm/s.

By tracking an uplink beacon, the knowledge of the ground station location is essentially perfect. The difficulty is distinguishing the beacon location from the Earth background. The laser output (5×10^6 photons/s) is clearly overwhelmed—especially considering the variability of the total signal from the Earth. Even with a narrowband optical filter of 0.2-nm bandwidth, the background photon flux is still 5 times higher than the beacon signal strength. In order to perform beacon tracking, therefore, accurate calibration of the Earth background will be required. A practical limit for the beacon tracking will be at a SEP angle of 30 deg when the Earth background is of more comparable strength to the beacon uplink.

Above 90-deg phase angle, the Earth background becomes less of a problem, and beacon tracking becomes more feasible. Depending on the phase model, the Earth background is about 10–25 percent of the beacon signal strength. At very high phase angles, (>160 deg), the Earth background is (depending on the model assumed) on the order of 5×10^3 incoming photons/nm/s. Here the flux from the laser beacon is significantly larger; eliminating any image-induced bias.

At smaller phase angles, there can be sizeable centroid error from the Earth background. Given two centroid measurements—(x_L, y_L) for the beacon and (x_E, y_E) for the Earth image, with intensities B_L for the laser and B_E for the Earth (the intensities are measured in the waveband)—the centroid of the system is shifted from the uplink beacon source by an amount

$$(x_L, y_L) + (x_E - x_L, y_E - y_L)\frac{B_E}{B_L + B_E} \tag{5.3-7}$$

With an Earth image that is 20 μrad wide, the separation between the Earth and beacon centroids $(x_E - x_L, y_E - y_L)$ could be as large as 20 μrad. Considering the brightness ratio, $B_E/(B_L + B_E)$, a 1:10 ratio would cause a 2.0-μrad shift in the estimated centroid location. This is much too large an error. A 1:100 ratio is acceptable in this case, since the laser spot would be located to 0.2 μrad, approximately the error for measurement error. Because of the possible atmosphere-induced intensity fluctuations of the laser signal, and the variability of the Earth background intensity, knowledge of the intensities values for B_L and B_E could have significant errors. Additionally, the centroid location for the illuminated portion of the Earth, (x_E, y_E) is susceptible to variation in Earth intensity (albedo variation). Some of this error can be taken out by knowing the position of the laser beacon relative to the lit Earth, but there still will be residual errors.

Because of the predictability of the laser spot shape, a better centroiding algorithm (such as the maximum likelihood algorithm or some other model based algorithm) might yield better results by working on the part of the signal away from the lit limb. There are some obvious operational complications, but at last resort, if the uplink beacon could be pulsed at slow intervals, Eq. (5.3-7) could be calculated with both $B_L = 0$ (laser off) and using the laser signal. The change in brightness (if the image transmission were controlled on the Earth) could be used to determine whether the laser image was combined in the signal, or the image included the Earth only. Knowing one of the terms in Eq. (5.3-7), as well as the sum, gives a more accurate location of the point (x_L, y_L). A procedure like this could also be used as part of a calibration procedure to attempt to correct for Earth's albedo variation.

For the case where the Earth is large, but less intense than the laser spot at the relevant wavelength, the expected size in pixels of the laser spot can be used to limit the centroid window area and the contribution of the Earth image.

5.3.5 Noncooperative Beacon Tracking

Three major noncooperative beacon ATP technologies have been developed at JPL in the past decade. (These methods are also sometimes

referred to as beaconless tracking.) The distinguishing factor is the type of reference sources, namely visible Earth images, long-wavelength infrared (LWIR) Earth images, or visible stars. A visible Earth tracker was conceived as the first beaconless tracker. Although several potential solutions have been proposed for the visible Earth tracker, albedo variation was identified as the major challenge. To solve the albedo variation problem, three centroiding algorithms based on the edge detection and maximum likelihood criteria were explored. Two different approaches, LWIR Earth tracker and star tracker, were later proposed to overcome the albedo variation problem, and the analysis shows that attitude jitter of less than 150 nrad (1-sigma, single axis) can be achieved, which would meet the requirements of the current deep-space optical communication pointing system.

5.3.5.1 Earth Tracker–Visible Spectrum. Visible Earth-image tracking appears to be attractive because of its high brightness, and importantly, it does not require an uplink laser beacon (which considerably simplifies link operation). However, visible Earth-image tracking requires accurate compensation for centroid shifts (bias) due to the Earth weather-induced albedo variations. Figure 5-31 illustrates this problem. Figure 5-31(a) shows Earth images taken by the Galileo spacecraft as it receded from the Earth. The image contains intensity variation due to the presence of cloud pattern. The same image as would be seen through diffraction limited optics is shown in Fig. 5-31(b). The diffraction-limited point spread function reduces the image contrast significantly. Finally, at Europa distance, the image is only several pixels in diameter, and the detector pixel quantization lowers the image resolution, as illustrated in Fig. 5-31(c). It has been shown that achieving centroid accuracy (1-sigma) of 0.1 pixel requires that the intensity need to be known to be within 10 percent of the true value [53]. Since the average Earth

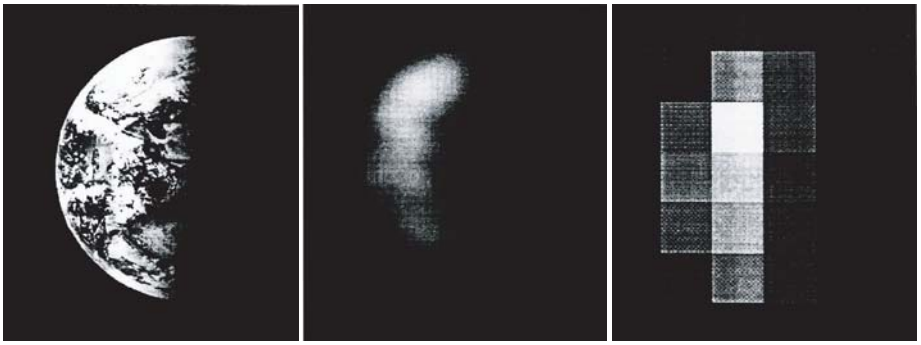


Fig. 5-31. (a) High resolution Earth image as seen by the Galileo spacecraft as it receded from Earth, (b) same image of Earth as seen through a diffraction-limited optical system. The image blurring is due to the diffraction-limited spread of the receiver optics, and (c) same image of Earth as seen through an array detector with a pixel FOV of 3.5 μrad.

albedo variation is much more than 10 percent, computing accurate centroids from the pixilated image with albedo variation presents an enormous challenge to the design.

Conceptually, the optical communication subsystem can mitigate the albedo variation problem by performing periodic imaging of the Earth image with other celestial references such as the Moon or nearby stars which have a more predictable distribution of light. Since the distance between the Earth and other celestial references are accurately known, the albedo offset of Earth can be deduced. The limiting factors in the albedo offset include the measurement error of the celestial reference sources, spacecraft jitter, and stray-light noise. For the case of using the Moon for bias compensation, the centroid measurement uncertainty increases as the mission range becomes large due to the dim Moon image. For the stars, it is independent of the mission range, and the details of analysis are presented later in the star tracker section.

The concept of the ATP system using the Earth tracker is illustrated in Fig. 5-32. First, Earth is imaged on FPA and the centroid is computed. Then, the ground position is deduced from the known distance between the geo-centroid of the Earth and the ground receiver position. The pointing vector to drive the fine-steering mirror (FSM) is the difference between the current transmit laser position and the ground receiver position with the addition of point-ahead vector to account for the two-way light travel time between the receiver and the transmitter. When the Earth image cannot provide sufficient signal for high-rate tracking, inertial sensor measurements are used to propagate the knowledge of the optical boresight at a higher rate between FPA updates.

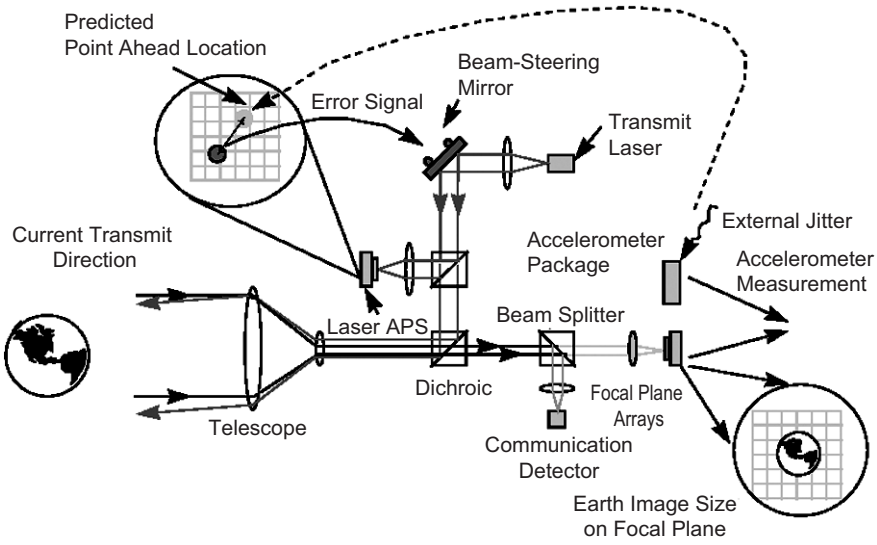


Fig. 5-32. The ATP concept using the visible Earth tracker.

Functionally, FSM jitter is measured using the reflection of the downlink laser off the FSM, while telescope jitter or S/C vibration is measured using a combination of FPA and inertial sensors.

One of the key considerations for the Earth image tracker to provide an accurate pointing is the Earth signal level and availability. The signal level affects the tracking rate, and the signal availability directly influences the communication link availability. Earth signal (reflected sunlight) has about the same spectral distribution as the Sun with most of the energy in the 400–900 nm band, which requires the tracking detector to have high spectral response for 400–900 nm. At low phase angle, an Earth image provides sufficient photons for tracking even at a 2-kHz frame rate. However, the signal from the Earth image has a wide variation, both in total and spatial distribution. The current best estimates show that the total signal follows a phase law between that of the Moon and a Lambertian sphere. There is some direct evidence the total Earth signal can be as bad as the Moon model under some weather conditions. Table 5-12 shows signal estimates versus phase angle with no optics loss, and it assumes a low quantum efficiency (Q.E.) of 25 percent, such as from a photogate (PGT) active pixel sensor (APS) design. Assuming the minimum requirement of 10,000 electrons per frame [54] and Earth at Jupiter distance, maximum frame rates of 6.2 kHz, 100 Hz, and 20 Hz are available for the phase angles of 90, 160, and 170 deg, respectively. Therefore, use of inertial sensors is required somewhere between 90 and 160 deg of phase angle. For the signal availability, it is limited by the angular separation between the Earth and the Sun, and between the Earth and the Moon. Based on conic elements, the Earth as seen from Jupiter, for example, nearly always has sufficient separation from the Moon, since the Moon's orbit is inclined to the ecliptic by about 5 deg. The angular separation between the Earth and the Sun is limited by straylight considerations, rather than overlapping images.

Other key considerations include stray light from the Sun. When the stray light rate becomes high, not only does the added photon noise cause additional centroid error, but the pixel non-uniformity also becomes much more significant and requires pixel response calibration. Two to three degrees separation is considered as the current requirement.

In the next sections, we present three centroiding methods to mitigate the albedo variation problem.

5.3.5.1.1 Maximum-Likelihood Method. The maximum likelihood method is considered to be an optimal solution if one can assume the existence of a perfect reference image. The acquisition process using this approach has been developed for the cases of rigid translation movement between the two image frames under static conditions [58,59,60]. It has been shown that the optimal spatial acquisition requires solving two nonlinear equations to estimate the coordinates of the transceiver from the received camera image in the

Table 5-12. Estimated Earth signal at Jupiter distance, assuming QE of 25 percent of photogate mode APS and 100 percent optical efficiency.

Phase Angle	Distance	Total Photons, 400–900 nm	Electrons, No Phase Law, No Optics Loss, PGT Device	Electrons, Lambertian Model, PGT Device	Electrons, Moon Model, PGT Device
90	5.2 AU	5.7×10^9	7.0×10^8	1.7×10^8	6.2×10^7
160	4.3 AU	3.9×10^9	1.0×10^9	2.8×10^7	1.0×10^6
170	4.3 AU	3.9×10^9	1.0×10^9	7.0×10^6	2.0×10^5

transformed domain when the uncertainties between the reference image and the received image are modeled as additive white Gaussian disturbances. The optimal solution can be obtained iteratively by solving two linear equations. Numerical results using a sample Sun-lit Earth as a reference image demonstrate that sub-pixel resolutions can be achieved in a high disturbance environment. Spatial resolution is quantified by Cramer-Rao lower bounds [61].

The above process was applied to acquire a Sun-lit Earth image. The image is assumed to be detected by a CCD array and corrupted by additive white Gaussian disturbances such that $(S/N)_\ell = 1$, where $(S/N)_\ell$ is the average signal to noise ratio at time t_ℓ defined by

$$\left(\frac{S}{N}\right)_\ell = \frac{1}{\sqrt{M_d N_d}} \frac{\left\|\vec{\mu}_\ell\right\|}{\sigma_\ell} \tag{5.3-8}$$

Table 5-13 lists the lower bound for the estimation variances for different sizes of CCD arrays. As an example, consider a requirement to achieve a sub-pixel resolution of 10 percent during acquisition. The corresponding σ_ℓ computed to be 16.68 for a signal-to-noise ratio of unity for a 4×4 detector array. From Table 5-13, the standard deviation of $(\hat{x}_\ell - x_\ell)$ is evaluated to 0.13 pixels. If the standard deviation is used as a measure of the resolution capability of the process in the spatial coordinate, the 4×4 detector array will not be able to meet the requirement. Following the same argument, a sub-pixel requirement of 10 percent can be achieved by using an 8×8 CCD array, which indicates that better accuracy can be achieved with higher resolution reference image.

5.3.5.1.2 Edge Detection Method. The edge detection method uses the fact that, even though the Earth albedo varies widely across the Earth surface, there

is a relatively high contrast between the solar illuminated surface and the dark space background [62]. Furthermore, since the distance between Earth and the spacecraft is known, the radius of the limb can be calculated easily. If the Earth limb can be accurately extracted from the focal plane image, the precise orientation of Earth can be derived. The receiver location can then be calculated using the spacecraft ephemeris and the relative orbital geometry. Although both maximum likelihood and edge detection methods are technically sound, in practice their performances are influenced by the varying Earth albedo, solar illumination, as well as the receiver point spread function and detector pixel quantization. The combined effects of these factors on the accuracy of extended imagery process are very difficult to analyze. On the other hand, it is necessary that the algorithms be fully characterized before the method can be proposed as a replacement for a beacon-based tracking system. A software simulator approach in which the algorithms are tested against a large number of test images was taken as the logical solution to the algorithmic verification process.

The test images were generated with realistic parameters such as the proper Sun-Earth-Probe (SEP) angle to the spacecraft, representative albedo contrasts and spatial correlation properties, the blurring due to the receiver optics point spread function, and discrete quantization due to finite CCD pixels. Figure 5-33 is the histogram of the simulation results of the maximum likelihood method. The error in estimating the image offset is normalized to the full-width half maximum of the Gaussian point spread. Note that most of the simulations resulted in a normalized error of 0.2 to 0.5 range. An improved correlation algorithm using iterative steps and nonlinear estimations was shown to provide improved results for the limited number of cases tested [62]. Further tests are needed to characterize the RMS error in estimating the image shifts.

Table 5-13. Performances of various detector array sizes.

CCD Array Size		Normalized Lower Bound	
Md	Nd	$\frac{\mathrm{var}(\dot{x}_\ell - x_\ell)}{\sigma_\ell^2}$	$\frac{\mathrm{var}(\dot{y}_\ell - y_\ell)}{\sigma_\ell^2}$
2	2	5.72×10^{-4}	3.64×10^{-4}
4	4	6.13×10^{-5}	2.61×10^{-5}
8	8	1.06×10^{-5}	2.77×10^{-6}
16	16	3.59×10^{-6}	4.39×10^{-7}
32	32	9.52×10^{-7}	7.32×10^{-8}

Because the Earth limb forms an arc of a circle, a simple curve-fitting routine can be used to derive the Earth location. When there is a high contrast between the limb and background, edge detection and curve fitting provides an accurate way of estimating the Earth centroid. When the spatial contrast is reduced, such as when excessive blurring by the receiving optics is experienced, or when the limb point is not directly illuminated by the Sun, then the edge extraction routine may have problems identifying the correct limb points. The focus of the edge-detection algorithm development was on adapting (modifying) the edge-detection method such that an accurate edge could be extracted under such conditions.

As the plots in Fig. 5-34 show, The results of the edge-detection method were more encouraging than the correlation algorithm, but they still did not quite reach the 0.1 normalized error. These results are plotted in Fig. 5-34. The results for the crescent (large SEP angle) cases show that 95 percent of the cases producing a centroid estimate below the 0.1 beam divergence. The combined results for all cases were not as good, with approximately 80 percent of the cases below the 0.1 normalized error and more than 95 percent of the cases below 0.2 normalized error.

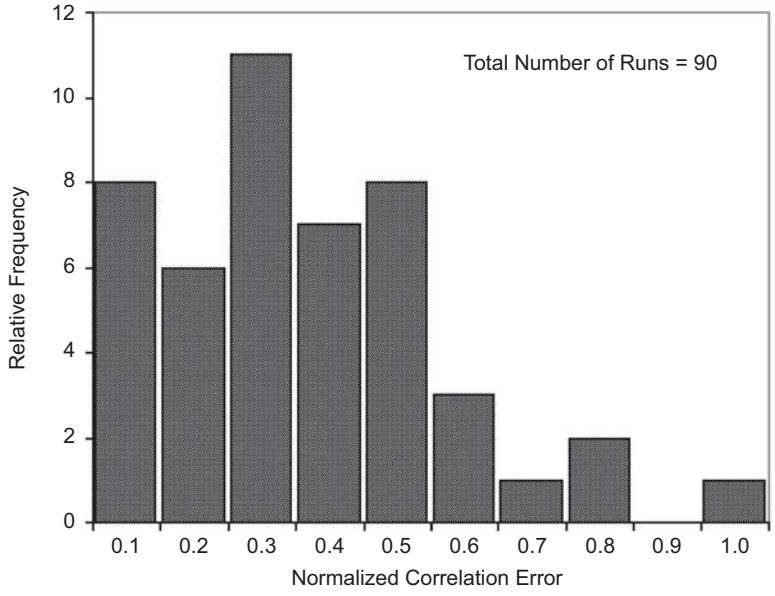


Fig. 5-33. Simulation results for the correlation tracker.

Both maximum-likelihood and edge-detection algorithms show potential for extended-source image acquisition. Initial results from the correlation algorithm do not reliably satisfy the 0.1 normalized error, but improvements in the algorithm using an iterative nonlinear estimator have shown an improved performance. Initial results from the edge-detection algorithm also do not satisfy the 0.1 normalized error, but the normalized error was steadily reduced as the algorithm was refined. For the simulated Earth images, the edge-detection method exhibits superior performance relative to the correlation algorithm. Future improvements in the correlation algorithm include better reference template generation, hard limiting the image, and addition of nonlinear estimating routines.

5.3.5.1.3 Subpixel Scanning Method. As one of the variants of the edge detection method, the concept of subpixel scanning was used [63,64]. The technique increases the resolution of the Earth image with subpixel scanning using FSM: a sequence of images is captured as the Earth image is moved across a CCD array in subpixel increments using FSM scanning. Using this high-resolution data, the algorithm then locates the geo-center of the Earth regardless of the illumination by the Sun. The algorithm relies on two parameters for its solution; cutoff frequency of the low-pass filter and a threshold used to extricate the Earth image from the background of the image. The optimal cutoff frequency of the low-pass filter is estimated using cross validation. For the threshold, a Monte Carlo simulation is used to determine the threshold that minimizes the mean squared error between the estimated and true locations of the ground-based terminal. Simulations indicate that the parameters can be determined very accurately using these methods. Numerical results of

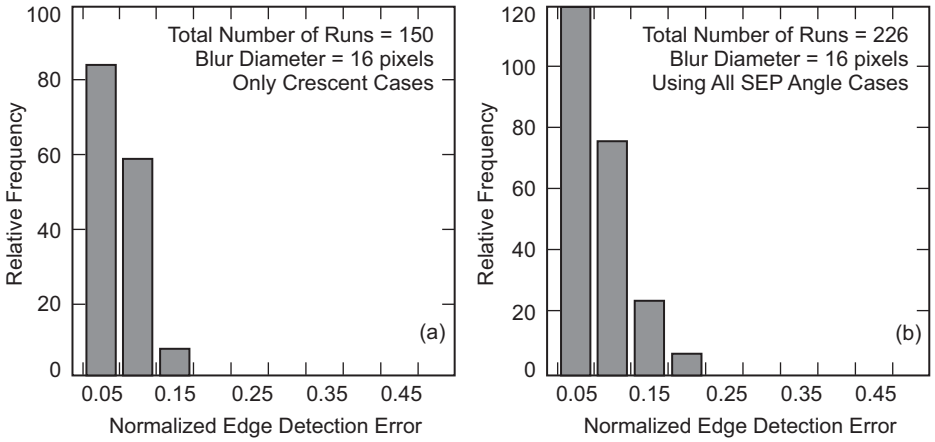


Fig. 5-34. Simulation results for the edge detection tracker for (a) large SEP angle (crescent Earth) cases and (b) for all SEP angle cases.

the acquisition algorithm, including parameter estimation, demonstrate that subpixel resolutions can be achieved regardless of the observed shape of the Earth or the presence of noise. Results on the centroid accuracy were achieved to within 1/10th and 1/20th of a pixel accuracy using partially to fully illuminated Earth images with an SNR of 1 to 20 used to test the algorithm [64].

5.3.5.2 Star Tracker. As stated earlier, the use of a star tracker for the ATP system is one of several options that can potentially solve the albedo variation problem of the visible Earth tracker [54,65]. The perceived potentials of this concept are twofold: a) range independent signal level and, b) point source that does not have an albedo-variation problem. Since the star tracker provides low-bandwidth signal, high-bandwidth gyros and angle sensors need to be combined for the high-bandwidth pointing knowledge estimation. The analysis which is presented later in Fig. 5-36, shows that pointing knowledge of 150 nrad (single axis, 1 sigma) can be achieved with an 8-cm diameter telescope aperture with assumptions of centroiding NEA of 1/25 pixel accuracy per star, Space Infrared Telescope Facility (SIRTF) class gyros (angle random walk, ARW = 0.0001 deg/root-hr), 5 Hz star trackers with ~5.0 degree FOV, detector of 1000 by 1000 pixels, and stars of roughly 9 to 9.5 magnitudes. This 150-nrad pointing knowledge is well below the typical deep-space optical communications requirements of about 300 nrad. The star coverage study shows that the average link availability is above 98 percent with a single star tracker.

The general approach is to determine both the attitude of the optical communications terminal (OCT) coordinate frame, and receiving station location in inertial space, such as relative to the J2000 coordinate frame (or the newer International Celestial Reference System (ICRS) frame), and point to a derived location in the optical communications frame. This approach is akin to that used to point science cameras on a Voyager-, Galileo-, or Cassini-type spacecraft. A specific approach to attain this attitude (or pointing) knowledge accuracy is to use a combination of high-precision star tracker measurements and high-bandwidth inertial sensor updates: gyros for medium bandwidth (< 50 Hz) and angle sensors for high bandwidth (> 50 Hz). An attitude estimator integrates this information with any additional data (such as position data from navigation (ephemeris), alignment data from in-flight calibrations, or data provided from communication with the ground) for the attitude knowledge estimation. Adding the point-ahead to the attitude knowledge to account for the two-way light time gives the pointing knowledge. The pointing knowledge is used to drive the fine-steering mirror to transmit the downlink laser beam.

5.3.5.2.1 The Pointing Knowledge Estimation Method. Star trackers are very accurate and provide accurate (absolute) pointing knowledge with typically a low update rate (depending on the star magnitude). On the other hand, inertial

sensors can provide high bandwidth (relative) pointing knowledge. Here the mathematical derivation is given on how to combine the star tracker and inertial sensor measurements to estimate the attitude of star tracker using star tracker and gyro measurements. Here we treat the attitude of star tracker equivalent to the pointing knowledge since the point-ahead, coordinate transformation from star tracker to optical communications terminal, and the receiver position do not affect the uncertainty of the final pointing knowledge. This method is based on an "averaging" technique, which mainly reduces the jitter of the pointing knowledge estimate. Angle-sensor measurements are to fill in between the two successive gyro measurements for higher update rate, and the total RMS error (or jitter) increases in a root-sum-squared (RSS) sense.

Assumptions for the pointing knowledge (attitude) jitter estimation procedure (after the initial acquisition of stars).

- A one-dimensional discrete time example is assumed. The estimation works the same way in two or three dimensions, except that it is mathematically more complex.
- The gyro is assumed to be high bandwidth and to have random noise and bias consistent with high-precision gyros. The "high bandwidth" is assumed to be high enough compared to the platform disturbances for sufficient disturbance rejection control.
- The equations shown depend primarily on the information rate, which is a measure of the SNR per unit time. Information rate is a method where, that to first order, we can compare the net effect of star trackers with different accuracies and update rates, such as noting that 100 measurements/second with $\sigma^2 = 100$ are the same as 1 measurement with accuracy $\sigma^2 = 1$ (= 100/100).
- The star tracker is at lower frequency. In this example, a tracker measurement is made at every *N* steps of the gyro. The tracker has random noise s_k with variance (assumed as a constant in this case), σ_S^2, and the star tracker measurement propagated to the correct time for inclusion in the estimate (and that the σ_S^2 includes any additional noise due to propagation, such as gyro bias contributions).
- In this example, the initial estimate of the attitude, x_0, with variance, $\sigma_{x_0}^2$, is assumed to be based on the star tracker measurement.
- The gyro angle measurement is g_k. We assume there is an estimation state (not shown) for bias, and the measurement is compensated for the estimated bias;
  - The random error (angle random walk, assumed to contain a priori uncertainty due to gyro bias) w_k, with variance σ_w^2 (degrees^2/hr)

- Bias error b_k (degrees/hour), with mean value b over the short interval in question.
- The only bias error contribution is due to the residual bias error.

- Assume uniform time spacing Δt seconds between successive gyro measurement updates, and uncorrelated error sources; with N measurements of the gyro between every star tracker update. The star tracker measurements are then $T = N\Delta t$ seconds apart.
- x_k will be the pointing estimate after k steps for the gyro.

Basic equations

To estimate the attitude of the one dimensional telescope boresight, x_k

$E[w_k] = 0$	white noise
$\theta_0 = x_0$	initial state
$\theta_{k+1} = \theta_k + (g_k + w_k)$	gyro based state equation

and the change in error which includes residual bias error,

$$e_{k+1} = e_k + w_k + \Delta t \times b_k \tag{5.3-9}$$

and the total error becomes (with bias and noise)

$$E\left[(e_{k+1})^2\right] = \Delta t\, \sigma_w^2 + (\Delta t\, b_k)^2 + E\left[(e_k)^2\right] \text{ (degrees}^2\text{)} \tag{5.3-10}$$

Then for N steps, the total error becomes

$$E\left[(e_{k+1})^2\right] = N\Delta t\, \sigma_w^2 + (N\Delta t\, b_k)^2 + E\left[(e_k)^2\right] \tag{5.3-11}$$

The variance (after taking out the mean error, $(N\Delta t\, b_k)^2$) is

$$\text{Var}[e_k + N] = \text{Var}[e_k] + N\Delta t\, \sigma_w^2 \tag{5.3-12}$$

The increase in the total error is approximately $N\Delta t\, \sigma_w^2$ (jitter) + $(N\Delta t\, b)^2$ (bias) between measurement intervals of N and $N + k$.

Without any star tracker measurements, the estimated attitude angle is

$$x_k = \theta_k, \quad \text{Var}[x_k] = \text{Var}[e_k] \tag{5.3-13}$$

After a star tracker measurement, using standard type Kalman/least squares weighting to incorporate the star tracker measurement, and with a defined by

$$a = \frac{\sigma_S^2}{\sigma_S^2 + \mathrm{Var}\left[e_k\right]} \tag{5.3-14}$$

the update is (note the superscript +) made by computing

$$x_k^+ = \left(S_k\right)\left(1 - a\right) + a \times x_k \tag{5.3-15}$$

$$\mathrm{Var}\left[x_k^+\right] = \left(1 - a\right)^2 \sigma_S^2 + a^2 \mathrm{Var}\left[e_k\right] = \frac{\mathrm{Var}\left[e_k\right]\sigma_S^2}{\sigma_S^2 + \mathrm{Var}\left[e_k\right]} \tag{5.3-16}$$

Then, for $k = 0$ and time $t = 0$, the total error variance is σ_S^2 with $T = N\Delta T$.

For $k = N$, before the next tracker update,

$$\mathrm{Var}\left[x_N\right] = \mathrm{Var}\left[x_0\right] + T\,\sigma_w^2 = \sigma_S^2 + T\,\sigma_w^2. \tag{5.3-17}$$

And the corresponding bias error is $T\,b$.

After the update with a star tracker measurement, substituting $\mathrm{Var}\left[x_N\right]$ in Eq. (5.3-17) for $\mathrm{Var}\left[e_k\right]$ in Eq. (5.3-16) gives the new variance,

$$\mathrm{Var}\left[x_N^+\right] = \frac{\sigma_S^2\left(T\,\sigma_w^2 + \sigma_S^2\right)}{2\sigma_S^2 + T\,\sigma_w^2} \tag{5.3-18}$$

For accumulated bias error, substituting $\mathrm{Var}\left[x_N\right]$ in Eq. (5.3-17) for $\mathrm{Var}\left[e_k\right]$ in denominator of Eq. (5.3-16) and $T\,b$ for $\mathrm{Var}\left[e_k\right]$ in numerator of Eq. (5.3-15) gives the new estimate,

$$\frac{\sigma_S^2\,T\,b}{2\sigma_S^2 + T\,\sigma_w^2} \tag{5.3-19}$$

The derivation process for Eq. (5.3-18) is graphically shown in Fig. 5-35.

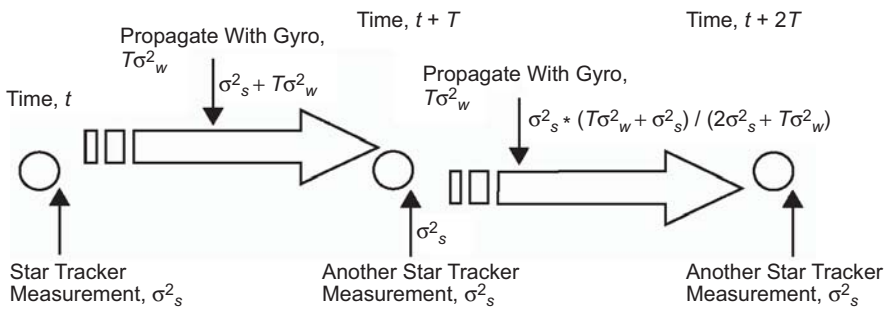


Fig. 5-35. Iterative averaging process for the pointing knowledge estimation variance.

At time, $t+T$, the new estimate is average of the star tracker measurement at time, *t*, (which has 1-sigma error of σ_S^2) propagated using gyro angle measurements (the propagated star tracker measurement taken at time *t* has a 1-sigma variance of $\sigma_S^2 + T\,\sigma_w^2$ at time $t+T$) and the new star tracker measurement at time, $t+T$ (1-sigma variance of σ_S^2). The equally weighted average of the attitude estimate with gyro propagated, x_N, and star tracker update, x_S, is

$$x_N^+ = \frac{x_N + x_S}{2}$$

Under those assumptions, the variance is

$$\begin{aligned}\mathrm{Var}\left(x_N^+\right) &= \frac{\mathrm{Var}\left(x_N\right) + \mathrm{Var}\left(x_S\right)}{4} \\ &= \frac{\sigma_S^2 + T\,\sigma_w^2 + \sigma_S^2}{4} \\ &= \frac{2\sigma_S^2 + T\,\sigma_w^2}{4}\end{aligned}$$

However, a more reasonable assumption is to give more weight to the estimate with smaller variance. One method to perform the weighted averaging is, noting that

$$
\begin{aligned}
x_N^+ &= \frac{\mathrm{Var}(x_N)x_S}{\mathrm{Var}(x_N)+\mathrm{Var}(x_S)} + \frac{\mathrm{Var}(x_S)x_N}{\mathrm{Var}(x_N)+\mathrm{Var}(x_S)} \\
&= \frac{\left(T\,\sigma_w^2+\sigma_S^2\right)x_S}{\left(T\,\sigma_w^2+2\sigma_S^2\right)} + \frac{\left(\sigma_S^2\right)x_N}{\left(T\,\sigma_w^2+2\sigma_S^2\right)}, \text{ then} \\
\mathrm{Var}\left(x_N^+\right) &= \frac{\left(T\,\sigma_w^2+\sigma_S^2\right)^2\sigma_S^2}{\left(T\,\sigma_w^2+2\sigma_S^2\right)^2} + \frac{\left(\sigma_S^2\right)^2\left(T\,\sigma_w^2+\sigma_S^2\right)}{\left(T\,\sigma_w^2+2\sigma_S^2\right)^2} \\
&= \frac{\sigma_S^2\left(T\,\sigma_w^2+\sigma_S^2\right)}{2\sigma_S^2+T\,\sigma_w^2}
\end{aligned}
\tag{5.3-20}
$$

For a more general case where the attitude estimate at time, *t*, is star tracker measurement propagated with gyro measurement, let the attitude estimate be x_u (with the corresponding 1-sigma variance σ_U^2), then the variance of x_N^+

$$
\mathrm{Var}\left[x_N^+\right] = \frac{\left(T\,\sigma_w^2+\sigma_U^2\right)^2\sigma_S^2}{\left(T\,\sigma_w^2+\sigma_U^2+\sigma_S^2\right)^2} + \frac{\left(\sigma_S^2\right)^2\left(T\,\sigma_w^2+\sigma_U^2\right)}{\left(T\,\sigma_w^2+\sigma_U^2+\sigma_S^2\right)^2} \tag{5.3-21}
$$

After additional *N* gyro measurements, the error variance grows to

$$
\mathrm{Var}\left[x_{2N}\right] = \mathrm{Var}\left[x_N^+\right] + T\,\sigma_w^2 \tag{5.3-22}
$$

Equation (5.3-21) is plotted in Fig. 5-36 for the Gyro ARW of 0.0001 deg/rt-hr and three star tracker NEAs of 0.7 μrad/frame, 1.0 μrad/frame, and 2.0 μrad/frame, respectively. Notice that the combined jitter of a star tracker and a gyro of smaller than 150 nrad can be obtained after averaging of 5 s.

The 0.7-μrad NEA can be achieved with 25 stars of visual magnitude of 9 or brighter. Centroiding NEA for each star is assumed as 1/25th pixel. 1/25th pixel NEA can be achieved with roughly 3500 e/frame using the centroid window of 5x5 pixels for a read noise of 10 e^- and 3500e/frame for a 5-Hz star tracker is feasible with 8-cm aperture with 5-deg FOV [54].

5.3.5.2.2 Star Tracker Configuration Trades. In this section, four star tracker configurations are discussed in terms of attitude estimation accuracy and star coverage. Depending on the requirements on the accuracy and the link availability, one can select among these options.

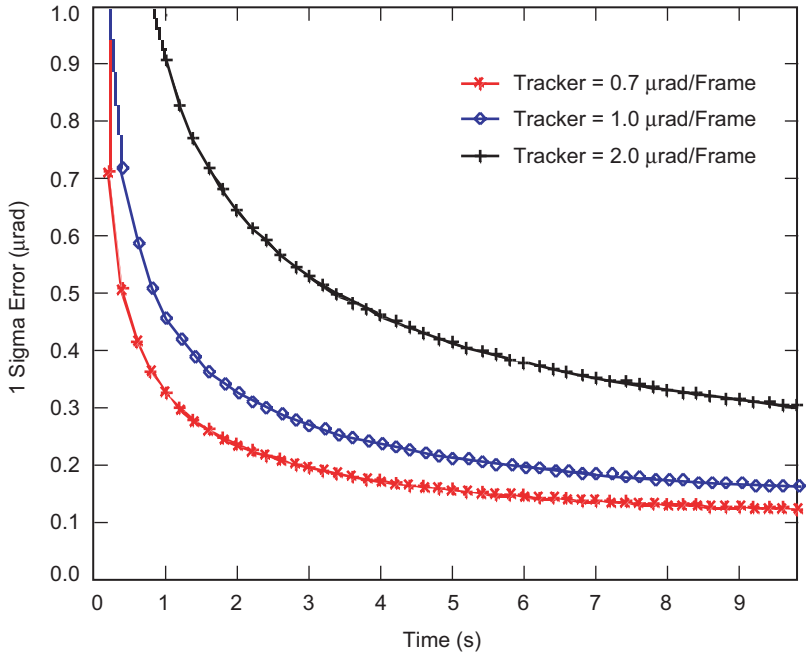


Fig. 5-36. Star tracker-gyro combined jitter as a function of integration time for the ARW = 0.0001 deg/root-hr.

The single star tracker approach (antipodal or boresighted configuration). A single star tracker gives two good attitude estimates (around the x axis and the y axis) and one poor attitude estimate (the twist around the boresight) because of the lack of star separation from the center of FOV. Typically for narrow-angle star trackers, the attitude estimates on the twist around the boresight is about 10 to 20 times worse than that of the other axes. Therefore, the star tracker orientation relative to the optical comm. terminal is critical. To take advantage of this fact, the star tracker is aligned along the telescope's optical axis either facing the Earth receiver (bore-sighted) or 180 deg away from it, facing the opposite direction (antipodal). Due to the Sun's stray light issue, antipodal is the preferred configuration. One potential issue with the single star-tracker approach is the relatively low star availability due to the limited star-search area. However, this availability heavily depends on the actual mission profile. As shown later in link availability analysis, the average sky coverage with a single star tracker can be better than 98 percent.

The two (or multi-FOV) star tracker approach (normal or gimbaled configuration). Since the boresight-twist estimation is large, a common approach is to use multiple star trackers. Some implementations use separate independent trackers. There are other designs that use mechanically integrated trackers, and others that use multiple FOVs (such as the mini-owls) on a single

detector. Another advantage of the two-star-tracker approach is the enhanced star coverage compared to the single star tracker. Table 5-14 summarizes the four configuration options with the pros and cons of each configuration.

5.3.5.2.3 Link Availability Analysis. The link availability, which directly depends on the star coverage, is one of the most critical mission parameters that affect the success of the mission. Previously, a link availability of 98 percent was assumed for a single star tracker with a FOV of 5 deg × 5 deg. The assumption on the visual magnitude of star is 9 to 9.5. The data analysis presented here is based on star position and visual magnitude data that were extracted from the Tycho II star catalog. A computer program was constructed that centered a circular FOV diameter of 3.5, 4.0, 4.5, 5.0, 5.5, and 6 deg at each point of an RA (α) and declination (δ) grid defined on the celestial sphere. The spacing of the grid was 0.1 deg in both RA and declination. For each combination of positions (α, δ), FOV size, and cutoff magnitude, the number of stars found within the FOV was computed. Finally, as a function of FOV size,

Table 5-14. Pros and cons of the four configuration options for the star-tracker-based ATP system.

Star Tracker Orientation (relative to Optical Comm Telescope)	Pros	Cons
Antipodal (single star tracker)	Looks away from Earth, more stars without Earth in FOV. Generally no Sun problems for outer planets, lower twist error contribution to pointing. Best pointing direction in the telescope direction.	Places more restrictions on mounting, requiring unrestricted viewing area, especially if mounted on the same platform as optcomm terminal. Requires accurate alignment with telescope pointing optics and need to develop calibration procedure.
Boresighted (single star tracker)	Shares channel, lower twist error contribution to pointing. Common optics relaxes alignment requirements.	Earth blocks out stars in partial field of view. Sun increases background and stray light so that long baffle may be required.
Normal (two star trackers)	Can be rotated with spacecraft for greatest star coverage field of regard.	Large error in one of LOS axes (can be reduced with two star trackers) due to large star-tracker twist error
Gimbaled (two star trackers)	Can be pointed and slewed to particular star or celestial body.	Requires gimbal mechanism, twist error about star tracker boresight may cause large pointing errors as boresight moves away from antipodal pointing. Knowledge of gimbaled position introduces additional error source.

the fraction of the sky where the specified number of stars can be obtained was computed as a function of visual magnitude. Tables 5-15 and 5-16 thus provide a snapshot for the star magnitude cutoff needed to achieve 98-percent sky coverage and predicted performance. In Table 5-15, for the given FOV and the number of stars, the star magnitude cutoff was computed such that the average sky coverage is 98 percent. In Table 5-16, the corresponding centroiding NEA was computed using the results of Table 5-15 assuming 1000 pixels across the detector and 1/25 pixel random error /star /axis attached to each measurement. The performance is reported in microradians 1σ, per axis, worst case RMS for 98 percent of the sky. For example, there are 25 stars of magnitude 9 for an FOV of 5 deg. The NEA using 25 stars is computed using the assumption of a single star NEA of $1/25^{th}$ pixel. Due to averaging of 25 centroids, the centroid NEA using 25 stars becomes

$$\text{NEA using 25 stars} = \frac{\text{NEA of single star}}{\sqrt{\text{number of stars}}}$$

$$= 85\ \mu\text{rad (pixel FOV from 5 deg and } 1000 \times 1000) \times \frac{1}{25} \times \frac{1}{\sqrt{25}}$$

$$= 0.7\ \mu\text{rad}$$

Table 5-15. Star magnitude (Mv) cutoff required to achieve 98 percent of sky coverage for various star counts and FOV sizes.

Count	FOV (deg)					
	3.5	4.0	4.5	5.0	5.5	6.0
5	8.5	8.0	8.0	8.0	7.5	7.5
10	9.0	8.5	8.5	8.5	8.0	8.0
16	9.5	9.0	9.0	8.5	8.5	8.5
25	9.5	9.5	9.5	9.0	9.0	9.0
36	10.0	10.0	9.5	9.5	9.5	9.0
49	10.5	10.0	10.0	9.5	9.5	9.5
64	11.0	10.5	10.0	10.0	10.0	9.5
81	11.0	10.5	10.5	10.0	10.0	10.0
100	11.5	11.0	10.5	10.5	10.5	10.0
121	11.5	11.0	11.0	10.5	10.5	10.5
144	11.5	11.5	11.0	11.0	10.5	10.5
169	***	11.5	11.5	11.0	11.0	10.5
200	***	***	11.5	11.5	11.0	11.0

Table 5-16. Performance estimate based on Table 5-15 and having 1:1 correspondence with Table 5-15.

Count	FOV (deg)					
	3.5	4.0	4.5	5.0	5.5	6.0
5	1.09	1.25	1.40	1.56	1.72	1.87
10	0.77	0.88	0.99	1.10	1.21	1.32
16	0.61	0.70	0.79	0.87	0.96	1.05
25	0.49	0.56	0.63	0.70	0.77	0.84
36	0.41	0.47	0.52	0.58	0.64	0.70
49	0.35	0.40	0.45	0.50	0.55	0.60
64	0.31	0.35	0.39	0.44	0.48	0.52
81	0.27	0.31	0.35	0.39	0.43	0.47
100	0.24	0.28	0.31	0.35	0.38	0.42
121	0.22	0.25	0.29	0.32	0.35	0.38
144	0.20	0.23	0.26	0.29	0.32	0.35
169	0.19	0.21	0.24	0.27	0.30	0.32
200	0.17	0.20	0.22	0.25	0.27	0.30

In summary, the star-tracker-based ATP approach can provide high accuracy on the order of 150 nrad (1-sigma, single axis) in pointing knowledge estimation with reasonable assumptions on the aperture size and star visual magnitude. The average link availability is estimated at over 98 percent using the FOV of 5 deg. As for the implementation, misalignment error between the optical communications terminal and the star tracker can be a major issue if they are not co-aligned, which may require sophisticated calibration.

5.3.5.3 Earth Tracker—Long Wavelength Infrared Band. Earth image tracking in the long wavelength infrared (LWIR) band is another option that can mitigate the albedo variation problem of the visible Earth image tracker for its low emissivity variation [66]. Low emissivity variations of thermal images is due to the relatively slow thermal changes of the Earth surfaces compared with rapid changes of reflectivity of the Earth surface for the visible wavelength.

Additionally, a full Earth image can be maintained even for high phase angles with the thermal imaging, which gives higher centroiding accuracy. The recent release from the Mars Odyssey program shows that the entire (full) Earth thermal image was successfully taken, whereas the visible-light image shows the thin crescent Earth viewed from Odyssey's perspective (Fig. 5-37, [67]). These images, taken at a distance of 3,563,735 km on April 19, 2001, as the

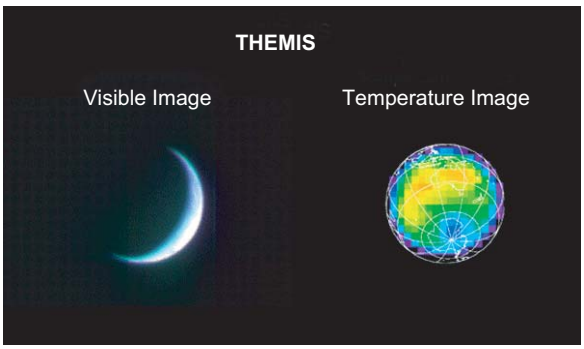


Fig. 5-37. Visible Earth image vs. thermal image.

Mars Odyssey spacecraft left Earth, support the potential of Earth thermal images in tracking/pointing applications.

5.3.5.3.1 Wavelength Selection. The advantage of using LWIR Earth images is the ability of the Earth-tracking sensor to observe energy emission from the Earth rather than reflected solar energy. To take a full advantage of steady Earth thermal emission, the optimum spectral band needs to be selected. The optimum spectral band involves a trade-off between the need for more signals to overcome the detector noise and the pointing bias introduced by thermal variations across the Earth's surface. The spectral signal variations for black bodies of different temperatures vary much more at the shorter wavelengths, so the longer wavelengths are preferred to minimize any bias error resulting from this effect. An additional advantage in the use of longer-wavelength radiation is attained from the reduction of direct solar background and solar scattering from the Earth's surface. The ability of the LWIR tracking system to generate an accurate Earth centroid value is related to the signal-to-noise ratio (SNR) in the sensor used to image the Earth. If the noise level on each pixel is not much lower than the signal value, the effect will be a shift of the final centroid value from the true value. Several different types of noise exist in detectors, and the amount of each varies from detector to detector. The aggregate noise on detectors tends to grow more slowly with integration time than does the accumulated signal, affording some potential to improve the SNR by operating the sensor at lower speeds, but this approach reduces the Earth centroid update rate.

The approach for selecting an optimum spectral band is to start at the 13-μm end of the 8- to 13-μm waveband and integrate the signal down to the spectral point at which the improvement to centroid estimation from higher signal-to-noise ratios is offset by the centroid shift inherent in the use of shorter wavelengths. The rationale for using the 8- to 13-μm waveband is its high atmospheric transmission and higher black-body radiation [68]. The available

radiance for this approach is shown in Fig. 5-38. The roll-off of the curve at shorter wavelengths further demonstrates the lack of value in extending the spectral band to shorter wavelengths. At shorter wavelengths, the variation of the radiance is larger over the entire phase angle. This radiance variation directly affects centroid bias. Based on this result, three spectral bands were selected for comparison. The 3- to 5-μm transmission window of Earth was evaluated and appears to be a rather poor choice. Only a very small fraction of the thermal energy is emitted in this band, and because of the short wavelengths, that means even fewer photons. Additionally, the blackbody response also shows that the background from scattered solar radiation will be much, much higher than would be found around 10 μm. Finally, since Rayleigh photon scattering is proportional to λ^{-3}, there will be over 20 times more Rayleigh background in the 3- to 5-μm band than around 10 μm. The 10- to 13-μm band looks quite good. There is a strong signal integrated over a full 3-μm band. Increasing the width of the band to 8–13 μm increases the signal strength by about 2 dB, according to this model. However, the model does not take into account the strong atmospheric absorption of the CO_2 band that reduces the signal gain rather significantly. Also, there is more diurnal variation in the 8- to 13-μm signal and, consequently, more spectral shift of the Earth centroid. Finally, for the 8- to 13-μm waveband, there is about 25 percent more solar-induced background than there is in the 10- to 13-μm waveband, resulting in a much greater reduction in the background level than in the signal level by limiting the band.

In summary, the 8–13-μm band was selected as the optimum band after several trades, and the subsequent analysis is based on this band.

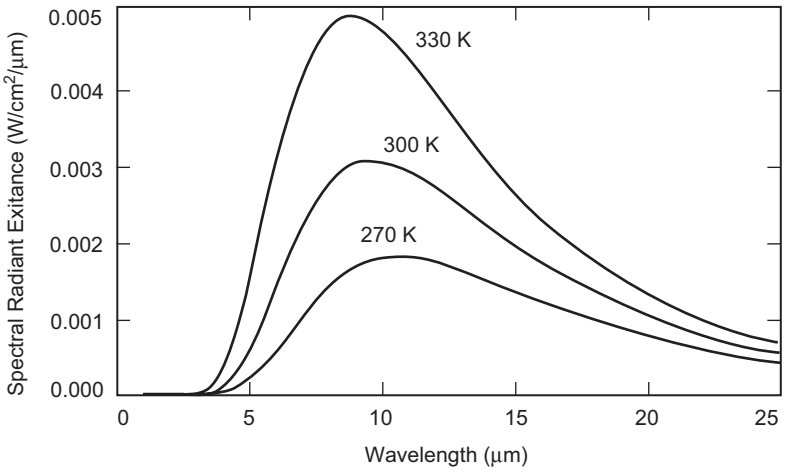


Fig. 5-38. Spectral radiance as a function of waveform for three temperatures: 270 K, 300 K, and 330 K.

5.3.5.3.2 Estimated Signal Level. A simple computer model was generated to estimate the effects of thermal variations across the surface of the Earth. The model calculates a very crude surface temperature profile based on expected diurnal, seasonal, and latitudinal variations. The model generates smoothly varying temperature profiles, without the steep gradients expected from weather patterns or surface topography. To obtain an available signal level, the model integrates the spectral radiance in the specified waveband for a selection of spatial points across the visible Earth surface. Multiplying by the known area of the Earth's surface yielded the radiant intensity of the Earth (photons/s/sr) (Fig. 5-39). Since the temperature of the visible portion of the Earth varies by season and by aspect angle (Sun–Earth–probe (SEP) angle), these values were plotted in Fig. 5-40 as a function of hour (of local time) and for solstice versus equinox.

5.3.5.3.3 Centroid Jitter Estimate. The photon radiant intensity values in Fig. 5-39, with the assumed parameter values in Table 5-17, were used to estimate the centroid jitter or noise equivalent angle (NEA). For simplicity, the center of brightness centroid is discussed here. Since the Earth image is a relatively uniform disk, it is assumed that each pixel in the sum experiences the same noise due to sensor-plus-signal photon statistics. All quantities are assumed to be measured in electrons. *DT* is the integration time used for dark current noise, and *DC* is the dark current rate in electrons/second. The basic center-of-brightness centroiding formula for a uniform signal (pixel value $D_{i,j}$ after background compensation) with a variance formula (assuming very large total signal, *T*, to take it outside the expectation) given by variance S_x (S_y is similarly computed) is approximately given by

$$\left.\begin{aligned} T &= \sum_{j=1}^{M}\sum_{i=1}^{N} D_{i,j} \\ S_x &= \sum_{j=1}^{M}\sum_{i=1}^{N} i \times D_{i,j} \\ S_y &= \sum_{j=1}^{M}\sum_{i=1}^{N} j \times D_{i,j} \end{aligned}\right\} \qquad (5.3\text{-}23)$$

Then

$$\left.\begin{aligned} x_e &= \frac{S_x}{T} \\ y_e &= \frac{S_y}{T} \end{aligned}\right\} \qquad (5.3\text{-}24)$$

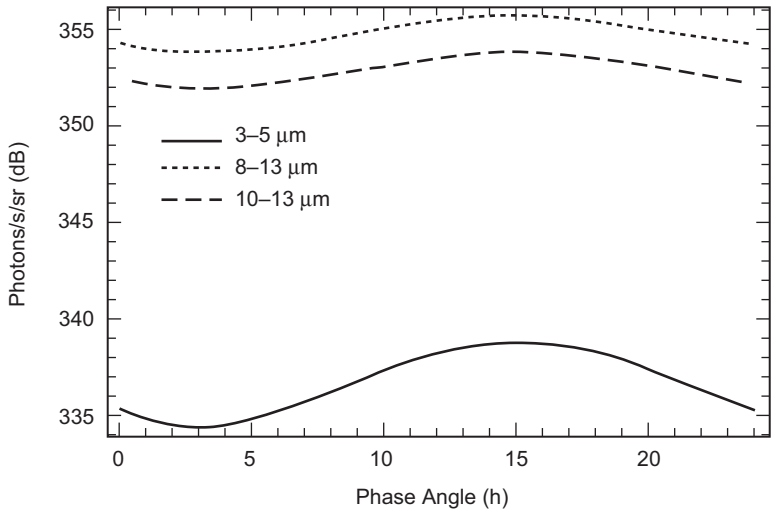


Fig. 5-39. Spectral signal for three wavelength bands.

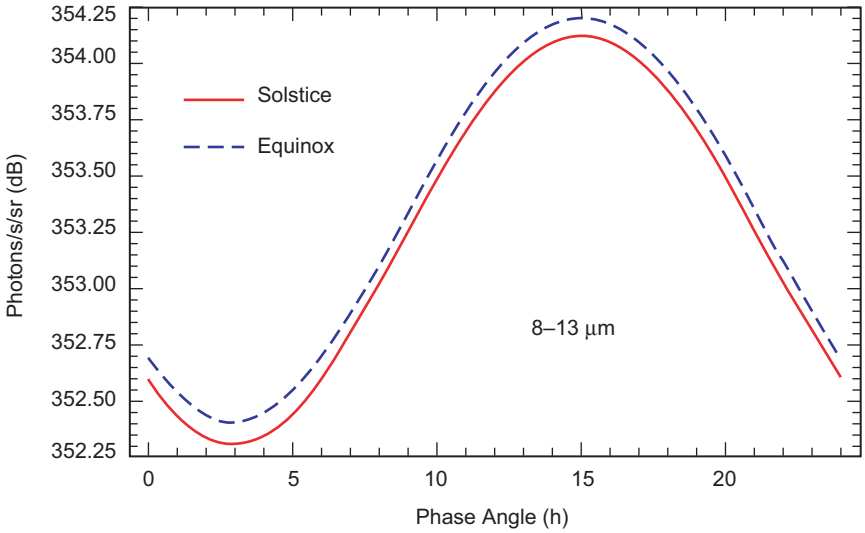


Fig. 5-40. Earth thermal signal level for solstice vs. equinox.

Table 5-17. Assumptions for NEA estimations.[a]

Parameter	Value	Rationale
Centroid window size	9 × 9 pixels to 25 × 25 pixels	Earth image of 60–220 μrad, 10 μrad/pixel, extra 3 pixels
Focal plane array (FPA) full well	20×10^6 to 30×10^6 electrons	Specification of DRS Technologies IR FPA of 25- to 40-μm pixels
ADC effective bits	14 bits	14 bits
System noise (1 sigma)	200 electrons	Read noise + electronic noise + background noise
Frame update rate	10 Hz to 1 kHz	10 Hz (inertial sensor assisted) to 1 kHz (optical only)
Aperture size	10 cm to 30 cm	Previous baseline for optical communication for deep space
Range	0.5 AU to 2.7 AU	0.5 AU to 2.7 AU (Mars missions)
Optical transmission efficiency + detector QE	10 percent	Detector QE of 80 percent and optical transmission of 13 percent

[a]Underlined values were used for the simulations presented in Figs. 5-41 and 5-42.

$$\left.\begin{aligned}\mathrm{Var}(S_x) &= \frac{N}{2} \times \frac{N/2+1}{3} \times \frac{T + N \times M \times (\mathrm{Var}(\mathrm{Read\ Noise}e) + DT(DCe))}{T^2} \\ \mathrm{Var}(S_y) &= \frac{N}{2} \times \frac{N/2+1}{3} \times \frac{T + N \times M \times (\mathrm{Var}(\mathrm{Read\ Noise}e) + DT(DCe))}{T^2}\end{aligned}\right\} \quad (5.3\text{-}25)$$

Assuming $N = M$, and that the dark current noise is negligible, the variance simplifies to

$$\frac{N}{2} \times \frac{N/2+1}{3} \times \frac{T + N^2(\mathrm{Var}(\mathrm{Read\ Noise}e))}{T^2}$$

There are two classes of parameters: one consists of design values such as aperture size and detector full well; the other is mission-dependent parameters such as range and centroid window size (governed by beacon spot size). Consider the two tracking scenarios, optical-only tracking and inertial-sensor-assisted tracking. The beacon update rates of 10 and 1 kHz were used, respectively. For inertial-sensor-assisted tracking, the NEA is very small—on the order of 10 nrad (10 μrad/pixel) for 8- to 13-μm bands (Fig. 5-41). For the

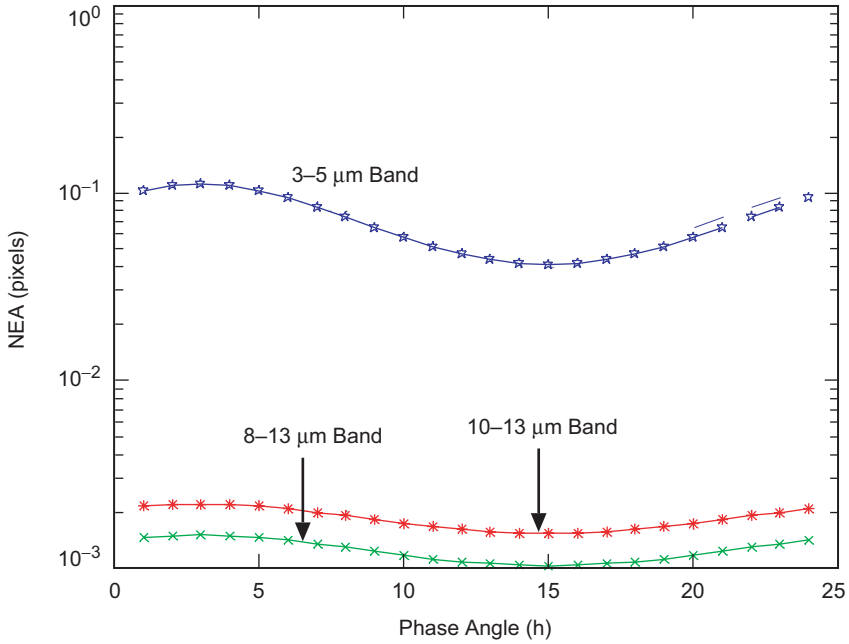


Fig. 5-41. NEA for inertial-sensor-assisted tracking update rate (10 Hz) using the worst-case scenario outlined in Table 5-17.

3- to 5-µm band, the NEA is up to 1 µrad, and with trade-offs on the detector full well and aperture size, this can be reduced to 70 nrad. For optical tracking only, the worst-case estimate of the NEA is more than 1 µrad. However, an NEA of better than 100 nrad can be achieved with the trade-offs on the smaller detector full well and larger aperture size (Fig. 5-42).

5.3.5.3.4 Centroid Bias Error Estimate. For bias estimation, the edge-detection-based centroiding can take full advantage of the Earth thermal image with all the edge pixels detected. An estimate of the bias error can be derived based on the single-edge pixel detection error and the number of edge pixels. The key concept is that the bias error decreases as more edge pixels are averaged to obtain the centroids. A *d*-pixel-diameter image will have a πd pixel edge length; for each axis, the accuracy will range, depending effectively on the half of the total number of edge pixels, $\pi d/2$ (using the side of square pixel) to $(\pi d/2)/\sqrt{2}$ (using the diagonal of a square pixel) edge pixels, where the majority of the information for a particular axis comes from edges with intensity gradients with large components in that axial direction. This gives an expected rms accuracy of

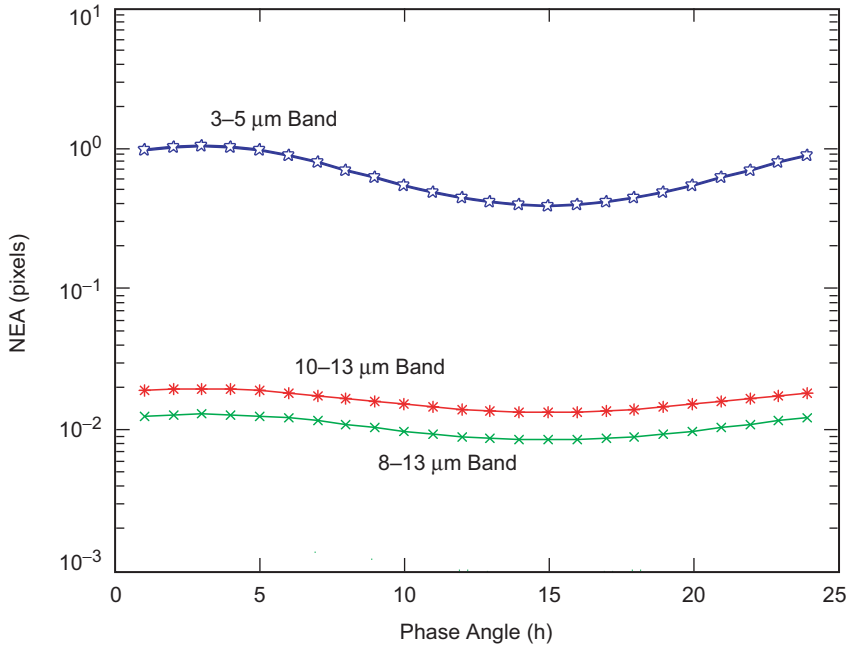


Fig. 5-42. NEA for optical tracking update rate (1 kHz).

$$\frac{\sigma}{\sqrt{\frac{\pi d}{2}}} \text{ to } \frac{\sigma}{\sqrt{\frac{\pi d}{2\sqrt{2}}}} \text{ pixel} \qquad (5.3\text{-}26)$$

where σ is the 1-sigma bias error in edge detection.

Assuming 10-μrad pixels and the range of 2.7 AU, the above estimation gives about 400- to 474-nrad bias error (1 sigma) with $\sigma = 0.1$ pixel. Bias terms arise because of non-uniformity of the surface emission near the edge of the Earth and spatial quantization (sampling) of the thermal image. To reduce the bias error, the number of edge pixels needs to be increased with smaller pixel FOV, as the Eq. (5.3-26) indicates.

In summary, the analysis on LWIR Earth image trackers indicates that the pointing knowledge jitter of less than 150 nrad for update rate of 1 kHz assuming the range of 2.7 AU is feasible. For larger ranges, the update rate decreases for the equal pointing-knowledge jitter. For example, the signal will be 100 times dimmer at 27 AU. If the update rate is reduced to 10 Hz, then the integrated total signal would be identical by integrating 100 times longer time. In this respect, the performance of the LWIR Earth image tracker is equivalent to that of star tracker for the Solar System range. For the bias error, the star tracker is independent of the range while the LWIR Earth image tracker

depends on the Earth size on the FOV of the FPA such that the pixel FOV needs to be specifically designed to meet the bias error requirement.

5.3.6 ATP Technology Demonstrations

5.3.6.1 Reduced Complexity ATP Architecture. A CCD-based spatial acquisition, tracking, and pointing subsystem architecture has been developed to perform both spatial acquisition and tracking functions for a lasercom terminal [69]. The focal plane array detector can achieve both the wide field of view required for spatial acquisition and the high update rate needed for effective platform jitter compensation by operating the CCD in the "windowed" read mode. Furthermore, this spatial tracking subsystem based on the CCD tracker requires only one steering mirror to perform both line-of-sight stabilization and point-ahead functions, and to provide means to optically close the point-ahead control loop without additional sensors. When incorporated into the lasercom system designs, the array tracking concept leads to a reduced complexity system and hence a lower system cost.

Previous designs of lasercom systems generally achieved the desired pointing accuracy by using a directionally sensitive detector (such as a quadrant avalanche photodiode) to measure the angular error between the detector line-of-sight and the beacon direction [70,71]. The error was then fed back to a high-bandwidth steering mirror to stabilize the detector line-of-sight along the beacon direction. A second point-ahead mirror in the transmit beam path was then used to provide the required pointing offset between the transmit and receive signals. Since the quadrant detector had a limited FOV, a separate, larger format detector was usually required to provide the wide field of view coverage during the acquisition process. Furthermore, in order to properly relay the optical signals between the steering mirrors and the detector focal planes, additional optical relay elements were required that further increased the system complexity. However, with the advances in array detector technology, conceptual simplification of the lasercom tracking and acquisition subsystem can be achieved without sacrificing the system performance. This is because a pixilated detector can provide wider FOV coverage, thus simplifying the spatial acquisition process. A wide-FOV system can also permit tracking of the receiver beacon off axis, thus permitting the system to be implemented with only one steering mirror.

A conceptual block diagram of the array-based tracking system is shown in Fig. 5-43. A remote beacon laser is imaged by the telescope optics onto the focal plane array. By reading out the area of the detector containing the beacon signal and calculating the image centroid, the angular direction of the beacon can be accurately deduced relative to the optical axis of the system. A small amount of the transmit signal can also be imaged onto the acquisition detector and the location of the transmit signal can be measured relative to the optical

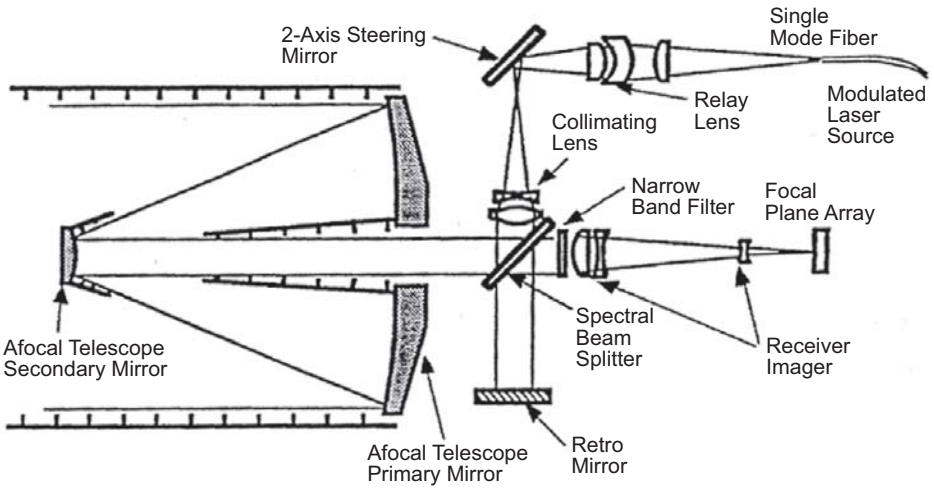


Fig. 5-43. Reduced complexity two-spot spatial tracking using a single focal plane array detector.

axis. The distance between the two image spots in the focal plane is a direct measure of the relative angular offset between the transmit and beacon signals. By sensing any difference between this measured, instantaneous point-ahead angle and the desired point-ahead value, the instrument can derive a real-time control signal to maintain pointing of the transmit signal. The subsequent beam-steering control can be achieved using a tandem of high-bandwidth steering mirror and wide-dynamic-range spacecraft. Large amplitude disturbances (such as the dead-band cycle of the spacecraft) are first removed using spacecraft attitude control. A fast steering mirror in the optical path is then used to compensate for the high-frequency, small-amplitude disturbances (micro-vibrations). The spacecraft removes the bias and maintains the steering mirror at the middle of its dynamic range. During the initial acquisition, the spacecraft is also used to orient the instrument line of sight for acquisition.

Shown in Fig. 5-44 is a block diagram of the two-spot tracking control loop. The detector images both the beacon signal and a portion of the transmit signal. The output digital data are then relayed to a control processor, which computes the positions of the image centroids and hence the instantaneous point-ahead angle. This point-ahead angle is then compared to a reference point-ahead angle, and the difference is fed into a compensation filter, which calculates the control needed for the fine-steering mirror. At the same time, the position of the beacon signal is fed to the gimbal control circuit, which stabilizes the position of the gimbal spot on the focal plane.

This simplified ATP architecture was implemented and demonstrated in the Optical Communication Demonstrator [72], which was verified in laboratory [73] and field experiments [74]. It was also implemented (in its second

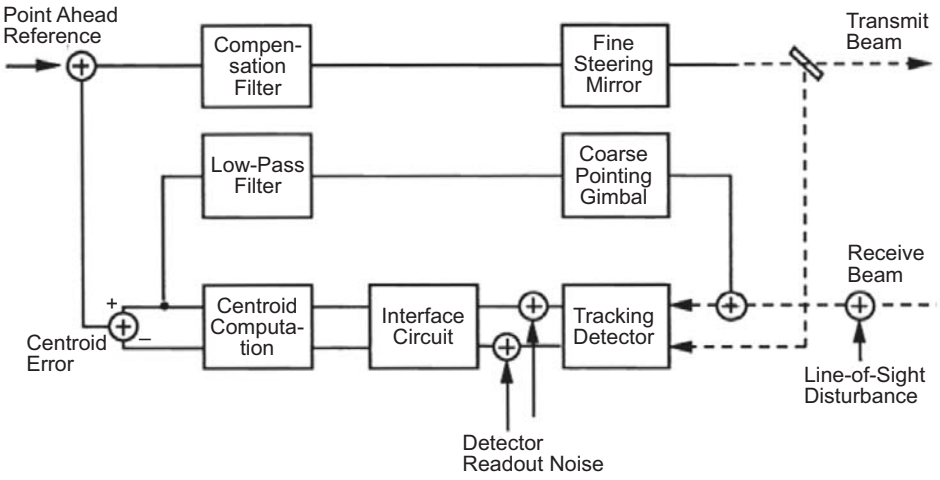


Fig. 5-44. Block diagram of a CCD-based tracking-control loop.

generation) for an Altair unmanned aerial vehicle (UAV)-to-ground 2.5 gigabits per second (Gbps) optical communications experiment as the optical communication terminal [75,76].

5.3.6.2 Centroiding Algorithms—Spot Model Method. Accurate centroid estimation is critical for free-space optical communications where the number of photons from the reference optical sources (such as stars or an uplink beacon) is limited. It is known that the centroid accuracy is proportional to the SNR. The presence of various noise sources during the exposure of the CCD can lead to significant degradation of the centroid estimation. The noise sources include CCD read noise, background light, stray light, and CCD-processing electronics. One of the most widely used methods to reduce the effects of the noise and background bias is the thresholding method, which subtracts a fixed threshold from the centroid window before centroid computation. An improved centroiding method that utilizes the spot model to derive the signal boundary (which is used to truncate the noise outside the signal boundary) has been developed and is presented in this section [77]. This process effectively reduces both the bias and the noise. The effectiveness of the proposed method is demonstrated through simulations.

Accurate centroid estimation is a critical task for a beacon-based pointing system. Past studies have shown that the centroid error (random and bias error) for deep space optical communications needs to be less than 1/20th of a pixel, whereas the total pointing error allowed (1 sigma) is 1/15th of a pixel [78.79]. Two types of centroid errors, random and bias, are affected by various sources. A random error is caused by noises, such as CCD read noise, shot noise, dark

current, and ADC quantization noise. A bias error occurs when non-uniform background light, such as stray light and Earth background image, exists.

Conventional methods to reduce the noise and bias include thresholding and centroiding of the normalized zero-crossings [80,81]. For the thresholding method, an estimated threshold is subtracted from the centroid window, which equivalently performs a bias subtraction and eliminates the noise. This method can be effective when the threshold value eliminates most of the bias and the noise. However, a simple threshold, in general, is not effective since the threshold value is dependent on the brightness of the image, and the number of pixels forming the object may be altered by the thresholding process. To avoid this problem, the use of zero-crossings for centroid estimation was proposed [81]. The limitation of that approach is the assumption of equal weighting on every pixel.

For the same objective of reducing the effects of noise, there were suggestions to use only nine pixels around the signal peak [82,83,84]. This truncation can simplify the centroid calculation without affecting the centroid accuracy if the signal is limited to this small local region. As was indicated in [85], however, the truncation of the wide signal considerably affects the accuracy of centroid estimation. Therefore, the number of pixels used in centroid estimations needs to be carefully selected so as not to sacrifice the centroid accuracy.

In this section, we develop the use of a spot model to determine which pixels are used for centroid estimation. A spot model can be constructed from the characterization of the optical system point spread function (PSF). On the centroid window, which is usually several pixels larger than the beacon spot size to allow beacon motion, the approximate signal boundary of a beam spot can be estimated from the spot model and the measured noise level. Once the boundary is identified, the pixels to the outside of the signal boundary can be set to zero, effectively eliminating all the noise and bias outside the beam spot.

5.3.6.2.1 Effects of Noise and Bias on Centroiding Accuracy. The equations for centroids (centers of brightness) for spots on a CCD-type focal plane arrays are well known [79]:

$$C_x = \sum_i ip_{ij} / \sum_{ij} p_{ij}, \quad C_y = \sum_j jp_{ij} / \sum_{ij} p_{ij} \tag{5.3-27}$$

where i and j are the x and y axis coordinates, and p_{ij} is the output of the (i, j)th pixel value of CCD. From Eq. 5.3-28, it is clear that the noise or bias closer to the edges of the centroid window dominates the centroid error due to the larger weighting factor as coordinates increase toward the edges. This is one of the

most important motivations that led to the development of the spot model method.

The equation for the random centroid error (or NEA, noise equivalent angle) is summarized as follows,

$$NEA = \sqrt{\frac{S + N_P(\mathrm{Var}(R_F) + \Delta t\, R_T}{\frac{S_2 N(N+1)}{3}}} \tag{5.3-28}$$

where S = total signal.

Δt = the exposure time.

N = truncated half width of centroiding area.

N_P = number of pixels involved in the centroiding area, $N_P = (2N+1)^2$

R_F = fixed per-pixel noise (1σ), such as read noise.

R_T = per-pixel background signal (including straylight and dark current).

Equation (5.3-28) indicates that NEA is inversely proportional to SNR. Therefore, either the signal needs to be increased, or the noise needs to be decreased in order to reduce the NEA. This implies that the effect of the noise is small if the signal is relatively larger than the noise and vice versa. To illustrate this, let us take an example where the spot signal is low. For deep-space optical communications that may require stars as a beacon source, the minimum signal available from an 11th-magnitude star with a 30-cm telescope is 10,000 photons (with 25-percent system efficiency). Assuming a CCD QE of 0.5, this translates to 5000 electrons. In this example, the reduction of the centroid window size improves the centroid accuracy significantly if it does not truncate the signal notably. Response of the NEA versus the number of pixels used in the centroid estimation is shown in Fig. 5-45. The assumptions are (1) the same fixed per pixel noise ranging from 5 e^- to 20 e^- and (2) no background signal. Figure 5-45 shows that NEA increases more rapidly with larger fixed per-pixel noise as the number of pixels (used in centroid estimation) increases.

Bias error is caused by non-uniform signal distribution, which includes stray light and background image. This corresponds to the cases where the telescope is pointing toward the Earth or close to the Sun. Even if background subtraction were applied, there would be some bias left, especially if the threshold is below the maximum of the background signal. As Fig. 5-46 shows, even 0.1 percent of the peak spot value as the maximum bias value can cause considerable bias error if the centroid window size is large, such as 9 × 9 pixels as in this example.

5.3.6.2.2 Comparison of Algorithm Performance. To demonstrate the effectiveness and robustness of the model-based noise reduction method in centroid estimation, three cases were investigated. First, three centroiding

algorithms (including model-based) were compared at various noise levels given a total signal equivalent of 5000 e^-. Second, the bias error resulting from the three centroiding methods was compared. And last, three scenarios of using incorrect spot models were used. Incorrect models at 0.1 pixels, 0.2 pixels, and 0.3 pixels were used to show the robustness of the model-based method.

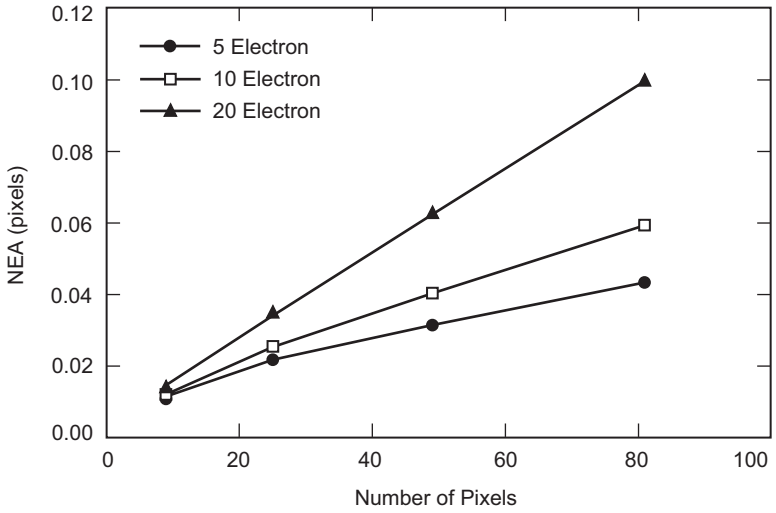


Fig. 5-45. Effect of the number of pixels used in centroid estimation. NEA increases as the number of pixels increases and more rapidly with larger noise level.

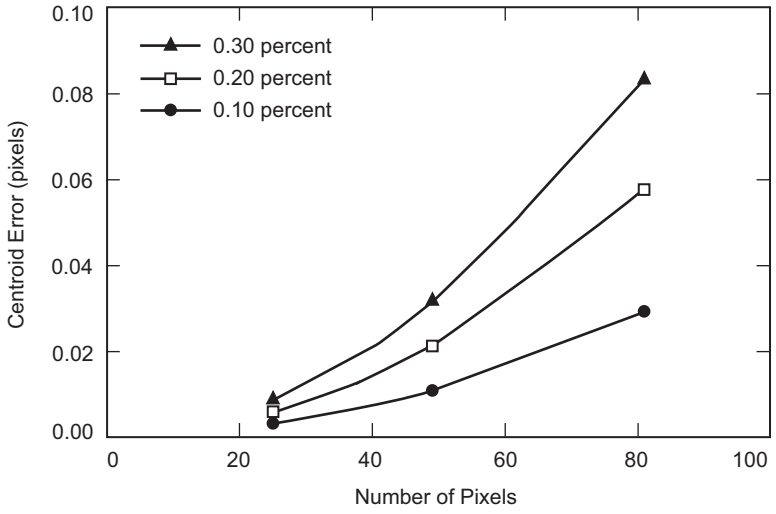


Fig. 5-46. Bias centroid error vs. the number of pixels used in centroid estimations at bias values from 0.1 percent to 0.3 percent of the peak pixel value.

Three algorithms were run 100 times for a fixed-noise value, and the noise was increased from 10 e^- equivalent to 100 e^- equivalent (Gaussian noise with 1 sigma value from 10 e^- to 100 e^- equivalent). The 1-sigma error and mean error were computed and plotted in Fig. 5-47. As is shown, the model-based algorithm outperforms the other two methods. The strength of the model-based method is not only the much smaller centroid error but also its insensitivity to the noise, as demonstrated in both plots. As the noise increases, the centroid error from the standard and thresholding methods also increases. However, the model-based method exhibits a nearly constant error.

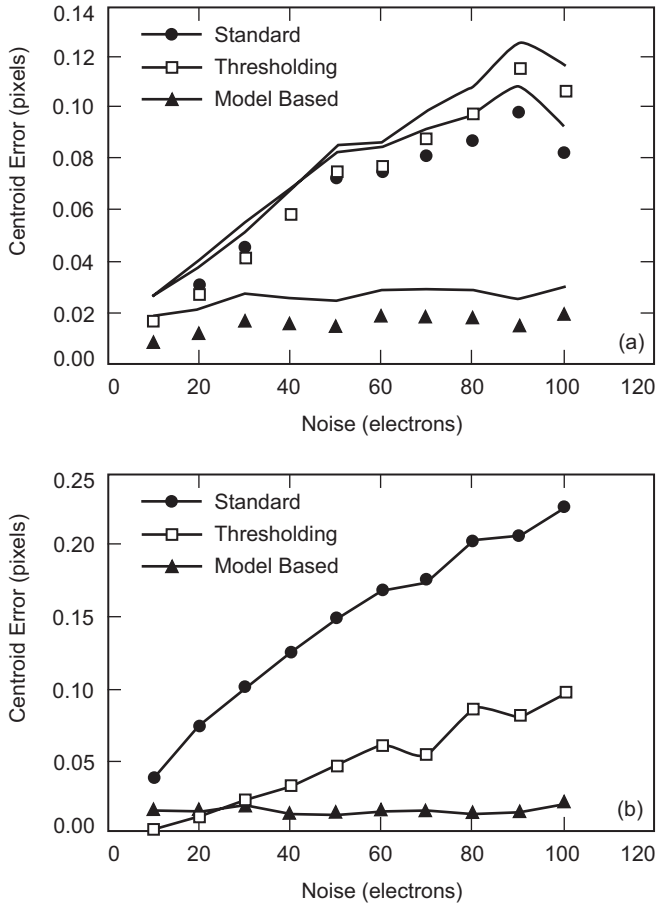


Fig. 5-47. Comparison of three centroid methods: standard centroid, thresholding method, and model based method. Plots show the centroid error vs. fixed per pixel noise (equivalent to electrons based on total signal of 5000 electrons): (a) 1 sigma random error and (b) mean error.

5.3.6.2.3 Bias Error. The effect of a bias in the spot image using the three centroid methods was also compared. As was evident in Eq. (5.3-27) above, even the model-based method would be affected by the presence of the background bias unless complete removal of the bias is conducted. The bias value was selected based on the peak pixel value that is 28.5 percent of the total signal. Maximum bias was varied from 0.1 percent to 1 percent of the peak pixel value. Figure 5-48 shows the results. As expected, the model-based method outperforms the other two methods, which exhibit a linear relationship between the bias value and the centroid error.

A method based on the spot model was proposed to improve centroid estimates of a point source image. The method assumes the spot model can be used to truncate noise and bias in the measured spot, thereby improving centroid estimates. Simulations were performed to demonstrate the effectiveness of the proposed method for noise and bias. Compared with the standard centroiding and the more advanced thresholding method, the model-based method was found to be superior in accuracy. From the simulations where the incorrect spot models were intentionally used, the effect on its performance was minimal, especially at high noise levels. Since this method was intended for low SNR signal, it could prove to be essential for deep-space missions, where a strong optical signal is not readily available. Notice that both the 1 sigma and mean errors are in the neighborhood of 1/50th of a pixel.

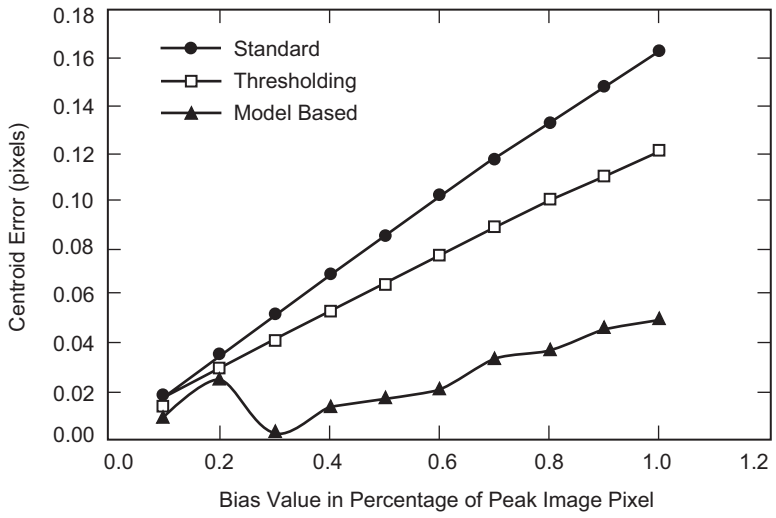


Fig. 5-48. Bias was added to the spot image, and the three centroid methods were applied. Bias values increased from 0.1 to 1 percent of peak pixel value of the spot image. The model based method has a noticeably smaller bias error contribution.

5.3.6.3 High Bandwidth, Windowing, CCD-Based Camera. In this section, we describe the functionality, architecture, and control methodology of a random-access, real-time, event driven (RARE) camera [86] as part of a real-time target acquisition and tracking platform. The camera implementation uses a Texas Instruments TC237 charge-coupled device (CCD) focal-plane array (FPA) and two TLV987 signal processors [87]. Control of the imager and signal processors is via custom logic in a field-programmable gate array (FPGA) that accepts user commands and provides region-of-interest pixel data to a host tracking processor. A message-passing paradigm is used to provide real-time imager control without knowledge of detailed imager operation.

Commercially available CCD cameras are not designed for a combination of single frame and high-speed streaming digital video with real-time control of size and location of multiple regions-of-interest (ROIs). To achieve low-level camera control with high-level system operation, a message-passing paradigm is defined. This functionality is achieved by asynchronously sending messages to the camera for event-driven operation, where an event is defined as image capture or pixel readout of a ROI, without knowledge of detailed in-camera timing. This methodology provides a RARE camera for adaptive camera control, and it is well suited for target-tracking applications requiring autonomous control of multiple ROIs. This methodology additionally provides for reduced ROI readout time and higher frame rates as compared to a predecessor architecture [88] by avoiding external control intervention during the ROI readout process.

5.3.6.3.1 Camera Requirements. The primary motivation for this camera development is to realize an adaptive sensor mechanism as part of a platform for real-time autonomous acquisition and tracking applications [89]. Such a platform requires both a sensor and a control philosophy that provides real-time adaptation of the sensor based on target characteristics and dynamics and environmental conditions. The requirements for the sensor in a deep-space application are generally for a low-noise, high-QE, high fill factor, large-format device. These requirements are currently best met by CCDs. To achieve this tracking goal requires a camera capable of frame rates of several hundred to several thousand frames per second with operating parameters that can be adjusted on a per-frame basis. High frame rates with adaptive imager control are achieved with a conventional CCD by reading out only the pixel regions of interest and discarding all other pixels. This mode of operation required the development of a customized local controller for the CCD imager to provide a tightly coupled mechanism for imager operation. Configuration of the controller is handled by a host tracking processor that loads initialization and tracking parameters into the controller to define imager operation. The initialization parameters are needed for defining the start-up mode of the controller, and the tracking parameters define detailed operation of the imager

during acquisition and tracking operations. A previous release of the camera used software control of low-level CCD operations as a first implementation but required tight coupling of the camera with the host tracking processor. That release of the camera was discussed previously [86]. Release 2.0 represents the current state of the RARE camera development.

5.3.6.3.2 Hardware Architecture. Release 2.0 is used as part of a real-time target-tracking apparatus for free-space optical communications and non-invasive eye tracking [89] and provides simplified high-level control of low-level camera operation on an intra-frame basis. The architecture of the release 2.0 RARE camera is shown in Fig. 5-49. This figure illustrates the three-component system of the RARE camera, consisting of a custom imager card with a low-voltage differential signaling (LVDS) cable assembly, a commercial of-the-shelf (COTS) field-programmable gate array (FPGA) card, and a commercial processor. The imager card was designed with a Texas Instruments TC237 CCD imager chip [90] with two TLV987 signal processor chips [87]. Two 987 processors were required to handle the dual-pixel-stream output capability of the TC237 CCD. Each processor accepts an analog pixel stream and provides transistor–transistor logic (TTL)-level output signals to the custom LVDS cable. The cable assembly provides single-ended TTL-level input–output (IO) signals to the CCD card and the FPGA card, but it runs differential signals through a pair of small computer system interface (SCSI) cables to allow for high-speed strobe operation over several feet of cable. The FPGA card is a TransTech PMCFPGA-01 card with a 300 kilo-gate Xilinx XCV300E FPGA for the low-level controller of the CCD imager. The host tracking processor is a general-purpose computer with a 32-bit peripheral

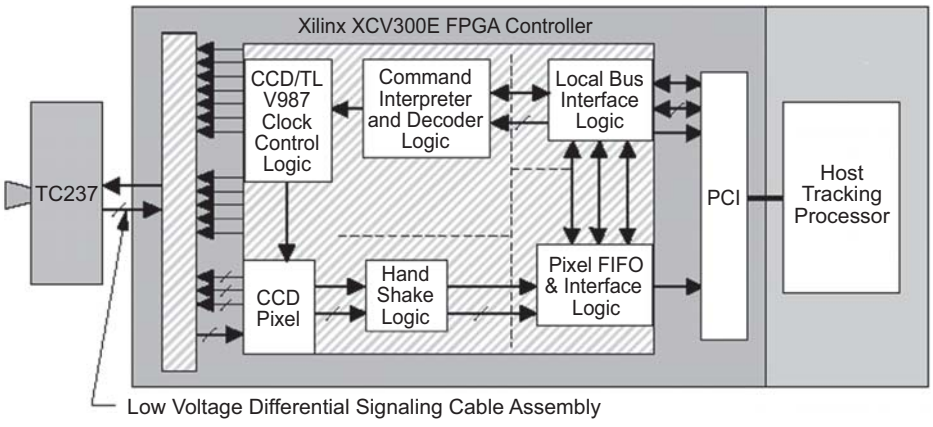


Fig. 5-49. Random-access real-time (RARE) camera tracking system architecture (FIFO = first in, first out; PCI = peripheral component interconnect).

component interconnect (PCI) bus used to provide FPGA control parameters, to read camera status, and to read pixel data from the FPGA card.

5.3.6.3.3 Multi-ROI Operation in Software. The host tracking processor loads ROI parameters into the FPGA controller to define the ROI size and location. These parameters are used to scroll through unwanted lines and unwanted pixels per line until the ROI is reached. They additionally define the number of pixels per line and the number of lines to read out for the ROI.

Dynamic adjustment of these parameters can be done on a per-frame or intra-frame basis to allow for enhanced system adaptation. They can be used to define an ROI for one or more frames, or they can be adjusted within one frame to allow for multiple ROIs within a single frame.

When readout of a ROI(s) is required, the tracking processor sets the mode of the FPGA controller to ROI readout and requests pixel data as defined by the ROI parameters previously loaded. The tracking processor then polls the FPGA controller for available pixel data to initiate the ROI readout from the pixel FIFO in the FPGA. If more than one ROI per frame is required, the tracking processor can load new size and location parameters for the next ROI without requesting a new frame transfer. The vertical location parameter of the next ROI is defined relative to the last line of the current ROI and denoted by the inter-window scroll area.

This methodology also allows for ROIs that overlap or share common rows of pixels with or without vertical separation. The primary difference as compared to the case above is in the definition of the ROI. For the case of common lines of pixels between ROIs, the tracking software must read out three different regions corresponding to the two areas where the ROIs do not share common rows of pixels and an additional third region containing pixels from both ROIs. This additional region must read out pixels for both ROIs and will have a width parameter defined by the left edge of the left-most ROI and the right edge of the right-most ROI. This methodology is illustrated in Fig. 5-50 for two ROIs and is applicable to N(>2) ROIs.

5.3.6.3.4 Dual-ROI Benchmark Results. The dual-ROI methodology was benchmarked to determine the achievable frame rates for this scheme. The "home" position of two 11×12 ROIs is defined by origins (320,236) and (320,248). These locations place the ROIs in the same columns and with a one-row separation at the center of the CCD FOV. The top ROI is allowed to move throughout any portion of the upper half of the CCD FOV, and the bottom ROI moves in an opposite sense throughout the bottom half of the FOV. The ROIs are moved in opposite directions to emulate the operational mode required in the acquisition and tracking platform. The reported frame rates include the time needed to perform frame transfers and read out the ROI data only. Figure 5-51 shows that frame rates from 900 to 1100 Hz are achievable for dual-pixel-

stream operation. There was no timing delay effect observed when shifting the ROIs horizontally. The reason for this result is that the total number of pixels read out from the CCD is constant. Consequently, the only parameter that affects the frame rate is the position of the last row of the bottom ROI. The reason for this result is that the only variable for this operational mode is the number of line-scroll operations performed, and the aggregate number of line scrolls increases as the bottom window is moved toward the bottom of the CCD FOV. A further increase in the frame rate is possible by using only a portion of the CCD FOV. This approach requires the defining of a sub-region of the CCD FOV and the centering of the ROIs in this region. The number of line scrolls and pixel reads is reduced by moving the "home" position of the two ROIs closer to row one, column one of the CCD FOV.

The RARE camera provides a key component for the real-time, adaptive tracking platform. We have developed this infrastructure by implementing a methodology to quickly extract pertinent pixel data using a commercially available progressive scan imager. This technology is also well-suited to adjusting the camera parameters to accommodate changing ambient and target conditions during tracking. In this section, we presented details of the RARE camera design based on the Texas Instruments TC237 CCD imager chip. The novel feature of this design is the use of an event-driven paradigm for imager control. This capability was implemented by developing a custom FPGA controller that converts a commercially available CCD imager into a smart pixel device. Communication to the FPGA controller is via commands from a host tracking processor. This combination of FPGA controller and host tracking

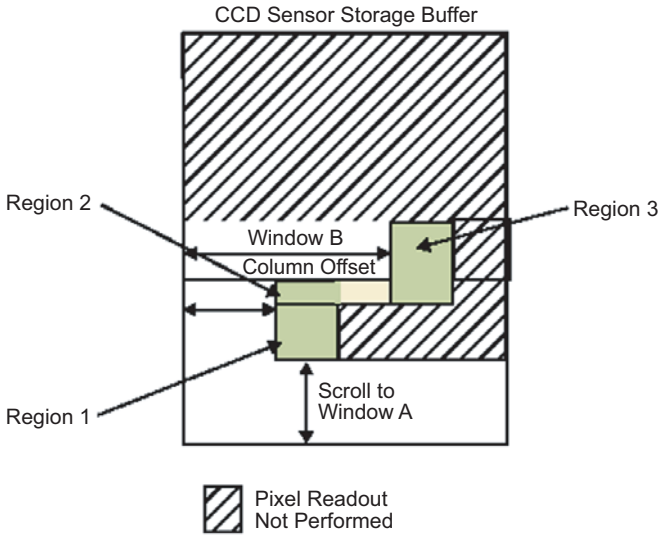


Fig. 5-50. Region-of-Interest (ROI) readout with common rows of pixels.

processor provides for higher-level commands to handle low-level imager operation for dynamically controlled ROI capability on a per-frame and intra-frame basis.

To assess the speed performance, several experiments were conducted for single- and dual-ROI operation. The first experiment illustrated the change in frame rate for a single fixed ROI origin and varying ROI size. The second experiment illustrated the change in frame rate for a single fixed-sized ROI with varying origin from the first row (top) and first column (left) of the CCD FOV. The results from these experiments achieved 75 percent of the theoretical best-possible frame rates for this CCD imager. The difference between the theoretical limit and the experimental results is due to a combination of both fixed and variable delays in the FPGA logic and host tracking processor software.

The dual-ROI results presented in Fig. 5-51 show that frame rates from 900 to 1100 Hz are achievable for two 11 × 12 ROIs centered about the CCD FOV. The two ROIs move in a counter-propagating fashion to emulate the operational mode of the envisioned acquisition and tracking platform. These frame rates will vary if the ROIs are allowed to move in a co-propagating fashion, as would be the case for the tracking of two targets within a scene. The frame rates for this case are governed by the ROI sizes and by the relative positions of the ROIs.

A nearly two times speedup is possible by running the imager at its maximum possible clock speed of 20 MHz per pixel stream. Additional speed

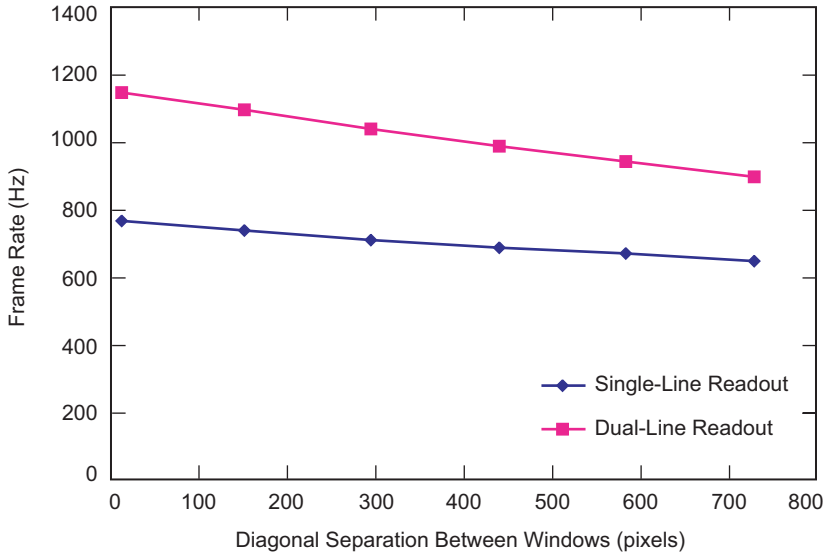


Fig. 5-51. Two-window ROI timing with diagonal shift. (Windows start at center of FPA and move to opposite corners.)

increases require using a different progressive scan imager with more than two pixel outputs to provide more parallelism in accessing the image data. Due to the generality of the RARE camera control scheme, the FPGA controller can be used for other commercially available CCD imagers to optimize system performance in terms of speed, image quality, or other parameters of interest.

5.3.6.4 Accelerometer-Assisted Beacon Tracking

5.3.6.4.1 Increasing Loop Bandwidth with Dim Beacon Sources. In deep-space optical communications, acquisition, tracking, and pointing are all challenging because of the stringent—on the order of sub microradian—pointing requirement. To achieve this level of pointing accuracy, one must maintain high-bandwidth tracking control. Feasible tracking sources (beacons) include uplink laser beams and celestial objects such as the Earth, the Moon, and stars. However, these tracking sources do not all provide the kilohertz tracking rate needed for pointing in deep space. One approach to enable a high tracking rate is to augment the tracking loop with inertial sensors to estimate high-frequency beacon movements [91]. In this section, we discuss the use of linear accelerometers, mounted in a configuration to measure angular displacement, to achieve high-bandwidth tracking with dim beacon sources. The advantages of linear accelerometers (or angular accelerometers) are their low cost, high bandwidth, and small size compared with other inertial sensors such as gyros. Simulation and experimental results show good agreement. A tracking bandwidth increase of 11 times has been demonstrated [92].

High-data-rate, narrow-beam optical communication imposes the challenge of pointing a downlink beam to a fraction of the beam divergence, typically sub-microradian in jitter and bias. This, in turn, requires a reference optical source, a beacon that can be used as a reference for closed-loop tracking/pointing control. In the past, ATP system design required a beacon-tracking rate of several kilohertz to maintain the link properly. The required tracking rate depends on the platform vibration amplitude and frequency contents. A typical tracking source has been a laser beacon, especially for short-range optical communications such as intersatellite optical links [93].

However, the kilohertz beacon-tracking rate is not readily available in most deep-space applications due to the long range that limits beacon energy collected at the spacecraft telescope. This is true even for Earth-image-based tracking and star tracking [54]. The challenge is to achieve high-rate beacon tracking, even with low-rate beacon centroid measurements, that is, to estimate accurately the relative beacon position movements between the measured beacon centroids. In the past, similar problems were addressed with the use of inertial sensors: spacecraft attitude control using star trackers and gyros [94], and (in the case of the Hubble telescope) pointing using star trackers and various inertial sensors [95]. Although these applications are slightly different,

the underlying principle is identical. Since the downlink target is moving very slowly in inertial space, all high-frequency motions come from the spacecraft. The high frequency movements of the beacon relative to the target can be deduced from the measurements of the source (platform) vibrations that cause movements of the reference beacon (either laser beacon or celestial objects) on a CCD array. If the error between the true and the estimated beacon positions is smaller than the error budget, a fast tracking rate can be maintained. Implementation of this concept requires accurate high-bandwidth inertial sensors. Among the possible inertial sensors are angular-rate sensors, angle-displacement sensors, gyros, and angular and linear accelerometers. Because of the low cost of linear accelerometers (as well as their accuracy over high bandwidth, small size, and availability) they make an attractive option for implementation. Using linear accelerometers requires double integration for the position estimation from acceleration measurements. Furthermore, linear accelerometers are not as sensitive to low-frequency vibration as are gyros, a feature that limits their usage in the case of very low beacon intensity. However, previous use of linear accelerometers suggests their promise in a range of ATP applications. Linear accelerometers have been used successfully, in the line-of-sight stabilization of a gimbaled imaging sensor suite [96] and in measuring the rotational and translational acceleration of a rigid body [97]. For deep-space optical communications, we sought to demonstrate that linear accelerometers could be used for beam pointing and control as well as for line-of-sight stabilization and for measuring the movement of a single body. For the double integration of accelerations to estimate displacements, some problems and solutions for zero-mean displacement signals have been addressed [98,99].

The key issue in high-bandwidth tracking is the availability of beacon centroids at high rate. Given the limited beacon intensity in deep space (thus, small number of beacon centroids available) and the fact that the beacon movements are caused by spacecraft vibration, it is essential to be able to estimate the beacon centroids (or movements) at times when the beacon centroids are not available. In this section, we give the three-accelerometer configuration for the two angular position estimations and describe the algorithm used. Detailed treatments of the trapezoidal method for the linear displacement estimation, initial velocity/acceleration bias estimation, and random error analysis are given in [92].

5.3.6.4.2 Algorithm Theory. Angular displacements on two axis (α, β) can be obtained using three accelerometers, as shown in Fig. 5-52. Let α (horizontal) and β (vertical) be the angular displacements of the x−z plane around the z-axis and the y−z plane around the y-axis, respectively. A_1, A_2, and A_3 represent three accelerometers, and d_1, d_2, and d_3 represent the corresponding estimated linear displacements. Linear displacement estimation from the acceleration

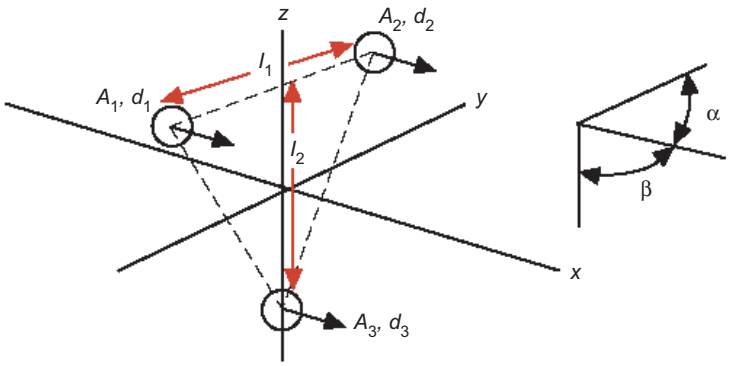


Fig. 5-52. Triangular configuration of three accelerometers.

measurements using the trapezoidal method is described in [92]. Then, the two angular displacements due to the three linear displacements are

$$\left.\begin{aligned}\alpha &= \frac{d_1 - d_2}{l_1} \\ \beta &= \frac{\frac{d_1 + d_2}{2} - d_3}{l_2}\end{aligned}\right\} \tag{5.3-29}$$

where l_1 is the separation between A_1 and A_2, and l_2 is the separation between A_3 and the middle point of the line connecting A_1 and A_2.

Angles at the Nth sample time can be represented as

$$\left[\alpha_N \beta_N\right] = B\left\{C\left[D_1 D_2 D_3\right]\right\}^3 \tag{5.3-30}$$

where

$$B = \begin{vmatrix} \frac{1}{l_1} & \frac{-1}{l_1} & 0 \\ \frac{0.5}{l_2} & \frac{0.5}{l_2} & \frac{-1}{l_2} \end{vmatrix}$$

$$C = \left[1, (N-1)\Delta t, \frac{1}{3\Delta t^2}, \frac{\Delta t^2}{6}\right], N = 2$$

$$C = \left[1, (N-1)\Delta t, \left(\frac{0.5N-2}{3}\right)\Delta t^2, (N-2)\Delta t^2, \cdots, (N-i)\Delta t^2, \cdots \Delta t^2, \frac{\Delta t^2}{6}\right], N > 2$$

Δt = sampling interval

$D_i = \left[d_{i1} v_{i1} \alpha_{i1} \alpha_{i2}, \cdots, \alpha_{iN} \right]^T$

d_{i1} = initial position of d_{iN}

v_{i1} = initial velocity of d_{iN}

d_{iN} = displacement estimation at the Nth sample time from $\alpha_i's$

$\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iN}$ = accelerometer outputs from accelerator $A_i, i = 1,2,3$

5.3.6.4.3 Algorithm Design. Figure 5-53 shows all the major signal flows, from three accelerometer measurements, two angle reference inputs (beacon position centroids), and the final outputs of the two angular position estimates of the angular position estimation algorithm (APEA). Additional inputs are reference signals in terms of beacon centroids (x-axis, y-axis). The linear displacement estimator produces three displacement estimates corresponding to the three accelerometer outputs. Combined with the three initial positions derived from the beacon centroids, three position estimates (p_1, p_2, p_3) are obtained. These are, in turn, inputs to the initial velocity and acceleration bias estimator. The estimated initial velocity and acceleration bias are fed back to the linear displacement estimator to improve the next position estimations. The final angular position estimations are obtained from the estimated three linear

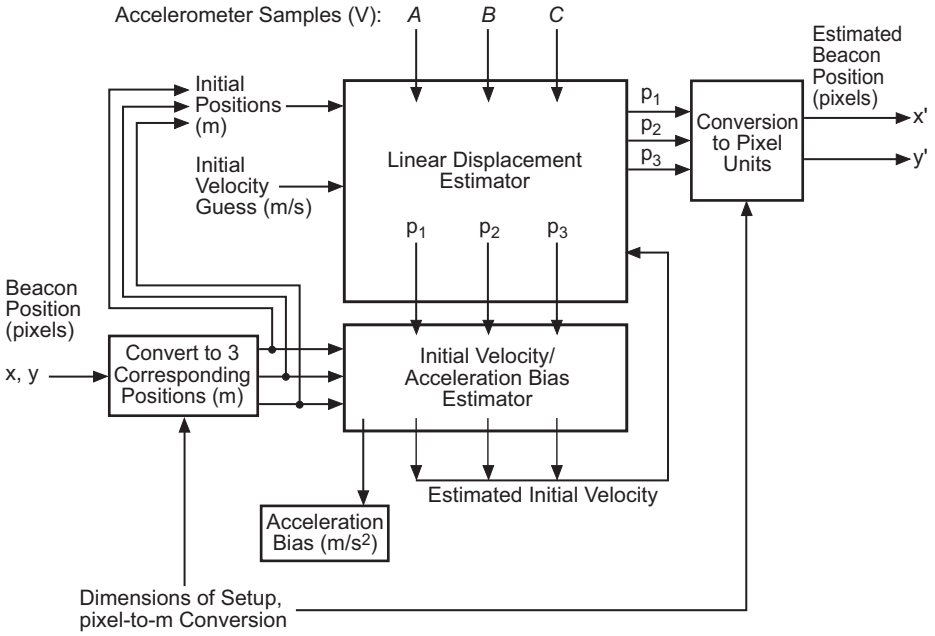


Fig. 5-53. Angular position estimation algorithm (APEA) block diagram showing the major signal flows.

positions, after performing the linear displacement-to-angle conversion [Eq. (5.3-29)]. The rate of beacon reference inputs to the APEA determines the reference reset period, *N*. For example, if $N = 2$, every second beacon position output is an estimation, while the other is the true beacon position. If $N = 5$, every 5th output is the true beacon position. For this experiment, we did not do any smoothing over multiple beacon samples due to the noise of the accelerometers.

5.3.6.4.4 Algorithm Simulation. A three-accelerometer configuration of the experimental setup was used. Sinusoidal signals of 1, 10, and 100 Hz for vibration were used with an assumption of zero measurement noise. The only error sources are the algorithm errors of the APEA. Figure 5-54 shows the displacement estimation results. As shown, the error increases with both the frequency of the vibration signal and the reference reset period.

5.3.6.4.5 Experimental Validation. For the experiment, three accelerometers were mounted around the optical communications terminal, and the entire setup was placed on a vibration table (Fig. 5-55) [100]. The experimental procedure was as follows.

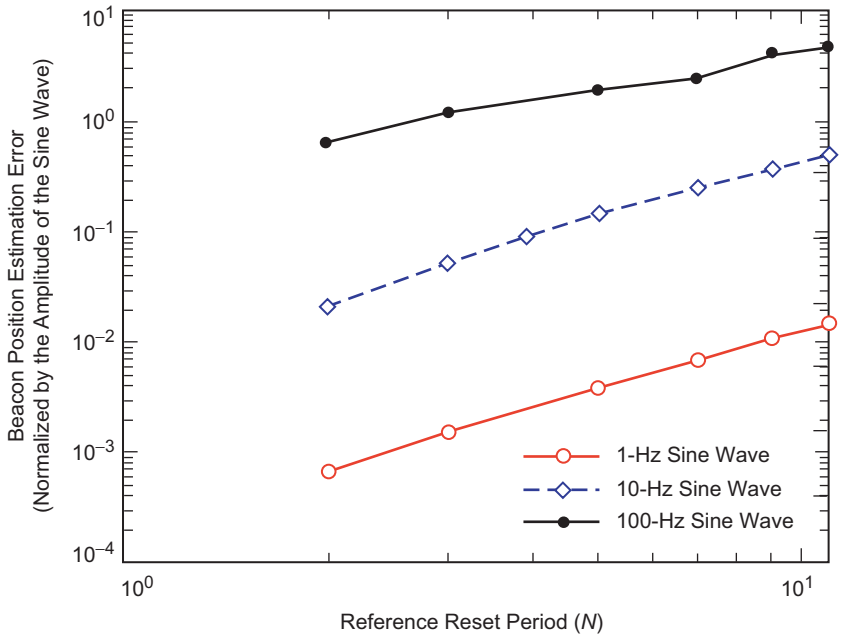


Fig. 5-54. Angular position estimation error as a function of reference reset periods for sinusoidal signals of 1, 10, and 100 Hz. The error is proportional to both the reference reset period and frequencies of the underlying vibration signals.

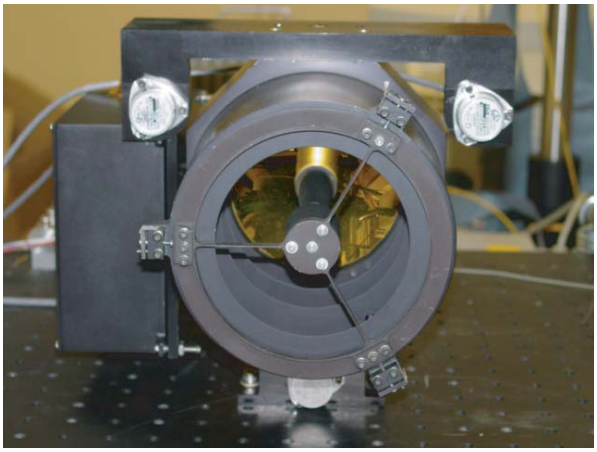


Fig. 5-55. Three accelerometers were mounted around the optical communications terminal, and the entire terminal was place on the vibration table. The piezo-actuator underneath was commanded to shake the table with the generated vibration signal.

1) Generate the angular spacecraft vibration signal to command the piezo-actuator. In this experiment, the laboratory-measured Cassini spacecraft vibration [linear acceleration power spectral density (PSD)] was used to derive the angular vibration signal for a more realistic frequency content of the expected deep-space vibration signal. The amplitude of the vibration signal was inversely proportional to the length of the interface plate of the optical communications terminal. For this experiment, 15 cm was used. The transformation of linear acceleration into rotational displacements was done following the procedure in [101].
2) Command the piezo-actuator to shake the platform vibration table with derived PSD.
3) Measure the angular motion (reference vibration signal or beacon centroids) using the optical communications terminal.
4) Run the angular position estimation algorithm with various reference reset periods.
5) Compute the angular position estimation error. The derived vibration signal was sampled with a CCD at a 625-Hz rate.

The experiment was done for reference reset periods of 2, 3, 5, 7, 9, and 11 frames. The beacon position estimation results for a reset period are given in Fig. 5-56. As was indicated in the simulation results, the error is proportional to the frequencies of the vibration signals. The error also grows for larger reference reset periods (Fig. 5-57).

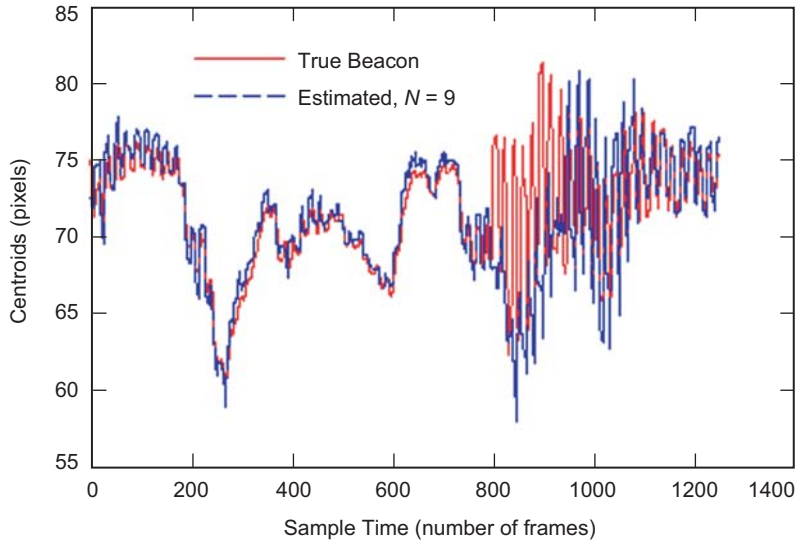


Fig. 5-56. Experimental beacon position estimation results for a reference period of 9 frames. As was shown in the simulation results, the larger error is noticeable for larger reset periods.

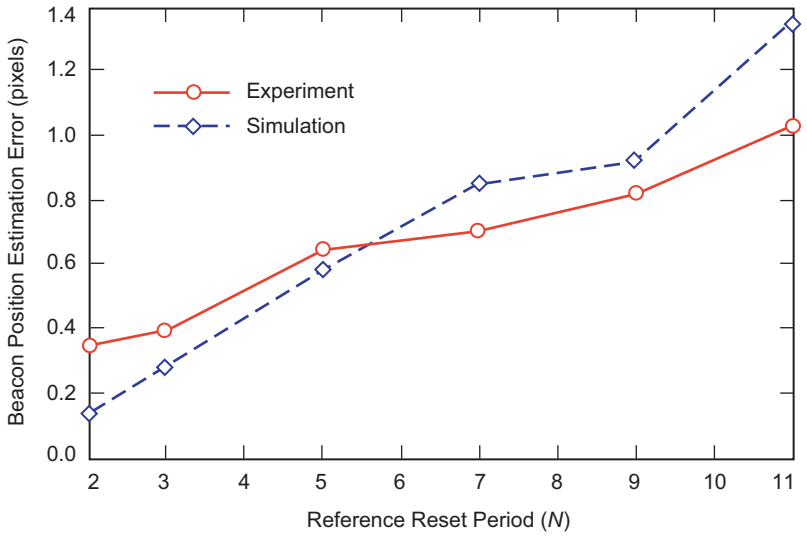


Fig. 5-57. Comparison of the experimental and simulation results. The measured vibration on the CCD of the optical communications terminal was used for the simulation with an assumption of zero noise in the acceleration data. Overall estimation (RMS) errors match between the two results.

Overall, the simulation and experimental results matched closely. The amplitude of the vibration signal shown in Fig. 5-56 is about ±10 pixels (or 36.1 μrad). The resulting angular position error is about 1 pixel (or 3.61 μrad) for the reference reset period of 11 (Fig. 5-57). Since the estimation error is directly proportional to the amplitude of the vibration signal, the desired sub-microradian pointing is achievable for the deep-space optical communications if the spacecraft vibration can be suppressed below a certain threshold. The threshold depends on the amplitude and frequency contents of the specific spacecraft vibration. For our experiment, about one-third of the given vibration amplitude, or ±12 μrad, would give microradian-level error in angular position estimation for the reference reset period of 11, as an example. This would increase beacon-tracking bandwidth by 11 times. Therefore, a tracking bandwidth of 1 kHz can be achieved with a beacon tracking rate of 91 Hz.

The concept of using linear accelerometers to increase the tracking bandwidth can be applied for deep-space optical communications tracking and pointing with a trade-off for the additional error in the beacon position estimations. Simulation and experimental results show good agreement in the beacon-position estimations with the various reference reset periods. The results also showed that the estimation error is proportional to both the reference reset period and the frequencies of the vibration signals.

5.4. Flight Qualification

Hamid Hemmati, William T. Roberts, and Malcolm W. Wright

During the early days of the Apollo program, there was great concern at NASA Headquarters over the effect on the program of a failure outside of earth orbit. Because of the obvious safety concerns, NASA leaders were insistent that the Apollo missions hold to a standard of 'three nines' (99.9%) reliability during the lunar phase of the mission. The NASA administrator questioned Dr. Von Braun, asking him if he could be assured of this level of reliability. Dr. Von Braun considered the question, and then put the question (in German) to each of his four lead engineers, "Can you think of any problems which might cause a catastrophic failure." Each engineer answered confidently "Nein!" At the completion of this questioning, Dr. Von Braun turned back to the contingent from NASA Headquarters, and said, "Gentlemen, I give you four 'neins' reliability!"

Anecdote related by Dr. Ernst Stühlinger
Former Associate Director for Science,
Marshall Space Flight Center, Huntsville, Alabama

5.4.1 Introduction

The expense of launching spacecraft, and the limited ability to repair or replace components on deep-space missions, leads to the practice of producing

flight hardware systems using the highest grade components available and proving their reliability and durability in the space environment through rigorous, repeated testing. This process is known variously as flight qualification or space qualification, and it is applied in varying degrees of rigor, depending on the mission. The flight qualification testing of electronic and electro-optical parts is often very expensive, requiring many tests on many sample parts to validate the reliability claims and assure in-flight performance.

This portion of the chapter is intended to provide an overview of the process of flight qualification and the various levels to which testing is carried out under typical flight qualification is described here. The description begins with an examination of the approaches available in qualifying parts for space flight, and describing the latitude to which one can reasonably tailor a flight qualification program. Section 5.4.5 goes on to explain the various conditions under which a deep-space optical communications terminal may be called upon to operate, and covers most of the environments of concern on a deep-space mission. The final section deals with the specifics of qualifying electro-optic detectors, lasers, and other optical components for space flight.

5.4.2 Approaches to Flight Qualification

There are two main approaches to the development of parts that are ultimately qualified to operate in space. In the first approach, flight parts are developed for the particular space environment from inception. In this case, a part can be designed with operating requirements in mind from the beginning; environment-tolerant processes, materials, and structures can be designed to accommodate operation at the required levels; and manufacturing systems with traceability and accountability can be implemented to achieve the goals and requirements of the program. This approach is obviously difficult to implement, and it requires an extremely long lead-time to set up the facilities and establish the procedures that will ensure part survivability. As a result, flight parts development for a particular space environment is generally expensive to implement. Programs have been developed along these lines; device and materials development for operation in space environments and in high radiation environments have been pursued for development of detectors and CCDs under various military programs. However, they are usually implemented only after existing products intended for terrestrial applications are tested and shown to be inadequate [102].

The more common approach to flight qualification focuses on the testing of commercial off-the-shelf (COTS) components, with the intent of determining the likelihood that the particular component will withstand the environment to which it is to be subjected. In many cases, COTS components can withstand many of the stringent requirements levied on them by the space environment. However, because they may never have been tested to this level, there is little

or no data upon which to rely. In this case, lots of components must be purchased and rigorously tested to determine the part's reliability under expected space conditions. This process is referred to as "up-screening," which is the main focus of the section on qualification of specific parts and materials (5.4.6 and 5.4.7).

Usually, preparatory research is invaluable in selecting a part that will require much less additional screening, and it will have a favorable impact on the program's cost and schedule constraints. For example, it may be that a particular detector was used on a previous flight program. Successful performance of the detector under the particular conditions of that mission is somewhat useful in providing confidence of the part's reliability, even though the vibration, thermal, or radiation environments may be different. It is likely that there will be test data on any similar components used to qualify the flight part on the mission. This data will be useful in flight qualification of the considered part, though the cognizant engineer must check with the manufacturer to assure that the materials, structures, and processes of component manufacturing have not changed in the interim. The most fortuitous outcome of this research may be that there are additional flight-qualified parts, either spares in flight storage or perhaps integrated into an engineering unit, which may be available to the program.

If the failure rate of a part is too high to be acceptable, minor modifications to the manufacturing processes and materials and testing methods are pursued first in an effort to effect a remedy. The screening tests should be helpful in indicating the source of the failure, and generating data that can be used to develop more environment-tolerant designs or assist in the selection of materials and processes. While this process is expensive and time-consuming, it nevertheless is still significantly less costly than setting up a dedicated manufacturing line for flight-qualified component development.

Here again, research is very important in identifying potential cost and schedule savings. There may be alternative parts that already incorporate the changes being considered to the product, and knowledge of the performance of those parts under the particular environment leading to failure may help to avoid a costly dead-end in the development of a qualified component.

Finally, if the changes required to bring a part up to the specifications will have a significant impact on the cost or schedule of the component integration, this must be communicated back to the project. In many cases, the component requirements can be relaxed with little or no impact on the overall performance of the system. For example, radiation-tolerance requirements may be relaxed by the addition of shielding to the system, or often simply by placing the component in a different location within the spacecraft. Vibration requirements can be relaxed by minor modifications to the structure or component placement to avoid particularly sensitive resonances. Addition of a thermal shield, addition

of a heater, or placing the component in a more benign location may afford a significant relaxation on thermal excursion and thermal cycling requirements.

5.4.3 Flight Qualification of Electronics and Opto-Electronic Subsystem

A number of different standard procedures and test methods have been established from which to draw upon in establishing a flight qualification program. These procedures and test methods have been adopted from the military standard (MIL-STD) test procedures [103]. Though these standards are generally provided for testing electronics components, many of them are equally applicable to optoelectronic components. A basic list of these is provided below.

5.4.3.1 MIL-PRF-19500. MIL-PRF-19500 is a generalized set of performance requirements for qualification of semiconductor devices, and they are therefore applicable to semiconductor detectors and lasers. These test requirements were developed for qualifying military components that often have very different specifications and requirements, and so much of the test flow is not applicable. It calls out five different 'quality' levels.

MIL-PRF-19500 does not actually describe the tests, but rather the sequence of tests to be applied and allowable lot failures for qualifying parts. It refers to MIL-STD test procedures, generally MIL-STD 750 tests for semiconductor devices, indicating which tests should be conducted under various circumstances, and in which order the tests are to be applied. These test procedures are quite extensive for electrical systems, but limited for application to electro-optical devices. For the testing of detectors, one needs to supplement these tests with additional optical characterization to assure adequate radiometric performance. Such tests would likely include spectral responsivity, total responsivity, response uniformity, temporal response, dark noise as a function of temperature, responsivity as a function of temperature, and detector isolation and crosstalk, to name but a few. The ultimate selection of tests should be decided based on the required operational characteristics of the detector.

5.4.3.2 MIL STD 750. MIL STD 750 is a comprehensive list of tests and procedures for general semiconductor devices. The list covers environmental, mechanical, digital electrical, and linear electrical tests and procedures. Only certain tests from this suite are applicable to detectors, detector materials, and diode lasers, which are similar to MIL STD 883.

5.4.3.3 MIL STD 883. MIL STD 883 is a comprehensive list of detailed descriptions of tests for electronic microcircuits. This standard covers tests for a wide variety of environments (e.g., space, high altitude, land, underwater),

mechanical tests for material and construction integrity and vibration and shock environment survivability, digital electrical tests to exercise and verify the operation of digital electronic devices under various electrical conditions, linear electrical tests to characterize the operating performance and range of electronic devices under various electrical conditions, and test procedures. Most of these tests are only applicable to certain types of laser diodes, detectors, or packages. It is up to the qualification engineer to select tests from this suite that are applicable and of significant consequence for the expected environment.

5.4.3.4 Telcordia. Telcordia is a source of testing procedures particularly relevant to opto-electronic systems [104]. Telcordia tests were developed to establish a certification standard for terrestrial fiber-optic-based telecommunication components. The procedures rely heavily on MIL-STD-883 tests but are tailored for terrestrial-based opto-electronic components to ensure the standardization and reliability across the entire telecommunications infrastructure. Telcordia standards are the evolved set of qualification standards originating under Telcordia's previous name of Bellcore.

In general, parts that are Telcordia qualified have already passed some of the stringent qualifications procedures. They have a very extensive application legacy for mission life assurance and operation under most of the conditions required for flight-qualified hardware. Radiation testing and outgassing (the materials requirements for space applications) are noticeably lacking in the qualification guidelines.

5.4.3.5 NASA Electronics Parts and Packaging (NEPP). The NEPP program was developed to establish flight-qualification guidelines for electronic parts, including opto-electronic and photonics devices that might be used in free-space optical communication systems [105]. These guidelines are being developed with data generated from multiple testing programs to assess the effects of space environments on recently developed COTS components. The program consolidated the work of previous programs in space environmental testing, and it is currently addressing the relative lack of standards and procedures for the testing of opto-electronic and optical components. Issues of radiation tolerance and parts reliability testing are at the core of the NEPP program, but other issues (such as low temperature exposure, thermal cycling, mechanical shock, mechanical vibration, and aging) are also considered.

5.4.4 Number of Test Units

Clearly, successfully testing more parts results in higher confidence levels in the selected flight units. The problem is, how many units must be tested to achieve the level of confidence required by the mission?

The probability of failure of n units out of a sample of N units is governed by the binomial distribution, and is calculated by

$$P_f = \binom{N}{n} p^n q^{N-n} \tag{5.4-1}$$

where

$$\binom{N}{n} \equiv \frac{N!}{n!\,(N-n)!} \tag{5.4-2}$$

In this notation, N is the number of units sampled, n is the number of units that fail the test, p is the probability of failure of a randomly chosen device, and q is the probability that the device does not fail. The difficulty arises in that, in general, the failure probability p is not known *a priori*. To determine the probability of survival or failure (q or p) exactly requires that an infinite number of trials be conducted, because it is only in the limit as $N \to \infty$ that the trials give the exact probability. To account for the uncertainty in estimating the probability from a finite number of samples N, one can apply the estimated probability $\hat{p} = n/N$. Then, assuming a normal distribution from which the sample N was taken, and associating some error ε with the uncertainty in the estimated probability, the confidence level α can be derived for the sample. The derivation of this is beyond the scope of this book, but excellent references are available. The confidence level is shown to reduce to

$$\alpha = erf\left(\sqrt{\frac{N}{2}}\,\frac{\varepsilon}{\sqrt{\hat{p}\left(1-\hat{p}\right)}}\right) \tag{5.4-3}$$

in the condition that the product $Np >> 1$.

At this point, it must be kept in mind that the failure probability of an essential part for which flight qualification is necessary had better be small. This forces a large number $\underline{N}$ of samples to be tested, just to attain a high certainty that the failure probability is close to the estimated value. Because of this, other methods of testing are generally performed.

In practice, devices are often procured and tested in quantities of 5 to 100, based on the experience of engineers with the parts and failure mechanisms on similar parts. It is helpful to have information on failure rates, but again, this requires testing many units, indeed often in the thousands, to establish the reliability that is expected on some missions. Though detector vendors may maintain this information, they are often shy about sharing it. The small sample of devices deemed adequate for testing by the qualification engineer is

submitted to inspections and screening tests, and all parts are assumed by similarity to withstand the same loads across the lot.

5.4.5 Space Environments

The requirements of any space mission must ultimately address the unique demands of operating in the space environment. To qualify a system to operate in a space environment, its constituents, namely the components and subsystems, have to be qualified in the first place. A deep-space optical communication system has to survive the pre-launch, the launch, the cruise, and the prime mission's environmental conditions that may vary widely from one stage to another. The most egregious stresses include high vibrational loads during the launch phase and thruster firings on orbit; high thermal and mechanical loads in orbit due to the extremes of solar heat loads alternating with radiating to extremely low deep space temperatures; radiation-induced damage; and contamination resulting from out-gassing of materials in a high-vacuum environment. The magnitudes of these effects can vary depending on the mission, but most must be considered to some degree for any space-flight system. For example, while the radiation field is particularly strong in a Jovian mission, and temperature cycling is more demanding for an Earth-orbiting system, both radiation and thermal effects must be considered in each of the different missions.

Implicit in the definition of the environmental requirements is the need to consider the required operational lifetime of the system. For most concerns, the failure probability dies off exponentially with operation; outgassing falls with time on orbit as the outgassed material is emitted; thermal cycling problems tend to lead to early failures, but it is rare for failures to occur after many cycles. The exception to this rule is in total accumulated radiation dose, which accumulates roughly linearly with time on station, up to a level where the device no longer operates as required. Finally, the availability of resources on the host spacecraft will have an effect on the overall requirements levied on the system. For example, a large spacecraft in geosynchronous Earth orbit (GEO) will generally be able to provide significantly more power for thermal management, and more mass allocation for radiation shielding than a mass-limited and power-starved deep-space probe. Given the range of missions and above considerations, the following environmental requirements provide a backdrop for detailing the qualification plan for any flight optical communication system.

5.4.5.1 Environmental Requirements. A component, a subsystem, or a system prepared for spacecraft use may experience a variety of environmental effects [106–109]. The environmental conditions with major effects on the instrument and its constituents include:

1) Ionizing radiation (in space)
2) Vibration (handling and launch)
3) Mechanical, thermal, and pyro shock (orbit, launch)
4) Thermal gradients (due to solar irradiation)
5) Depressurization (from humid atmospheric pressure to vacuum)
6) Electromagnetic field emission/susceptibility, coronal arcing, plasma bias
7) Outgassing (cleanliness)

Generally, the environmental requirements are designated at the assembly level and sometimes at the subassembly level. These requirements are often more difficult to attain at the device or component level since the detailed design affects the requirements, and might not be known early in the spacecraft development process. In that case and in the absence of any specific localized information, one may assume the same environmental specifications for the subsystem and component levels, as well. To verify that the environmental requirements are met or can be met, analysis is needed in conjunction with a set of well-defined tests.

5.4.5.2 Ionizing Radiation. The space radiation environment is composed of many different types of ionizing radiation, comprised of large fluctuations in particle density, energetic distribution, and spatial distribution. Fortunately, we on Earth are shielded from almost all of the ionizing radiation by the Earth's magnetic field and our atmosphere. In leaving these protective barriers behind, the damage from radiation exposure of all types must be considered.

Sources of ionizing radiation in space are summarized in Table 5-18. Radiation types include electrons, neutrons, protons, heavy ions, and cosmic rays. Radiation effects range from single event upsets (SEUs) and single-particle damage to displacement damage and dosage accumulation leading to

Table 5-18. Sources of ionizing radiation.

Source	Type and Effect
Earth's Van Allen belts	Electrons, protons in Earth's magnetic field 1,000 to 6,000 km range, seen as low as 100 km
Galactic cosmic rays	High energy (100s of MeV) ions Belts shield LEO satellites
Solar flares	Highly intense protons, electrons, and some heavy ion bursts pump up the Van Allen belts
Fields around other planets and moons	Sulfur and oxygen near Jupiter and Europa Fields as high as 10 Grad
Spacecraft-borne reactors	Effect more pronounced in smaller spacecraft Secondary reactions in shielding may occur

gradual or sudden fatal damage. As one proceeds up from the Earth's surface, the upper regions of the atmosphere no longer provide enough density for the dissipation of cosmic-ray energy, allowing for high-mass cosmic-ray nuclei to be a radiation source. At regions above 20 km, the ozone layer is left behind, and materials must be capable of withstanding degradation from solar ultraviolet (UV) rays. These rays are of approximately the same energy required to bind many materials, especially hydrocarbons. As a result, these molecular bonds can be disrupted by UV, leading to material degradation.

Radiation sensitivity is highly dependent on the environment and the component itself. For example, passive optical elements and fibers are particularly sensitive to most forms of ionizing radiation including electrons, protons, and gamma rays. In contrast, diode lasers, some photodetectors, and some fiber-optic components are sensitive to displacement damage caused by protons and neutrons [110]. Doped fiber and polarization maintaining fiber are usually more sensitive to total accumulated dose of radiation, but not to displacement effects [111]. Germanium-doped (Ge-doped) glasses and rare–earth-doped fibers are particularly sensitive to radiation-induced photo-darkening [112]. Annealing and photo bleaching help recovery from radiation exposure, particularly at elevated temperatures.

Photodetectors, particularly Geiger-mode APDs, are especially sensitive to ionizing radiation [113]. Protons and cosmic rays result in false signals. Protons and neutrons result in displacement damage leading to higher dark current from the photodiodes [110]. In photodiodes with on-chip circuitry, total ionizing dose can be harmful. CCDs are similarly sensitive to radiation effects. However, active pixel sensors (APSs) are more immune to harmful radiation [114]. Also, detectors made of III–V materials (e.g., indium gallium arsenide and gallium arsenide (InGaAs and GaAs)) are less sensitive to harmful radiation effects than those made of silicon [115].

Diode lasers are less sensitive to ionizing radiation and more sensitive to displacement damage from protons and neutrons [109,115]. GaAs-based devices are robust with respect to typical radiation doses present in LEO environments with photo-bleaching possible in high power devices [116]. Laser active elements (e.g., Nd: YAG) are not very sensitive to radiation; whereas, doped fiber active laser or amplifier mediums are more sensitive to radiation [117]. Visible-region (e.g., 800 nm) fiber and fiber-optic components have a lower damage threshold than the near-infrared (1550 nm) fibers. In general, fibers and detectors do better at longer wavelengths. Modulator materials (e.g., lithium niobium oxide ($LiNbO_3$)) have relatively higher damage threshold [118].

Typically, the reliability and radiation hardness assurance (RHA) documentation is not available for commercial-off-the-shelf (COTS) components. Also, it is difficult to predict the radiation effects without testing. Even with testing, one cannot rely on devices to behave similarly since the

manufacturing process may change from one batch to other. Shielding of components generally helps, but not always. Certain levels of shielding may in fact slow down the highly energetic particles and increase their interaction cross section in the devices being shielded, leading to greater radiation sensitivity. Optics and opto-electronic devices also require a certain unobstructed field of view that may mean shielding is impractical. Manufacturer's packaging practices often make testing more difficult to perform. To fully understand the effects, lids of opto-electronic devices should be removed with direct exposure of the devices to the ionizing radiation. Each of the above effects and their consequences on the devices or subsystems is described below briefly. The possible adverse effects of major radiation effects on components and devices are summarized in Table 5-19.

As discussed earlier, radiation effects range widely among differing locations of space. Even for a given orbit, the radiation levels could vary significantly if it is an elliptical orbit. Two other variables are the amount of shielding and the duration of exposure. For example, the accumulated radiation dosage accumulated over a few years in a Martian environment behind 4 mils (0.1 mm) of aluminum is about 20 krad, while at Europa (a moon of Jupiter), one of the worst radiation places in our Solar System, radiation levels are 1 to 2 Mrad behind 4 mils of aluminum accumulated over one month.

5.4.5.3 Vibration Environment. The vibration environment is the vibration experienced by the subsystems and the system during ground handling and launch of the spacecraft. Adequate levels of design practice, followed by structural analysis and testing on the engineering model are required to ensure survivability. Depending on the launch vehicle used, the sinusoidal or other

Table 5-19. Possible effects of ionizing radiation on components and devices.

Radiation Effect and Type	Adverse Effect
Single event effects (SEE)	
Protons and heavy ions	
Single event upsets (SEU)	Soft failure
Single event latch up (SEL)	Functional and hard failure
Single event functionality interrupt (SEFI)	Recoverable failure
Single event burnout (SEB)	Hard failure in power transistors
Single event dielectric rupture (SEDR)	Hard failure
Displacement damage	Bulk lattice damage to photodetectors, diode lasers, and analog devices
Protons, neutrons	
Total ionizing Dose (TID)	Gradual and cumulative
Electrons, protons, gamma rays	Parametric, sudden degradation, malfunction
Single particle	TID failure of single transistor
Heavy ions	

mechanical vibrations and resonances vary greatly from spacecraft to spacecraft and will have a time-dependent and frequency-dependent component during launch. Figure 5-58 shows an example of typical launch vibration environments.

5.4.5.4 Mechanical, Thermal, and Pyro Shock Environment. Mechanical, thermal, and pyro shock are sudden changes in the environment. Subsystems and components experience shocks from explosive (pyro-activated) release mechanisms during fairing separation or spacecraft separation. Proper design practices followed by mechanical and thermal cycling of a selected number of subsystem or the prototype model under the specified environments and commensurate with the established military specifications (Mil-Specs) and military standards (Mil-STDs) are part of the qualification process. Figure 5-59 shows an example of a typical launch acceleration environment.

5.4.5.5 Thermal Gradients Environment. The temperature of a spacecraft may vary greatly as it travels from Earth to deep space, or the temperature rises and falls during a spacecraft's orbit. Temperature ranges of –200 to 50 deg C may be experienced for deep-space missions. A system's or subsystem's

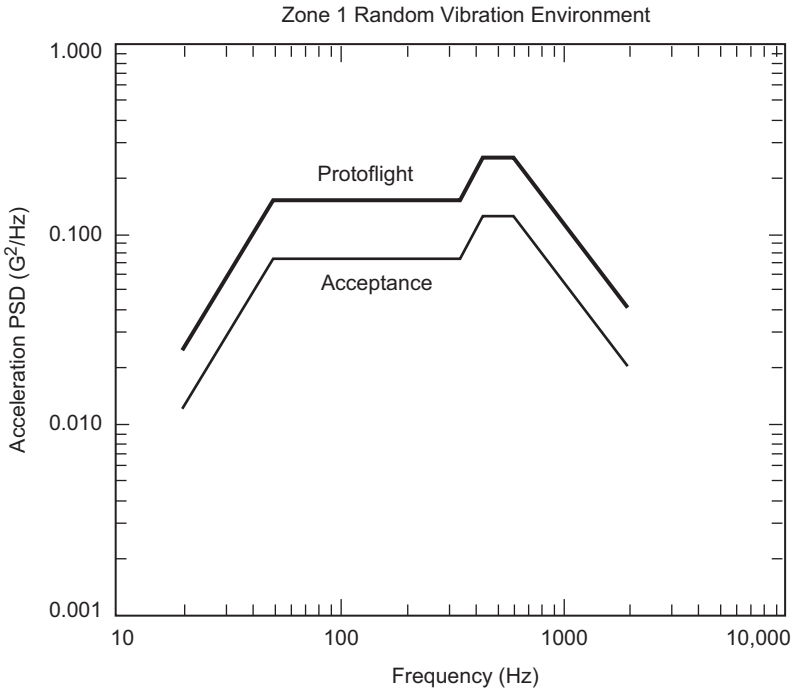


Fig. 5-58. Example of a launch vehicle vibration spectrum and pattern.

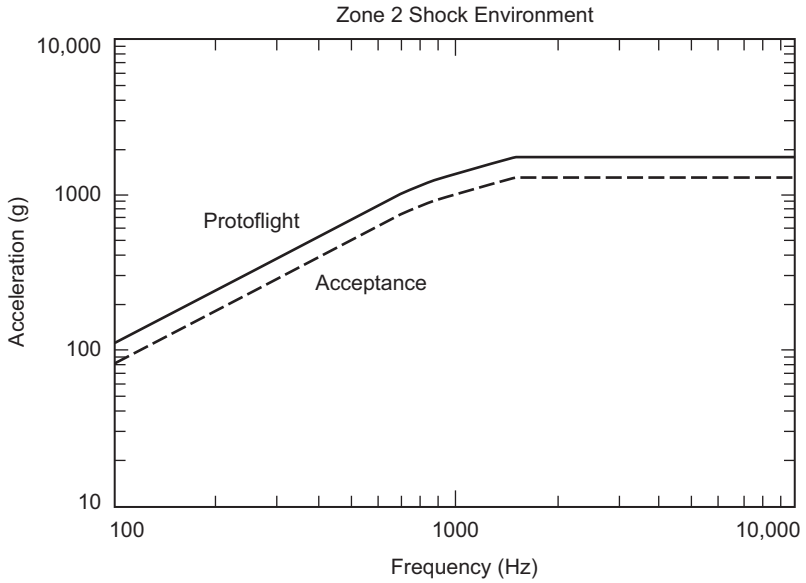


Fig. 5-59. Example of an acceleration profile during launch.

temperature may be maintained to within ±10 deg C if adequate mass and power are available for radiators and heaters. The concern is that thermal cycling and temperature gradients may induce misalignment of optics and laser resonators. Proper design practices, such as, selection of mechanical and optical parts, design of the laser resonator, and qualification at the prototyping stage, are required.

5.4.5.6 Depressurization Environment. During the ground assembly, the system operates under atmospheric pressure and a certain level of humidity. In space, the system should be either kept under pressure and hermetically sealed or operated under vacuum. Special design provisions, such as inclusion of windows, choice of very low outgassing materials, and allocation of vent orifices, should be practiced for each case. Again, vacuum testing at the prototype stage is necessary to insure integrity of optical alignment and functionality of the system. Hermetically sealed components have an advantage and are usually sealed with specific gases to either inhibit degradation (such as from oxidation effects) or act as getters for any impurities.

5.4.5.7 Electric and Magnetic Field Environment. The electromagnetic compatibility (EMC) is of interest primarily to electrical and electronics subsystems. Electric fields, magnetic fields, and electrostatic discharges could potentially pose a threat to the drivers, controllers, and the processor portion of a lasercomm terminal. Table 5-20 summarizes some of the known preventive

Table 5-20. EMC avoidance design practices.

Requirement	Design Practices
Electric field Emissions and susceptibility	Use of twisted shielded wires Control of grounding via single-point chassis reference Conductive mating surfaces between chassis and covers
Magnetic field Emissions and susceptibility	Use of twisted wires for high voltage and current leads Minimization of power and signal control loops Minimization of magnetic material content
Signal and control conducted Emissions and susceptibility	Use of optical isolation for controls
DC power conducted Emissions and susceptibility	Use of signal lead twisting and shielding Use of EMI filters and common mode choke Effective separation of signal and power wiring

design techniques, established to minimize the deleterious effects of electromagnetic interference.

Corona discharges can occur in vacuum between two sharp points held at sufficiently high voltages. Within a laser communication terminal, this can happen with the APD elements, the pump diode lasers, and the electro-optic Q-switcher or cavity dumper. Proper packaging typically avoids this problem.

5.4.5.8 Outgassing. Outgassing from materials used to fabricate the laser and from the surrounding material that can deposit and damage the optical coatings in the presence of strong beams within the laser cavity could be a major lifetime limiter for the laser. Telescope-mirror contamination with non-volatile residue can result in significant reduction of the Strehl ratio leading to major performance loss. Selecting and controlling the outgassing rate of materials and testing on the ground during the qualification process may minimize these. It should be noted that the vacuum surrounding a spacecraft could in fact be poor, with pressures on the order of 10^{-3} Torr (0.1 Pa). Also, there is the potential for particles and plasmas from outgassing and lack of proper cleaning procedures prior to launch as well as improper electrical grounding and charge build-up.

5.4.6 Flight Qualification of Detectors

In almost any electro-optical system, the performance of the detector is critical to the operation of the system, and often to the viability of the mission. Since optical systems are generally designed around the characteristics and performance of the detector, a reliable understanding of the operation, performance, and limitations of that detector is essential over the lifetime of the

mission, and under the various conditions to which the instrument will be subjected.

In this section, we define “detector” as a transducer in which high-frequency electromagnetic radiation (corresponding to visible, ultraviolet, or infrared light) is converted to an electrical signal. Various detectors operate in a variety of ways. Bolometers vary the current flowing through the device as a result of resistance changes in the detector material associated with their rise in temperature when they absorb energy. Photovoltaic and photoconductive detectors operate at a quantum level, promoting loosely bound charge carriers to the conduction band as a result of the absorption of a single photon. Detectors will be employed in sensor systems to perform various functions (Earth tracking, background measurement, signal detection), which will have differing levels of mission-criticality and sensitivity to degradation of performance.

Ideally, the detector performance and operation environment should be explicitly defined for the person responsible for sensor design and operational performance. This information typically comes from a detailed requirements flow down generated by the system engineer. This requirements flow down ideally should take into consideration the mission requirements, conditions, and constraints; and using reasonable estimates and experience, the engineer should develop an initial system design that allocates various performance requirements to each of the associated subsystems. In practice this is an iterative process, in which the system engineer communicates his initial design to the various subsystem engineers, who then respond as to the difficulty and cost of meeting these requirements under the specified conditions. Clearly, the initial design is a single point in a multidimensional trade space. As the system engineer negotiates with the responsible subsystem engineers, the design should quickly converge to a point that can be considered a local minimum in the cost/difficulty function in this trade space. For example, by using a more sensitive detector, the size of the collecting aperture can be reduced, resulting in what is generally a very welcome reduction in the size and mass of the system.

5.4.6.1 Flight Qualification Procedures. There are many different types of detectors which may be called upon to operate in a flight environment, The suite of flight detectors may include photoconductive detectors, bolometers, P-Type/N-Type (PN) or P-Type/Insulator/N-type (PIN)-type photodiodes, avalanche photodiodes (APDs), quadrant APDs, charged coupled devices (CCDs), active pixel sensors (APSs), or position sensing devices(PSDs), to name but a few. The remarkable sensitivity of modern APDs makes them the primary candidate for the receiver on a deep-space-based optical communication system. As the heart of the uplink receiver, the operational characteristics of these detectors must be thoroughly understood and controlled to obtain optimum performance of the uplink receiver system. APD failure

modes include all of those for normal photovoltaic detectors of the same semiconductor material, but the nonlinear mode of operation and the high potentials at which they operate tend to add additional concerns.

Testing of detectors for purposes of flight qualification should encompass tests that, to the degree possible, mimic the conditions and operational configuration of the part in the anticipated environment [118]. In addition to control of the environment, it is obviously most important to test the characteristics of the devices that the operational system will rely on. For example, though detector responsivity is frequently specified, in some circumstances the responsivity is less important than degradation of rise time or a change in operating voltage. This would be the case in a system that uses the detector primarily to record the timing of an event (such as the position of a communication pulse in a pulse-position-modulation format system, or range measurements in a laser remote-sensing system) rather than the radiometric intensity of the event. Of course, the system must still have the sensitivity to detect the pulse, and responsivity degradation will have some (generally minor) effect on the pulse timing.

Critical parts intended for long-term operation in space environments must pass a complicated and expensive screening flow that verifies that the part obtained will operate as required for the duration of the mission. The flight qualification process typically begins long before the parts are produced: the vendor corporate structure, manufacturing facility, process line, and operations are reviewed and inspected to certify that they adhere to good manufacturing, packaging, and testing processes [119]. Beyond the initial inspection, vendors and process lines must submit to periodic re-inspections to maintain their certification.

The first selection process for detectors begins at the wafer level (Fig. 5-60), where the material characteristics (resistivity, current/voltage (I–V) curves, etc.) are screened. This is intended to identify material that, through experience, tends to produce parts with the best operating characteristics and yield for space applications. Once good wafers have been identified and characterized, the fabrication facility makes use of these selected materials to fabricate detectors, the aggregate of which are referred to as a “lot.” From this point on, documentation on each individual part must be maintained. “Lot traceability” is required to amass the characteristics of the particular lot that, through statistical association, will be used to assess other devices coming from the same lot. It is usually advantageous to have devices from at least two lots, in case subsequent testing identifies a problem with a lot.

Each part must be electrically screened to assure that it meets basic operability. On a detector, this screen should consist of a current-voltage (I–V curve) measurement performed within the environmental (temperature, humidity, etc.) range of the final application. At this point, the entire population of detectors should be broken into groups that will undergo different types and

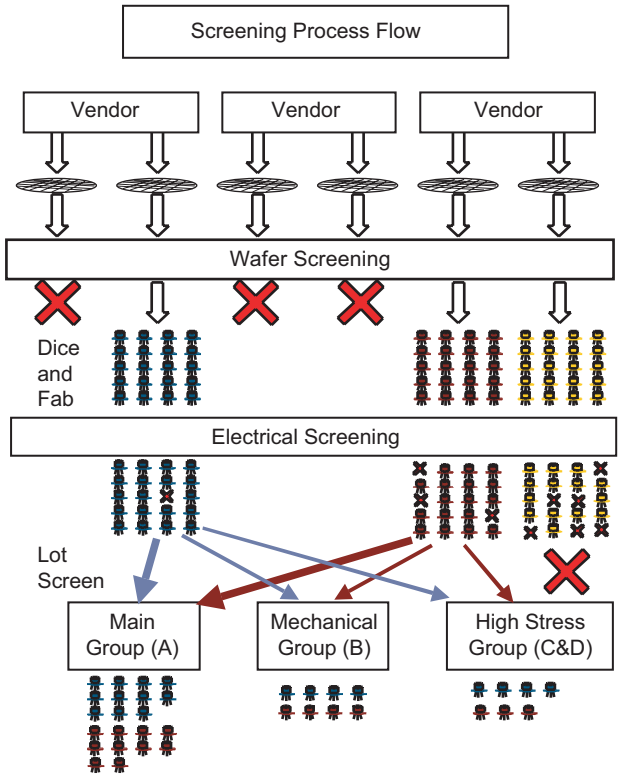


Fig. 5-60. Detector selection process (crosses signify rejected parts).

levels of testing. Each of these groups should contain samples from each of the different material lots used. The size of each of the groups will depend on the typical yield of such devices through the various qualification sequences. This will also determine the confidence level of the final parts chosen as flight units. Generally the largest group (Group A) containing roughly 2/3 of the individual parts will be the group from which the flight samples are drawn. Each of the detectors in this group will be subjected to a series of tests designed to eliminate non-performing or poorly performing parts, and to cause the failure of parts likely to fail early in operation through normally non-destructive tests. The second group of detectors (Group B) will be subjected to more stressing mechanical tests, designed to evaluate the limits of the parts to survive various stressing environments. Parts from this group will be subjected to conditions that can be expected to degrade their integrity and operability, and thus, the parts in this group should not be used for flight candidates. The third group (Group C) will be subjected to more destructive tests; these detectors will definitely not be candidates for flight parts. A fourth group (Group D) is usually included for space flight qualification, and includes parts from each lot that are

to be subjected to radiation testing. These parts can often be parts that have survived the Group C testing, reducing the total number of parts required. Qualification Flow of Main Sample Group (Group A) is shown in Fig. 5-61.

The first detector inspection is a simple visual screen, in which a technician visually checks the characteristics of a device, such as whether parts are properly aligned, if wire bonds are neatly formed, and if the detector materials or coatings are free from blemishes, residual manufacturing debris, etc. A comprehensive list of identifiable defects can be found in MIL-STD-883E 2008–2009. An experienced technician can identify such defects within seconds, making this a relatively high-throughput test with the ability to save a significant amount of time, effort, and cost in subsequent tests through early detection of defects and elimination of bad parts. Immediate feedback from the visual inspection will also assist the manufacturer in identifying the characteristics of parts coming from particular manufacturing equipment, processes, or personnel, who will quickly improve the overall manufacturing yield.

The parts that pass the visual inspection should be subjected to the first level of environmental tests, which would typically include an initial stabilization bake, the first round of low-level thermal cycling, and constant acceleration tests. These procedures are designed to cause the early failure or "infant mortality" of substandard parts. It may not be readily apparent after these procedures that a part has failed. This is commonly determined by the

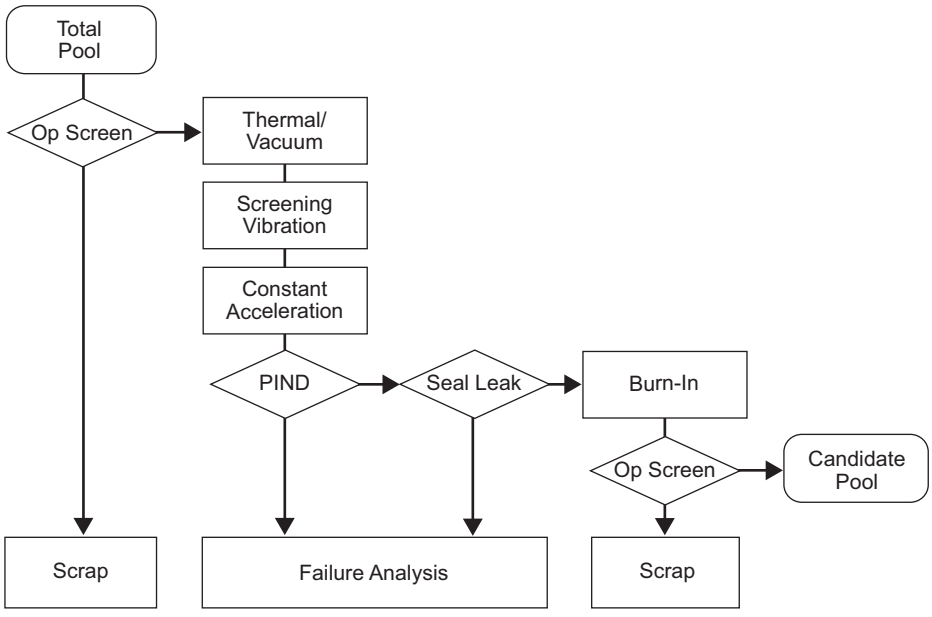


Fig. 5-61. Qualification flow of main sample group. PIND stands for particle impact noise detection.

particle impact noise detection (PIND) test and the fine and gross leak test. In the former, a microphonic transducer is attached to the detector, which is then subjected to vibration with acceleration of 10–20 g at 40–250 Hz. The transducer can then detect the impact of material that has been weakened or dislodged by the previous procedures, even though that material may not be visible through the detector window, or may appear to be attached. For top-level qualification for spacecraft parts, the lots or sub-lots are subjected to up to five PIND test runs. If the accumulated failures exceed 25 percent of the lot, the entire lot is rejected. Furthermore, if each particular PIND run continues to have failures after the failed devices were removed from the previous test, it indicates a continuous low-rate failure potential, and the entire lot is rejected. Only after a lot PIND test is run with failures of less than 1 percent can the lot be accepted.

The fine and gross leak test (MIL-STD-750D 1071) is used to verify the seal of the housing unit, which may have broken during the heating cycles, acceleration procedure, or PIND test. This test subjects the parts to pressure differences, during which gas flow from the sealed part is detected, or flow of gas or liquid into the part is observed. Various types of materials, including radioactive tracers, fluorocarbons, and dyes are used to detect leakage, depending on the type of part, volume of the enclosure, type of seal and detector characteristics.

Detectors that pass to this point are then nominally characterized electrically, principally through a measurement of the I-V curve, and detector responsivity under nominal operating conditions. An example of I-V curves for a photovoltaic detector is shown in Fig. 5-62. The top curve represents the un-illuminated detector; at low levels of forward bias current flows freely, whereas there is a limit to the amount of reverse-bias current that can be driven in the nominal operating voltage range. As the detector is illuminated, absorption of photons in the vicinity of the junction generates charge carriers that are swept away by the junction bias.

Damage to the detector will frequently show up at this stage of the testing, either through an open circuit, in which case little or no current will flow, regardless of the voltage, or a short circuit, in which excess current will flow, bypassing the detector junction. Short circuits are easily identifiable by observation of a change in the I-V curve as shown in Fig. 5-63.

Other types of damage or defects will present as a change to the detector responsivity. These tests can either by performed by making another I-V measurement under illumination conditions, or by simply operating the detector at a fixed set of nominal operational conditions (proper bias, temperature, and illumination). Significant variations (typically 2 dB or 37-percent variation from baseline) in detector response from the initial measured response should disqualify a detector from consideration.

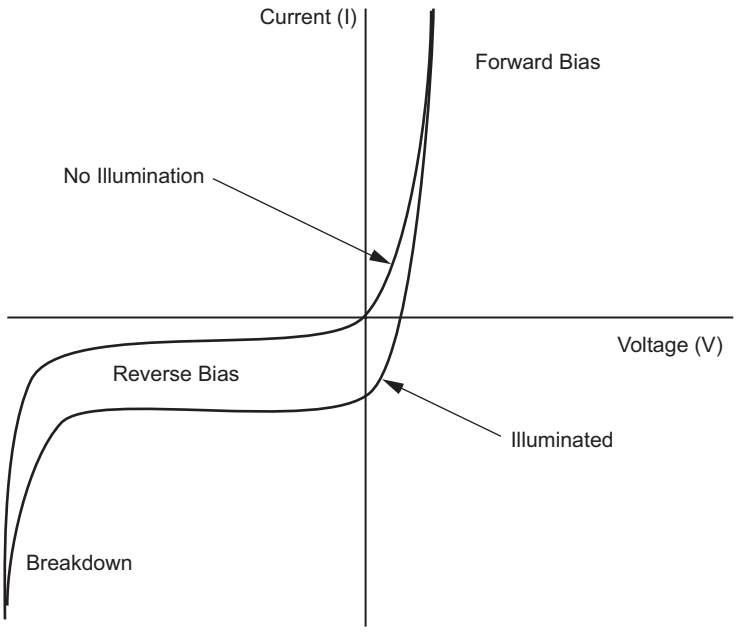


Fig. 5-62. An example of I-V curves for a photovoltaic detector.

The next procedure for the lot parts is generally a high-temperature burn-in, designed to induce the failure of time-dependent defects in materials and construction. The burn-in is generally performed under bias conditions for a period of at least 48 hours, and the part must be subjected to the suite of electrical tests within 24 hours of the burn-in procedure to prevent annealing from correcting any burn-in degradation. After the electrical tests, a second burn-in is performed, and electrical performance is characterized once again.

At the end of this testing, the fraction of detectors within the lot that continue to meet the specifications is calculated. This is referred to as the percent defective allowable (PDA) calculation. If the fraction of the lot that has failed exceeds a particular value (determined ultimately by the reliability and lifetime requirements) the entire lot is rejected. Otherwise, the detectors are subjected to a final suite of testing, designed to verify quality of the stock and to characterize the remaining electrical parameters. Failure of any part at this point simply eliminates it from further consideration, though the data is not used to disqualify the lot. Junction capacitance is measured, as well as spectral response (at the ultimate operational wavelength, if known). A final leak test is performed to verify hermeticity of the housing seal. The parts are then inspected, both through X-ray inspection (to identify internal mechanical defects) and an external visual inspection. The parts that pass these remaining tests are considered useful parts for stock, from which flight parts can be drawn.

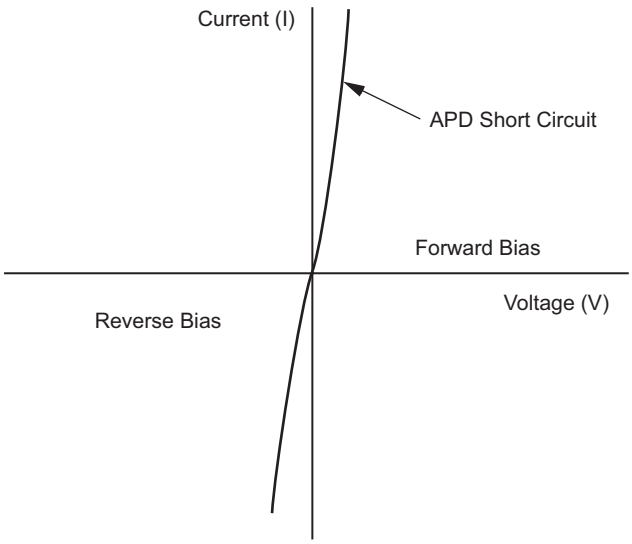


Fig. 5-63. Damage to the detector as evidenced by change in the I-V curve.

A flow diagram for the Mechanical Testing Group (Group B) is shown in Fig. 5-64. The following tests are performed on one of the smaller sample groups (Group B) culled out previously, and they generally consist of more rigorous, stressful mechanical tests unsuitable for parts that may eventually fly in space. Failure of these tests does not necessarily result in the automatic failure of the flight lot, but they are intended to generate information on the ultimate limits of the mechanical stresses to which the parts can reasonably be subjected and to identify unknown failure modes of parts.

This group is itself subdivided, with the majority of the parts going to the thermal-mechanical tests. Electrical testing (I-V curves, responsivity) is performed on these detectors at various intervals to track the device performance, similar to the testing used in the main qualification group. After characterizing the detectors, they are further subdivided, with most of the parts going to intermittent-operation lifetime testing. These tests consist of repeatedly turning the device on, and once the part has reached a stable temperature, switching the device off again. This test is intended to cause failures that would normally result from the electrical and thermal transients associated with normal operation. The remainders of the parts are subjected to thermal cycling, and then to thermal shock testing. This latter test subjects the parts to very rapid temperature changes, usually by immersion in liquids (typically water or perfluorocarbons) at various temperatures. The use of liquids speeds the thermal transfer, stressing the part beyond its maximum expected thermal change rate. This is designed to cause the failure of parts that may fail through mechanical stresses induced from large thermal expansion coefficients, or

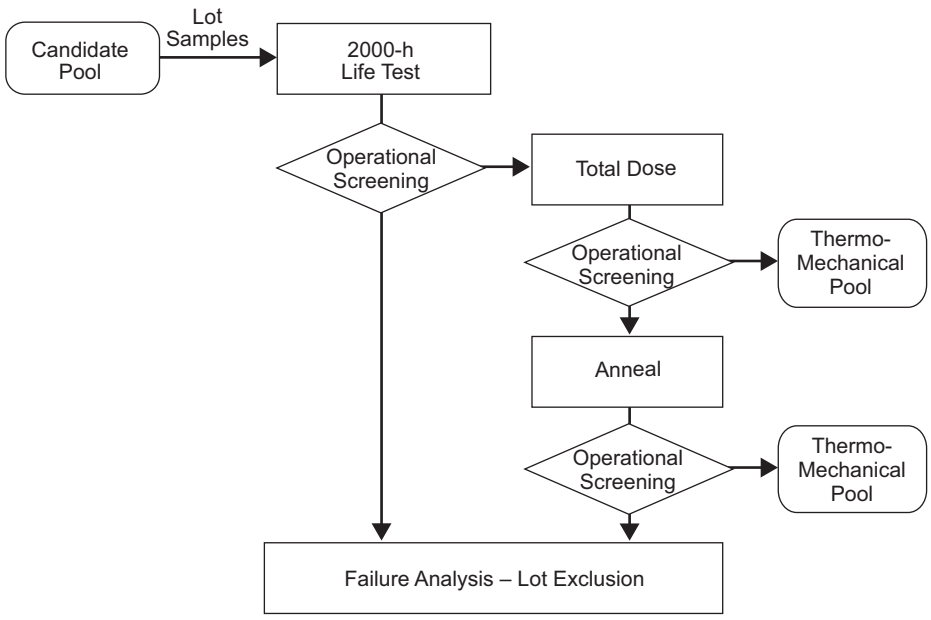


Fig. 5-64. Flow diagram of the mechanical testing group.

mismatches in thermal expansion coefficients. Electrical testing is used to determine failures that are not obvious to the testing personnel.

Some parts that pass the thermal cycling and thermal shock tests may then be subjected to destructive testing. The housing is removed, and a visual inspection is performed, principally to determine the previously undetected deleterious effects of thermal shock testing. Wire bonds are subjected to standard bond-pull testing, and integrity of the mounting of the detector is verified with a die shear test. Because of the destructive nature of these tests, they are clearly not available for final electrical characterization.

The remaining parts in this group (which may consist of rejects from previous tests) are subjected to a final set of mechanical tests. This set includes the measurement of the physical dimensions of the various parts, especially the sensitive surface area of the detector, the solderability of the part, and the solvent resistivity of the part. Solderability is intended to verify the integrity of the part after a typical soldering operation is performed, during which the temperature of soldering leads is quickly elevated to a high level, and subsequently dropped quickly at the end of soldering.

Figure 5-65 shows a flow diagram for the High-Stress Test Groups (Groups C and D). The remaining parts from the initial lot division are separated into high-stress test Groups C and D. These groups consist of detectors that are used to verify the performance degradation of detectors subjected to excessive mechanical shocks and ionizing radiation. The tests in

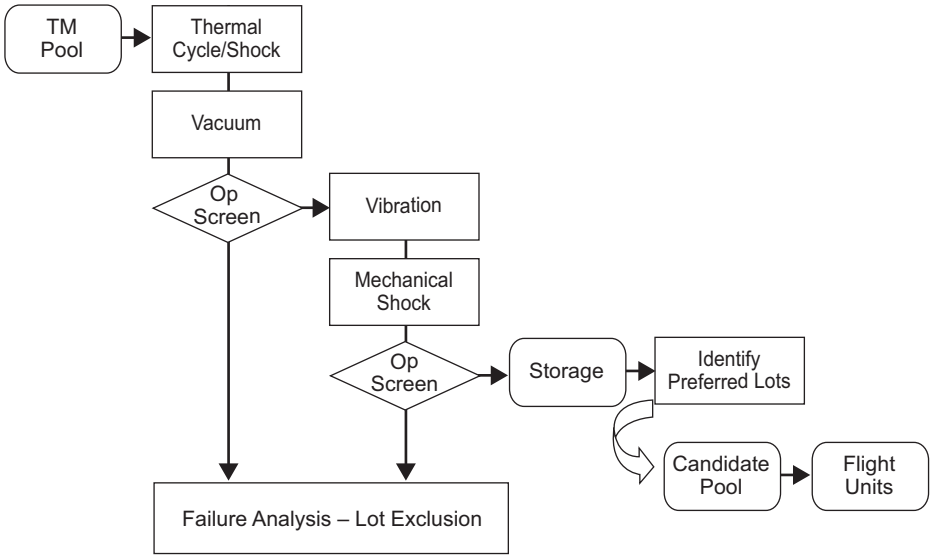


Fig. 5-65. Flow diagram for the high-stress test groups.

these groups again begin with electrical characterization, to establish a baseline against which detector degradation will be measured. Some of these detectors are subjected to high-stress mechanical tests, including the shock test, vibration testing, and constant acceleration testing. The first of these is designed to verify the integrity of the part that may be subjected to rather rough handling during integration and subsequent installation in the launch vehicle. Vibration testing is performed to establish the integrity of the part subjected to vibrations exceeding the amplitude, frequency, and duration expected to be experienced during launch. The amplitudes and frequencies are dependent on the type of launch vehicle to be used, and the subsequent effects of the structural-mechanical housing and support assembly. Because of the effects of damping and/or amplification of resonances in the surrounding support assembly, the expected vibration amplitude may be quite different from that of the spacecraft itself. Modeling and/or testing of the mechanical enclosure in order to derive the test requirements for this test should verify these levels. Finally, the parts should be subjected to a constant acceleration test, in which the ability of the part to withstand the acceleration of launch is verified. After the test detectors are subjected to these three sets of conditions, performance is once again verified by electrical testing.

5.4.6.2 Detector Radiation Testing. Space-based optical communications systems are expected to perform reliably in space for a certain required time period. Once in space, the detectors are typically subject to ionizing radiation that can degrade detector performance. The degradation mechanisms are still a

matter of intense research, but it is known that both responsivity and noise in certain opto-electronic detectors can be adversely affected by both the total accumulated ionizing radiation dose to which the detector is subjected, as well as the rate at which these doses are delivered [106].

The true radiation environment to which the detector is subjected is very difficult to calculate, and must take into account the varying spectrum of ionizing particles to which the detector will be subjected and the shielding of the detector resulting from the placement of various materials in and around the spacecraft. One might naively expect to test to an upper bound by merely using the conditions of the space environment, without the effects of shielding. Whereas this often will present a worst case, shielding can also slow down faster particles and thereby increase the probability that they will be absorbed in the photosensitive material.

A thorough analysis begins with the orbit (or alternately the deep-space trajectory) of a spacecraft and the anticipated energetic spectra of protons, electrons, neutrons, alpha particles, etc. Energetic protons are often the major consideration because of their tendency to be trapped in the Earth's magnetic field and the relatively high energies that they are capable of depositing into materials.

For interplanetary spacecraft, the type of radiation encountered during the long trip may dominate the radiation profile, or the radiation encountered at the final point of study may dominate. For example, during interplanetary flight, there may be relatively high probabilities of solar flares that can emit enormous volumes of highly energetic material into the interplanetary medium. Missions to Jupiter are also typically subject to high radiation doses because the planet's strong magnetic field traps extreme densities and energies of ionizing particles.

Example: As an example, consider the radiation testing required for flight qualification of a silicon avalanche photodiode (APD). The high operating voltage and high sensitivity of the device make it particularly difficult to qualify. Furthermore, the requirements for maintaining bias during radiation testing and subsequent optical characterization make it a particularly good case study.

Initially, the radiation effects engineer levies a requirement for the part, which takes into account the planned orbit of the spacecraft and models the attenuation effects of shielding, including spacecraft structure, materials, and instrument placement. For the particular case of interest, the requirement of 10 krad (Si) was levied for a total accumulating dose in the environment, and 51 mega-electron volt (MeV) protons were specified as the radiation type. This total dose level drove nominal testing to 20 krad (Si) for engineering margin. Furthermore, because it was hoped that the unit could be used on future Mars missions with longer operating lifetimes and slightly higher total accumulated doses, the decision was made to extend testing to a cumulative 40 krad (Si).

The method of qualifying the flight parts is to test similar parts, subjecting them to similar environments to those expected by the flight part. In this case, three Si APDs were subjected to a direct 51-MeV proton beam from a cyclotron. Prior to irradiation, each APD was optically characterized in a portable chamber of the design shown in Fig. 5-66. A Nd:YAG laser of the same wavelength as the flight laser (1064 nm) was used as a light source, and it was directed into an integrating sphere for polarization, randomization, and generation of a uniform extended radiance source. A short distance from the output port of the integrating sphere, two APDs were affixed side-by-side, each separated from the centerline of the integrating sphere output by small, equal distances.

At one APD port, a reference APD was used to verify output continuity from the integrating sphere. At the second APD port, the test APD was subjected to an identical optical input. Baseline current-voltage measurements were made on both APDs, both with and without laser illumination. Subsequently, radiation dosing was applied to one of the two APDs, and new I-V curves were obtained at cumulative levels of 5 krad, 10 krad, 20 krad, and 40 krad. Because the APD was expected to be biased throughout the mission, a bias voltage was maintained on the detector constantly throughout testing. The nominal bias voltage during irradiation was 275 V, at the lower end of the I-V curve voltage measurements. The upper voltage was typically the point at which avalanche breakdown was observed. The detector was not allowed to remain at breakdown for more than a few seconds at any time.

In this particular application, the APD was to measure the time of flight of a laser pulse, so the detector's temporal response was also an issue to be tested. To measure this, a separate chamber was designed in which the pulse from a passively Q-switched microchip laser at 1064 nm could be observed with the APD. A fast oscilloscope was used to measure the pulse rise time and fall time at different total irradiation dose levels.

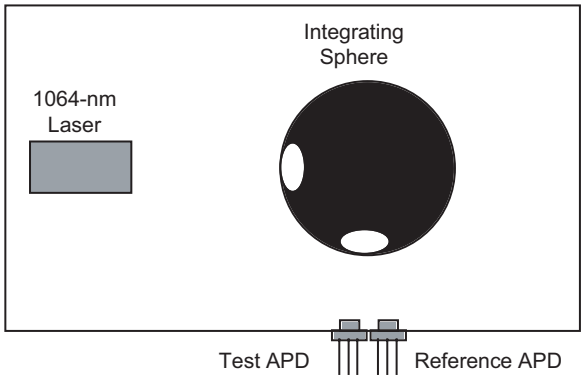


Fig. 5-66. APD optical characterization chamber.

Summary: Flight qualification of detectors is still very much an art, and draws heavily upon the experience and judgment of the cognizant engineer in selection of parts, selection of tests to be performed, and risk-tradeoffs in assigning test levels, test sequences, and numbers of parts to be subjected to tests. While the application of rigorous statistical methods can be used to derive confidence levels and assign numbers of detectors to be subjected to various tests, the relative success of the "art" of flight qualification, and especially at reduction in cost and schedule, prove the value of this more subjective method.

5.4.7 Flight Qualification of Laser Systems

Among the various types of lasers available, it is fortunate that space-based optical communication sources are based on semiconductor diode lasers, either as the seed or pump source. This is not coincidental since the terrestrial communication infrastructure requires similar performance and has invested significantly to ensure ruggedization for long-term reliability. Due to their compact size and power efficiency, diode lasers lend themselves well to communication applications. Other types of lasers such as gas-based systems are being investigated for space applications and bring unique challenges for space qualification [120]. Carbon-dioxide gas lasers have also been investigated for communication sources but will not be considered here for space application.

To develop a comprehensive space-qualification procedure for the laser system, a detailed knowledge of lasers is required with particular attention given to their potential sources of failure. One is referred to the many excellent texts on the principles and applications of lasers [20]. For a given space flight project, the qualification process has to be taken into account in the early stages of mission design along with the performance requirements. Failing to address the environmental requirements and their impact on the device design can lead to lengthy delays and budget overruns as workarounds are implemented late in the mission development cycle to mitigate problems evident during testing.

A process or guideline for qualifying laser systems is outlined in this section. The same rationale given in the above section for detectors can also be applied here. Several past missions have successfully deployed laser systems, some even for communications purposes. However, the qualification process has been somewhat arbitrary based on the individual mission requirements and budgets. As a background, where it is known, the qualification process is described for a selection of the past laser systems flown on different missions. Following that, the design and fabrication of semiconductor lasers is discussed in light of the stringent demands of the space environment. Given their ubiquity in communication sources, recommendations for a fully space qualified design are then given. These design options align with what is currently performed for high reliability in ground telecommunication fiber-optic systems. Prior to

addressing the actual testing flow, it is also helpful to understand the potential points of failure for laser systems. Understanding the various degradation mechanisms for the active diode and packaging along with the fiber interface allow one to tailor the tests, and this leads into the testing flow. The flow-down of tests has been customized from several programs as discussed in the final section.

5.4.7.1 Past Laser Systems Flown in Space. Lasers have been used extensively as sources for a variety of spaceborne instruments [121,122]. However, their qualification has been more arbitrary due to the unique technology involved. A partial list of lasers flown in space or being qualified for flight, along with a brief description of characteristics for each are given in Table 5-21.

The qualification of the above devices has varied from project to project. What is remarkable is the high degree of successful laser generation in the above missions, lending confidence in the qualification and reliability test procedures developed for each case. Most of the solid-state lasers have been custom built and tested as an assembled system through a series of vibration, thermal cycling, and vacuum tests. The diode-based systems (including the pump laser diodes of the Nd: YAG lasers) have relied on commercial devices upscreened through a further series of tests. Due to the varying mission and spacecraft requirements, it is not feasible to give a generic qualification procedure that covers all laser configurations from past missions. Also, the solid-state lasers can be thought of as optical systems with the inclusion of an active-gain media. Qualification then would be built into the design similar to complex optical systems, and the assembly and test would follow standard procedures for any flight system. The only remaining aspect unique to all the above lasers is the diode laser—either as the pumping source or as the complete transmitter. Hence we will focus on the past qualification procedure for diode lasers in this section. Even with that constraint, there is still no common plan available. As a point of reference, the qualification procedure for a representative number of systems above is discussed.

The TES (Tropospheric Emission Spectrometer) project is a technology demonstration for an interferometer test-bed. The laser is used as a stable frequency source and consists of a non-planar ring oscillator Nd: YAG crystal pumped by an 808-nm multimode diode laser. The reliability of the pump diodes was the key risk so 17 devices were placed in an accelerated 2500-hour life test. The failures were analyzed and traced to specific lots that were avoided for the final flight device. Other qualification tests on the diodes included an extended wafer burn-in, wire-bond pull, and die sheer tests. The complete laser system was then subject to vibration and thermal vacuum tests. The TES is flying on the Aura spacecraft, which was launched in July 2004.

Table 5-21. A partial list of lasers already flown in space.

Laser	Application	Mission/ Instrument	Qualification/ Developer	Key Parameters
Diode	Spectroscopy	Cassini/PIRLS	Commercial JPL	Mid-IR
Diode	Spectroscopy	Mars Polar Lander/DS-2	Commercial	1 μJ, 20 ns, 20 kHz, 890 nm
Diode	Spectroscopy	Mars Pathfinder	Commercial/ JPL	
Diode	Range finder	Rendezvous & Capture	Commercial	Multi-mode, CW
Nd: YAG	Altimetry	MOLA	McDonnell-Douglas	40 mJ, 10 Hz
Nd: YAG	Altimetry	Shuttle Laser Altimeter	MOLA spares	
Nd: YAG	Altimetry	Vegetation Canopy Laser	GSFC	10 mJ, 290 Hz, 5 ns
Nd: YAG	Altimetry	NEAR	McDonnell-Douglas	15 mJ, 12 ns, 8 Hz
Nd: YAG	Altimetry	Calipso	Fibertek	115 mJ, 27 Hz, 24 ns
Nd: YAG	Altimetry	ICESat/GLAS	GSFC	40 Hz
Nd: YAG	Altimetry	Clementine	LLNL	180 mJ, 1 Hz, 10 ns
Nd: YAG	Interferometry	TES	JPL/Lightwave Electronics	Single-longitudinal mode
Diode	Fiber optic	SRTM-Phase Calibrator	Commercial JPL	
Diode	Fiber optic	ISS Network Modules	Commercial	
Diode	Free-space communications	LCE in GEO orbit	Commercial/ NASDA	1 Mbps modulation
Diode	Free-space communications	STRV-II	Commercial (SDL)	1 Gbps modulation
Diode	Free-space communications	SILEX	Commercial (SDL)	50 Mbps modulation

DS-2 = Deep Space-2, GSFC = Goddard Space Flight Center, ICESat/GLAS = Ice Cloud Land Evaluation Satellite/Geoscience Laser Altimeter System, ISS = International Space Station, LCE = Laser Communications Equipment, LLNL = Lawrence Livermore National Laboratory, MOLA = Mars Orbiting Laser Altimeter, NEAR = Near Earth Asteroid Rendezvous, PIRLS = Probe IR Laser Spectrometer, SDL = Vendor, now JDS Uniphase, SILEX = Semiconductor Laser Experiment, SRTM = Shuttle Radar Topography Mission, STRV-II = Space Technology Research Vehicle II, TES = Tropospheric Emission Spectrometer

The diode laser pump bars in the Mars Orbiting Laser Altimeter (MOLA) laser were subjected to numerous electro-optic and screening tests throughout the fabrication process. The wafers were selected based on wavelength as a function of temperature for select devices to ensure a correct match to the Nd: YAG absorption band. After screening the bars visually for cracks, anomalies in metallization, stripe and coatings, the devices were processed from the selected wafers and tested electrically for threshold current, slope efficiency, linewidth, emission uniformity, power at high current and optically for pulse energy, boresight alignment, far-field, spot size, beam divergence, and pulse stability. The devices then underwent a burn-in for 44 hours and were retested with a further 22 hours of burn-in. The acceptance test for the devices was a maximum 10-percent degradation in threshold current, slope efficiency, and emission uniformity after the final burn-in. The MOLA laser transmitter successfully functioned for more than two billion pulses.

The qualification tests of the 10-W arrayed devices consisted of three-axis vibration tests (one sweep axis or one min/axis) with sine excitation corresponding to on-orbit operating conditions and random excitation corresponding to non-operating launch conditions. Optical tests were performed after each axis as well as optical power monitored during the operational test. A thermal vacuum test was then performed to ensure the devices operated correctly under vacuum, in particular the beam divergence was noted and the operating set points determined. The devices were not operated over the temperature extremes, four hours at $T_{\min.\max}$, but just their survivability was determined with three cycles. The only other tests were EMC and EMI tests to ensure the devices were not susceptible to power-line effects.

SILEX (Semiconductor Laser EXperiment) was deployed on a LEO Spot 4 and on a GEO Artemis satellite. The single-mode 850-nm diodes with output power of 100 mW were amplitude modulated around 1–40 Mbps. Commercial packages were baselined with a small sample pre-tested under vacuum with removable windows. A common lot of commercial devices was built and screened with some additional process monitoring. Preliminary environmental testing included MIL-STD-883 for the required shock, vibration, and temperature cycling. Reliability was based on life testing under vacuum up to 2000 hours at room temperature and other samples from three different epitaxial wafers operated at full power up to 3000 hours at 50 deg C [121]. No net degradation was noted. Although initially SILEX operated from a MEO, due to launch problems with the Artemis platform, it has now operated successfully with a GEO to LEO communication link.

In contrast to the above high reliability and thus significant cost programs, the Marshall Space Flight Center (MSFC) Autonomous Rendezvous and Capture experiment used multi-mode laser diodes from a commercial vendor as an illumination source. The devices were fiber coupled, and the vendor selection was based on the diode-to-fiber coupling scheme. A cylindrical lens

epoxied to the diode was deemed more reliable than a spherical ball lens under vibration. Apart from the fiber coupling, there was no individual pre-select of the devices. Several devices were vibration tested. and eight devices have successfully survived two Space Shuttle launches with no degradation. The packages were sealed with epoxy, and thus, they are not hermetic as the epoxy can bleed. Although this was cause for concern, no problem was encountered. The devices were vacuum and radiation tested to LEO type conditions with no problems. Care was taken to ensure the chamber was back filled with nitrogen during vacuum testing. The devices operated successfully on multiple Space Shuttle missions.

5.4.7.2 Design of Semiconductor Lasers for High Reliability Applications. In diode-based laser transmitter systems, the semiconductor diode can be either a low-power seed oscillator or a high-power pump diode used to pump a fiber amplifier or solid-state laser. A space-qualified design would require each step in the manufacture and packaging of the device to be compatible with the spacecraft environmental requirements. Although this is possible to undertake, there is not the commercial market to warrant large-scale production of such devices. However, there exist significant markets for low-cost devices that have a given lifetime in terrestrial application with the ability to replace the devices when the lifetime is exceeded. The difference in this case is primarily in the packaging and mounting of such a device, although rad tolerant does imply some alternate fabrication procedures may be advisable. The diode is grown by Molecular Beam Epitaxy (MBE) or Metal-Organic Chemical Vapor Deposition (MOCVD) layer by layer using, for example, alternating layers of GaAs and AlGaAs with varying Al concentration. There are various techniques for enhancing the reliability of the bare diode, and these techniques have been employed in the commercial devices. For instance, if a wavelength of 808 nm is required, then using a structure without Al in the active layers seems to prolong the device lifetime. Since Al oxidation is one degradation mechanism, avoiding Al in the region of the high intensity optical mode makes sense. Other designs have kept Al in the growth but have used passivated coatings on the facets of the device where the oxide growth would form. Hence, known mitigation strategies have been employed in the commercial devices to increase the lifetime. Once these devices are grown, cleaved, and metallized for bonding, they can be mounted to a submount to allow for heat transfer during operation. Commercial devices are typically not constrained by the large thermal ranges present in space environments and so a soft, or low melting point, indium solder is used to mount the bare diode to a copper or high thermally conductive mount. Due to this low melting point of around 150 deg C, the bond is susceptible to plastic shearing at temperatures approaching the melting point.

In space-based systems that are limited in their cooling budget due to available spacecraft DC power limitations, temperatures approaching this

100 deg C are not unforeseen. Indium is also known to creep from the submount joint up to the active laser region and cause catastrophic optical damage. To avoid these effects a hard solder such as gold/tin (Au/Sn) can be used that bonds the bare diode to a lattice-matched substrate. This process is significantly more complex since the device is now stressed, which can lead to stress birefringence in the optical mode of the semiconductor laser. Terrestrial fiber-based communication systems have used this technique in low power devices where the diode is mounted with the active region away from the mount. However, hard solder mounting of high-power devices at the wavelengths required for pumping solid-state lasers is not readily available.

The remainder of the pump diode package may include optics or fiber (to transmit the pump laser light) that are typically epoxied in place. Although the commercial packages can be hermetically sealed, the epoxies may not meet the out-gassing requirement for operation at elevated temperatures in a vacuum. Replacing the epoxies requires the packages to be reworked, which may not be an available option from companies that strive to meet high volume commercial markets.

Finally, none of the components, including the diode, in the commercial device are designed for high-radiation environments. Fortunately, this can be mitigated by shielding and will not significantly impact the small pump laser diodes packages. GaAs-based devices have shown a low susceptibility to radiation damage under moderate test conditions.

To design a pump laser diode to survive space qualification would then require hard soldering the semiconductor diode with the active side down to a lattice matched high thermally conductive submount, preferably in an hermetic package with no or low outgas epoxies and with sufficient mechanical robustness to ensure long lifetime. The commercial equivalent part can be upscreened by focusing on tests appropriate to the known degradation causes in the fabrication and packaging process, namely the mount and optical or fiber alignment.

5.4.7.3 Degradation Mechanisms. Laser systems for communication sources can be broken down into four main sections:

1) Semiconductor diode for optical pumping or oscillation
2) Fiber or solid state crystal for amplification
3) Optical components for the cavity and coupling to optical fiber
4) Electronics for driving the pump lasers or high-speed electronics to deliver the modulation

Here, we will focus on the degradation mechanisms that are unique to the laser, such as the semiconductor diode lasers and the solid-state gain media (whether in the form of fiber or as a bulk crystal).

Potential laser diode degradation mechanisms are:

- Mechanical: Die shear, wire bond fail, fiber pull—these mechanisms apply to the mechanical mounting of the die to the submount, the electrical connection to the metallized die via multiple wire bonds and the mounting of the optical fiber to the output of the diode laser.
- Metal electrode and solder stability: Soft or hard—a soft solder for the submount to the laser package can diffuse under moderate temperatures and pressures whereas a hard solder may produce instabilities in the mounting process that can stress the diode laser.
- Device dislocations and defects: these relate to non-uniformities in the material composition of the diode laser that can produce high current densities.
- Facet damage: oxidation or catastrophic optical damage (COD) can occur if the facet is not protected.
- Bandgap shrinking: nonlinear current densities near the facet can lead to distortion of the semiconductor bandgap, producing facet heating effects.
- Optical mode quality: in broad-area diodes, non-uniformities in the spatial optical modes can produce filaments that lead to local hot spots in the diode-laser gain region.
- Radiation damage: semiconductor materials are susceptible to displacement damage of ionizing radiation.

Fiber degradation mechanisms apply to passive fiber, where the fiber is just used to route the light, as well as active fiber, where the fiber is doped and acts as a gain media. Photo-darkening is the predominant degradation mechanism and arises from color centers formed from high-energy particles or gamma radiation. Basically, light is attenuated independent of wavelength, but the effect is reversible with high optical intensity able to bleach out the losses. This annealing effect is more effective while the laser is operating rather than when the fiber is cycled on and off. Radiation-induced damage has been extensively investigated in optical fibers to determine the optimum fiber composition that minimizes the induced damage [123]. Other mechanisms, such as stress fracturing of the fiber and outgassing of the jacket materials, have to be accounted for during qualification but will not be presented.

5.4.7.4 Qualification Process for Lasers. A general space-qualification or reliability-assurance procedure is now presented. As indicated earlier in the section on qualification of electronics and opto-electronics, 5.4.3.5, the NASA Electronic Parts and Packaging Program (NEPP) has formulated guidelines for the space qualification of opto-electronic and photonics devices. Here, we provide an overview of a qualification methodology from the fabrication process through product acceptance along with failure modes and test

descriptions. Figure 5-67 illustrates the qualification flow from the NEPP guidelines. However, it assumes the fabrication process can be customized for the mission requirements. Typically, for economic reasons this is not possible.

Reliability methodologies for fiber-optic components have also been detailed as applied to terrestrial telecommunications applications. The important distinction is that in a qualification program, the tests are well defined, and once the test criteria are met the devices pass. In a reliability program, failures are required to quantify the device reliability. The Telcordia documents (5.4.3.4, [104]) really satisfy the qualification aspect, and they have become the industry standard in the acceptance and testing of opto-electronic devices. For active components the Telcordia general reliability assurance requirements are detailed in GR-468-CORE, the fiber amplifier requirements are in GR-1312-CORE, and the passive optical component requirements are in GR-1221-CORE. Combining the NEPP guidelines with the Telcordia reliability assurance performance criteria and test procedures along with the above actually flown laser systems and their qualification processes allows us to develop a good space qualification plan for any diode-based laser system.

The first point to note is that, where possible, extensive use should be made of components developed and tested using Tecordia procedures. This addresses the device packaging principally. If such devices are not available, and assuming the devices have been commercially manufactured, the following summarizes a baseline approach.

Once the performance requirements have been determined, the suitability of the material composition with respect to radiation effects is tested. There may be design options to choose from that are more radiation tolerant. At the same time, an accelerated life test is performed on several samples to ensure that the

Lot Acceptance Test for Die
Die Mounting and Wire Bonding
Nondestructive Bond Pull Test
Pre-Cap Visual Inspection
Serialization and Marking
IR Scan
Seal
Temperature Cycle Test
Vibration
Particle Impact Noise Detection
Temperature/Electrical
DC Measurement
Burn-In
Hermetic Seal Test
Radiographic Inspection
Final Electrical
External Visual

Fig. 5-67. Qualification flow from NEPP guidelines.

design is compatible with the desired reliability. The life test should be at or above the performance level during nominal operation, but accelerated with higher thermal loads to reduce the testing time. Appropriate reliability models are used to translate the testing time to anticipated lifetime at lower temperatures. An example of such a model is the standard Arrhenius model where the lifetime is given by the mean time to failure from the exponential relationship:

$$T = T_o \exp - \frac{E_a}{kT}$$

Here E_a is the activation energy unique for each material system, and kT represents the thermal constant.

A destructive parts analysis (DPA) is also performed to characterize the integrity of the construction of the package. The tests listed in Fig. 5-68 are usually done on a sample basis so that accumulative effects are avoided. The test conditions are listed in MIL STD 883. If the device is fiber coupled, the test should be performed initially so that the device can be operated afterwards to check the degree of misalignment. The fine and gross leak tests, along with the residual gas analysis, are only applicable if the device is claimed to be hermetic. The results are documented. If any failures occur, either the part is excluded or waivers must be obtained from the project if the failure is deemed unrelated to those conditions expected during the mission.

Following the DPA, several representative device samples are chosen to undergo a battery of qualification tests as in Fig. 5-69. Again, there should be separate samples for each test to avoid any cumulative effects of the testing. A larger sample size is usually desired, such as 11 typically used to undergo Telcordia certification for laser modules. However, if the costs for testing and devices are prohibitive, then a sample size of 2–3 per test is sufficient. These devices should be randomly chosen from the same lot if possible to ensure that the results are representative of a particular batch process. If the flight devices come from many or unknown lots, then the same criteria should be used for the qualification testing.

The tests are derived from MIL-STD-883 and tailored to the opto-electronic design, such as those used in GR-CORE-468 for Telcordia certification. PIND or particle impact noise test is a short mechanical shock followed by an acoustic vibration to check for loose parts in a package. The thermal cycle test should cycle the temperature beyond the range that the device will experience in space and with an appropriate number of cycles to gain confidence that no cumulative degradation has occurred. Different thermal rates and number of cycles may be needed with a baseline of 50 cycles and a rate of 2 deg C/min. The vibration and constant acceleration tests should be performed separately on each axis. A sine test is sufficient for vibration testing the part since the

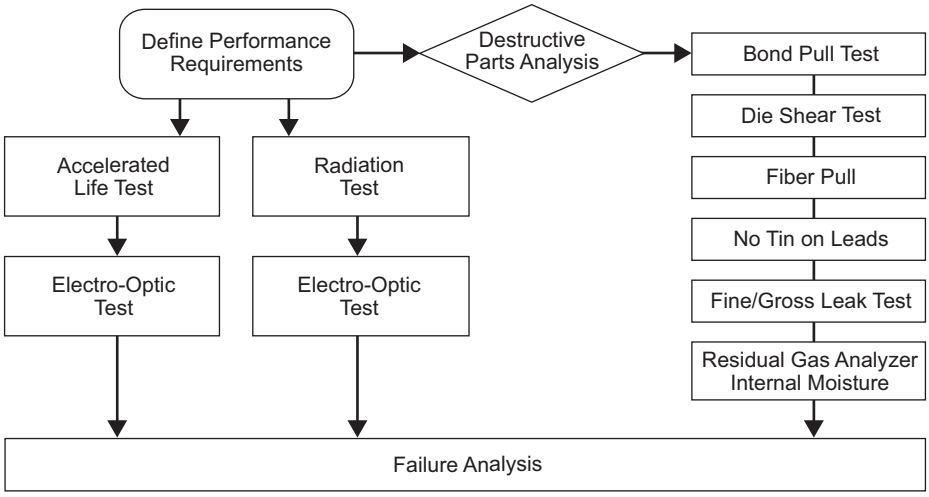


Fig. 5-68. Pre-screening destructive parts analysis.

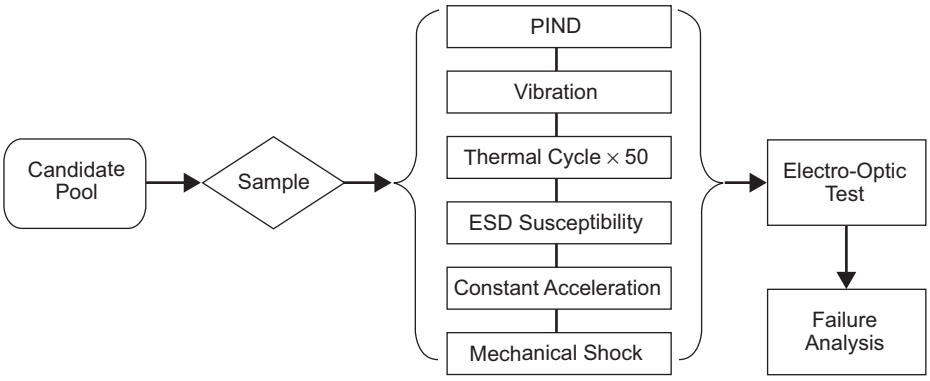


Fig. 5-69. A qualification flow chart example.

assembly usually undergoes a random vibration test at the next level. The electrostatic discharge (ESD) susceptibility test is fairly self-explanatory. All the tests are performed with the devices non-operational for qualification. However, it is important that the device be characterized after each test to measure the pertinent parameter critical to the device. For semiconductor laser diodes, these are typically the output power, wavelength, and perhaps mode quality in certain applications. If any failures are evident from testing, a complete failure analysis should be performed. Based on these findings, recommendations may be made to modify the environmental or performance requirements, reduce the lifetime, go to an alternate design, or continue with a higher level of risk.

The final testing flow shown in Fig. 5-70 provides the samples from which the flight units will be chosen. The candidate pool is the same from which the qualification samples were tested. Serialization basically gives the lot traceability of each device in case any anomalies are evident in the testing. The X-ray or C-SAM (C-mode scanning acoustic microscope) test should only be used if it is known that they will not degrade the device. These tests are used to check for voids and cracks in solder or bonds, to probe the chip attachment process, and to check the overall package integrity. A mini-accelerated burn in of the devices then occurs to mitigate any infant mortality that may be present. Although at elevated temperatures, the time scale needed is only several hours. Finally, a reduced set of thermal cycles, 8–10, is performed with the similar conditions as those used in the qualification testing. Following the 100-percent screening of the candidate pool, a flight device can be chosen along with backups as needed for engineering models, etc. Some results from an example space qualification of pump laser diodes are given in [124].

In summary, we have presented the approach to the qualification of lasers, in particular semiconductor lasers pertinent to a deep space optical communications system. This approach assembles the work of many projects in flying laser systems in space along with the experience of the

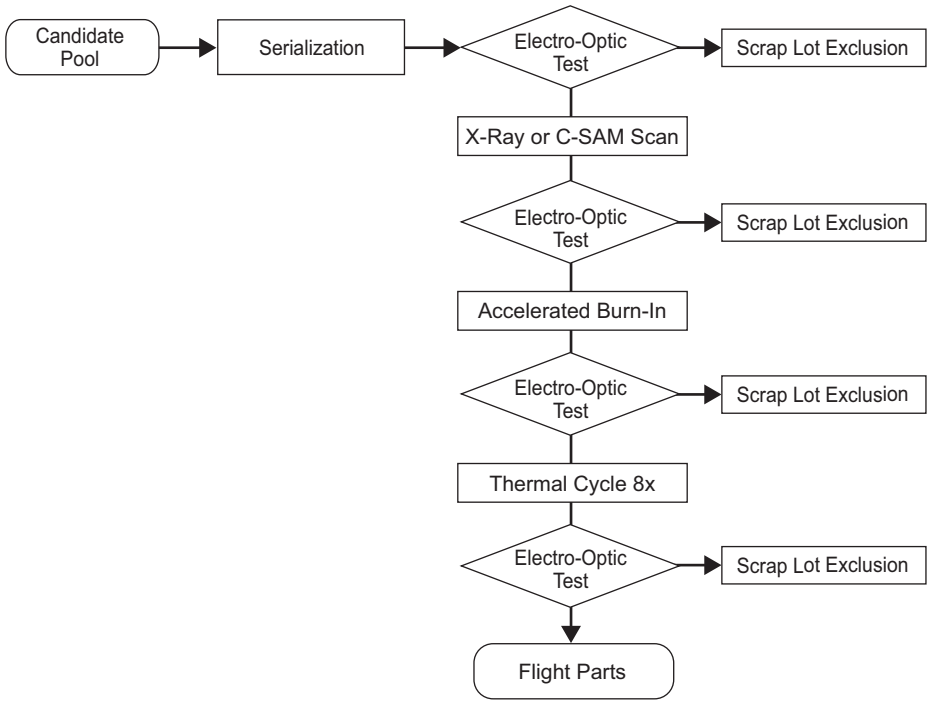


Fig. 5-70. Screening flow for flight units.

telecommunications industry in ensuring long-term reliability of lasers in terrestrial fiber-optic systems. Each project will have its own flavor, driven not only by the budget, but also by the class of mission as well. The qualification approach can thus be tailored accordingly. By highlighting the issues involved, it is hoped that a more robust qualification process can be followed, leading to a more reliable demonstration of the laser subsystem on any space borne applications.

5.4.8 Flight Qualification of Optics

Critical optical elements in an optical communication system are the front transmit and receive aperture (typically reflective), the aft optics (reflective or refractive), those in acquisition, the ATP subsystem, and the communication transmitter and receiver components. Many of these components have successfully flown in space following flight qualification through a rigorous testing procedure. The ATP subsystem's optics are built around a sensitive high-speed camera. Similar cameras are used in a variety of space flight imaging systems and will not be elaborated here.

Typical optical-system degradation mechanisms are known to be photo-darkening of the refractive optics with excessive radiation and misalignment due to lack of mechanical integrity in the mounting.

Qualification issues and cost drivers can be identified early in the mission design phase. For example, cost drivers might dictate use of selected commercial units wherever possible. The design of each unit within the optical system typically includes reliability engineering for vibration, shock, thermal/vacuum, EMC/EMI, outgassing, radiation, and safety [125–127]. Development of test plans and the actual performance of the environmental tests then follows.

To increase system reliability, the optical system development process may include development of different versions of the system: a Brassboard model, an Engineering model, and a Proto-flight model—each with an increasing degree of reliability. Definitions of various models vary from organization to organization. Roughly, a brassboard model includes flight qualifiable parts, without the need for use of qualified parts. The engineering model is a form, fit, and function system and uses partially qualified parts. The proto-flight model is the exact form, fit, and function and utilizes fully qualified parts only. The system as a whole then undergoes reliability testing.

References

[1] N. A. Page and H. Hemmati, "Preliminary Optomechanical Design for the X2000 Transceiver," *Proceedings of the SPIE,* vol. 3615, pp. 206–211, 1999.

[2] C. C. Chen and J. R. Lesh, "Overview of the Optical Communications Demonstrator," *Proceedings of the SPIE*, vol. 2123, pp. 85–94, 1994.

[3] N. A. Page, "Design of the Optical Communications Demonstrator Instrument Optical System," *Proceedings of the SPIE*, vol. 2123, pp. 498–504, 1994.

[4] H. Hemmati and J. R. Lesh "Laser Communication Terminals for the X2000 Series of Planetary Missions," *Proceedings of the SPIE*, vol. 3266, *Free-Space Laser Communication Technologies X*, G. Stephen Mecherle, editor, pp. 171–177, May 1998.

[5] P. R. Spyak and W. L Wolfe, "Scatter from Particulate-Contaminated Mirrors. Part 1: Theory and Experiment for Polystyrene Spheres and lambda = 0.328 μm," *Optical Engineering*, vol. 31, pp. 1746–1756, 1992.

[6] P. R. Spyak and W. L Wolfe, "Scatter from Particulate-Contaminated Minors. Part 2: Theory and Experiment for Dust and lambda = 0.8328 μm," *Optical Engineering*, vol. 31, pp. 1757–1763, 1992.

[7] P. R. Spyak and W. L. Wolfe, "Scatter from Particulate-Contaminated Mirrors. Part 3: Theory and Experiment for Dust and lambda = 10.6 μm," *Optical Engineering*, vol. 31, pp. 1764–1774, 1992.

[8] P. R. Spyak and W. L. Wolfe, "Scatter from Particulate-Contaminated Mirrors. Part 4: Properties of *S*catter from Dust for Visible to Far-Infrared Wavelengths," *Optical Engineering*, vol. 31, pp. 1775–1784, 1992.

[9] T. Weigel, "Stray Light Test Methods for Space Optical Components," *Proceedings of the SPIE*, vol. 2210, pp. 691–699, 1994.

[10] *Airborne Particulate Cleanliness Classes in Clean Rooms and Clean Zones*, available from General Services Administration (GSA) Regulation #FS-209E.

[11] *Product Cleanliness Levels and Contamination in Control Program*, MIL-STD 1246B, Sept. 4, 1987.

[12] H. Hemmati and N. Page, "Approaches for Efficient Coupling of Lasers to Telescopes with Secondary Mirror and Baffle Obscuration," *Proceedings of the SPIE*, vol. 4635, pp. 288–294, 2002.

[13] W. N. Peters and A. M Ledger, "Techniques for Matching Laser TEM00 Mode to Obstructed Circular Aperture," *Applied Optics*, vol. 9, pp. 1435–1442, 1970.

[14] B. J. Klein and J. J. Degnan, "Optical antenna gain—Transmitting antenna," *Applied Optics*, vol. 13, pp. 2134–2141, 1974.

[15] D. Vukobratovich, "Ultra-Lightweight Optics for Laser Communications," *Proceedings of the SPIE*, vol. 1218, pp. 178–192, 1990.

[16] R. Czichy, "Optical Design and Technologies for Space Instrumentation," *Proceedings of the SPIE*, vol. 2210, pp. 420–433, 1994.

[17] A. D. Pillsbury and J. A. Taylor, "Optomechanical Design of a Space-Based Diode Laser Transmitter Assembly," *Proceedings of the SPIE*, vol. 1044, *Optomechanical Design of Laser Transmitters and Receivers*, pp. 112–120, 1989.

[18] R. S. Bondurant, S. B. Alexander, V. W. S. Chan, A. A. Colao, J. E. Kaufmann, E. Lee, A. N. Madiwale, P. F. Martin, A. D. Pillsbury, E. A. Swanson, "An Opto-Mechanical Subsystem for Space-Based Coherent Optical Communications," *Proceedings of the SPIE*, vol. 996, *High Data Rate Atmospheric and Space Communications*, pp. 92–100, 1988.

[19] E. A. Swanson and R. S. Bondurant, "Using Fiber Optics to Simplify Free-Space Optical Communication Systems," *Proceedings of the SPIE*, vol. 1218, pp. 70–82, 1990.

[20] A. E. Seigman, *Lasers*, Chapter 26, "Laser Q Switching," University Science Books, Mill Valley, California, 1986.

[21] H. Hemmati and D. J. Copeland, "Laser Transmitter Assembly for Optical Communication Demonstrator," *Proceedings of the SPIE*, vol. 2123, pp. 283–291, 1994.

[22] R. B. Chesler, M. A. Karr, and J. E. Geusie, "An Experimental and Theoretical Study of High Repetition Rate Q-Switched Nd:YAG Lasers," *IEEE Journal of Quantum Electronics*, vol. 58, pp. 1899–1914, 1970.

[23] W. M. Grossman, M. Gifford, and R. W. Wallace, "Short-Pulse Q-Switched 1.3 and 1-micron Diode-Pumped Lasers," *Optics Letters*, vol. 15, pp. 622–624, 1990.

[24] A. Biswas, H. Hemmati, and J. R. Lesh, "High Data-Rate Laser Transmitters for Free-Space Laser Communications," *Proceedings of the SPIE*, vol. 3615, pp. 269–277, 1999.

[25] R. B. Chesler and S. Maydan, "Calculation of Nd:YAG Cavity Dumping," *Journal of Applied Physics*, vol. 43, pp. 1028–1030, 1971.

[26] L. Marshall, A. D. Hays, A. Katz, J. J. Kasinski, and R. L. Burnham, "High Power Pulsed and CW Diode-Pumped Mode-Locked Nd:YAG Laser," *OSA Proceedings on Advanced Solid-State Laser Conference*, vol. 10, Hilton Head, South Carolina, Optical Society of America, Washington, District of Columbia, pp. 240–246, March 18–20, 1991.

[27] M. Digonnet and C. J. Gaeta, "Theoretical Analysis of Optical Fiber Laser Amplifiers and Oscillators," *Applied Optics*, vol. 24, pp. 333–342, 1985.

[28] A. S. Kurkov, V. I. Karpov, A. Yu Laptev, O. I. Medvedkov, E. M. Dianov, A. N. Gur'yanov, S. A. Vasil'ev, V. M. Paramonov, V. N. Protopopov, A. A. Umnikov, N. I. Vechkanov, V. G. Artyushenko, and J. Frahm, "Highly Efficient Cladding-Pumped Fibre Laser Based on an Ytterbium-Doped Optical Fibre and a Fibre Bragg Grating," *Quantum Electronics*, vol. 29, no. 6, pp. 516–517, 1999.

[29] L. Goldberg, J. P. Koplow, and D. A. V. Kliner, "Highly Efficient 4 W Yb-Doped Fiber Amplifier Pumped by a Broad-Stripe Laser Diode," *Optics Lett*ers, vol. 24, pp. 673–680, 1999.

[30] C. C. Renaud, R. J. Selvas–Aguilar, J. Nilson, P. W. Turner, and A. B. Grudnin, "Compact High Energy Q-Switched Cladding Pumped Fiber Laser with a Tuning Range over 40 nm," *IEEE Photonics Technology Lett*ers, vol. 11, pp. 976–978, 1999.

[31] IPG Photonics Inc., web site, accessed July 11, 2005. (http://www.ipgphotonics.com/).

[32] P. A. Champert, S. V. Popov, and J. R. Taylor, "3.5 W Frequency-Doubled Fiber-Based Laser Sources at 772 nm," *Applied Physics Letter*s, vol. 78, pp. 2420–2421, April 13, 2001.

[33] P. A. Champert, S. V. Popov, M. A. Soladyankia, and J. R. Taylor, "Multiwatt Average Power Continua Generation in Holey Fibers Pumped by Kilowatt Peak Power Seeded Ytterbium Fiber Amplifier," *Applied Physics Letters*, vol. 81, pp. 2157–2159, September 16, 2002.

[34] Catalog of Mitsubishi Cable America, Inc., Ann Arbor, Michigan, 2003, web site accessed July 20, 2005. http://www.mcausa.com).

[35] W. Koechner, *Solid State Laser Engineering*, Springer Verlag, New York, 1994.

[36] W. K. Marshall, K. Cowles, and H. Hemmati; "Performance of Efficient Q-Switched Diode-Laser-Pumped Nd:YAG and Ho:YLF Lasers for Space Applications," *The Telecommunications and Data Acquisition Progress Report 42-95*, July–September, pp.168–173, November 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[37] L. B. Soldano and E. C. M. Pennings, "Optical Multi-mode Interference Devices Based on Self-Imaging: Principles and Applications," *Journal of Lightwave Technology*, vol. 13, pp. 615–627, 1995.

[38] H. Nakano, S. Tsuji, S. Sasaki, K. Uomi, and K. Yamashita, "10 Gb/s, 4-Channel Wavelength Division Multiplexing Fiber Transmission Using Semiconductor Optical Amplifier Modules," *IEEE/OSA Journal of Lightwave Electronics*, vol. 11, no. 4, pp. 612–617, April 1993.

[39] C. C. Chen, D. Robinson, and H. Hemmati, "A 100 Mbps Resonant Cavity Phase Modulator for Coherent Optical Communications," *Proceedings of the SPIE*, vol. 1635, pp. 135–143, 1992.

[40] H. Hemmati, C. Esproles, and R. T. Menzies, "Frequency-Stabilized Diode-Pumped Tm, Ho:YLF Local Oscillator with ±4 GHz of Tuning Range," *Proceedings of the SPIE*, vol. 3380, pp. 75–79, 1998.

[41] J. A. Taylor, A. Pillsbury, M. F. Richardson, and D. Welford, "Diode Laser Transmitter for Space-Based Heterodyne Communication," *Proceedings of the SPIE*, vol. 996, pp. 77–83, 1988.

[42] K. Nielson (Hughes Aircraft), "Active Energy Control for Diode-Pumped Laser Systems Using Pulse-Width Modulation," Patent #5291505, 1994.

[43] H. Hemmati and J. R. Lesh, "A 3.5 W Output, Diode Pumped, Q-Switched 532 nm Laser," *Proceedings of the SPIE*, vol. 2123, pp. 264–269, 1994.

[44] H. Bruesselbach, D. S. Sumida, R. Reeder, and R. W. Byren, "Low-Heat, High-Power Scaling Using InGaAs Diode Pumped Yb: YAG Lasers," *IEEE Journal of Quantum Electronics*, vol. 3, pp. 105–116, 1997.

[45] V. V. Bezotosnyi, K .K. Kumykov, and N. V. Markov, "Ultimate Output Parameters of Laser Diode Bars and Arrays," *Quantum Electronics*, vol. 27, pp. 481–486, 1997.

[46] H. Hemmati, M. Wright, A. Biswas, and C. Esproles, "High Efficiency Pulsed Laser Transmitters for Deep-Space Communications," *Proceedings of the SPIE*, vol. 3932, pp. 188–195, 2000.

[47] D. R. Hjelme and A. R. Mickelson, "Theory of Timing Jitter in Actively Mode-Locked Lasers," *IEEE Journal of Quantum Electronics*, vol. 28, pp. 1594–1606, 1992.

[48] K. Shaik and H. Hemmati, "Wavelength Selection Criteria for Laser Communications," *Proceedings of the SPIE*, vol. 2381, pp. 342–357, 1995.

[49] *Baseline Design Document for Europa-Orbiter Mission: X2000 Optical Communication Subsystem*, JPL D-30264 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 26, 1998.

[50] A. Biswas and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–38, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[51] D. M. Boroson, "Interplanetary Deep-Space Optical Communications Systems and Techniques," in *Special Symposium on High-Capacity Free-Space Laser Communications, 16th Annual Meeting of the IEEE Laser & Electro-Optics Society*, Tucson, Arizona, vol. 1, pp. 85–86, October 27–28, 2003.

[52] H. Hemmati and J. Hilland, *Mars Lasercomm Terminal Accommodation*, JPL D-32921 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 2003.

[53] C.-C. Chen, "Effect of Earth Albedo Variation on the Performance of a Spatial Acquisition Subsystem Aboard a Planetary Spacecraft," *The Telecommunications and Data Acquisition Progress Report 42-95*, July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 202–211, November 15, 1988. http://ipnpr.jpl.nasa.gov/progress_report/

[54] J. W. Alexander, S. Lee, and C.-C. Chen, "Pointing and Tracking Concepts for Deep-Space Missions," *Proceedings of the SPIE*, vol. 3615, *Free-Space Laser Communication Technologies* XI, pp. 230–249, 1999.

[55] JPL mission and spacecraft library website, Jet Propulsion Laboratory, Pasadena, California, site accessed June 24, 2005. http://leonardo.jpl.nasa.gov/msl.

[56] JPL Optical Comm Group Acquisition Tracking Link Analysis Software website, Jet Propulsion Laboratory, Pasadena, California, site accessed June 23, 2005. http://ultratung.jpl.nasa.gov/group6/atlas/atlas.htm

[57] *Accelerometers, DSES Redmond: QA3000 Q-Flex® Accelerometer*, Honeywell web site accessed August 16, 2005. www.inertialsensor.com/qa3000.shtml

[58] J. J. Pearson, D. C. Hines, S. Golosman, and C. Kuglin, "Video-Rate Image Correlation Processor," *Proceedings of the SPIE*, vol. 119, *Application of Digital Image Processing, IOCC*, pp. 197–205, 1977.

[59] C. Kuglin and D. C. Hines, "The Phase Correlation Image Alignment Method," The Phase Correlation Image Alignment Method," *Proceedings of the International Conference on Cybernetics and Society*, pp. 163–165, September 1975.

[60] E. Castro and C. Morandi, "Registration of Translated and Rotated Images Using Finite Fourier Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, pp. 700–703, September 1987.

[61] T. Y. Yan, "Extended-Source Spatial Acquisition Process Based on Maximum- Likelihood Criterion for Planetary Optical Communications," *Proceedings of the SPIE*, vol. 1635, *Free-Space Laser Communication Technologies* IV, pp. 273–285, 1992.

[62] C.-C Chen, E. Hui, and G. Okamoto, "Confidence Range Estimate of Extended Source Imagery Acquisition Algorithms via Computer Simulations," *Proceedings of the SPIE*, vol. 1635, *Free-Space Laser Communication Technologies* IV, pp. 300–308, 1992.

[63] B. E Marino, Haiping Tsou, Tsun Yee Yan, "Earth Beacon Acquisition and Tracking Under Various Illumination Conditions," *Proceedings of the SPIE*, vol. 4272, *Free-Space Laser Communication Technologies* XIII, pp. 190–199, 2001.

[64] B. E Marino, H. Tsou, T. Y. Yan, "Use of Cross-Validation and Monte Carlo Simulation for Estimating the Parameters of Beaconless Free-Space Optical Pointing and Tracking," *Proceedings of the SPIE*, vol. 4635, *Free-Space Laser Communication Technologies XIV*, pp. 84–94, 2002.

[65] S. Lee, G. G. Ortiz, and J. W. Alexander, "High Accuracy and High Bandwidth Pointing Knowledge Using a Star Tracker-Based Acquisition, Tracking, and Pointing System," *The Interplanetary Network Progress Report 42–161*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–18, May 15, 2005.

[66] S. Lee, G. G. Ortiz, W. T. Roberts, and J. W. Alexander, "Feasibility Study on Acquisition, Tracking, and Pointing Using Earth Thermal Images for Deep-Space Ka-Band and Optical Communications," *Interplanetary Network Progress Report 42-155, July–September 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–18, November 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[67] Mars Odyssey web site, Jet Propulsion Laboratory, Pasadena, California, accessed August 10, 2005.
http://mars.jpl.nasa.gov/odyssey/gallery/calibration

[68] W. L. Wolfe and G. J. Zissis, *The Infrared Handbook*, The Infrared Information Analysis Center (IRIA), Environmental Research Institute of Michigan, Ann Arbor, Michigan, pp. 3-52–3-55, 1993.

[69] D. Russell, H. Ansari, and C.-C. Chen, "Lasercom Acquisition and Tracking Control Using a CCD-based Tracker," *Proceedings of the SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI*, G. Stephen Mecherle, editor, pp. 294–303, 1994.

[70] R. B. Deadrick and W. F. Deckelman, "Laser Crosslink Subsystem—An Overview," *Proceedings of the SPIE*, vol. 1635, *Free Space Laser Communication Technologies IV*, pp. 225–235, 1992.

[71] *A Study to Define the Impact of Laser Communication Systems on Their Host Spacecraft, Hughes Aircraft Co., Final Report*, NASA-CR-175272, National Aeronautics and Space Administration, Washington, District of Columbia, April 1984.

[72] C. Chen and James Lesh, "Overview of the Optical Communications Demonstrator," *Proceedings of the SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI*, G. S. Mecherle, editor, pp. 85–95, 1994.

[73] M. Jeganathan, A. Portillo, C. Racho, S. Lee, D. Erickson, J. DePew, S. Monacos, and A. Biswas, "Lessons Learnt from the Optical Communications Demonstrator (OCD)," *Proceedings of the SPIE*, vol. 3615, *Free Space Laser Communication Technologies XI*, G. S. Mecherle, editor, pp. 23–30, 1999.

[74] A. Biswas, M. W. Wright, B. Sanii, N. A. Page, "45 km Horizontal Path Optical Link Demonstrations," *Proceedings of the SPIE*, vol. 4272, *Free-Space Laser Communication Technologies XIII*, G. S. Mecherle, editor, pp. 60–71, 2001.

[75] G. G. Ortiz, S. Lee, S. Monacos, M. Wright, and A. Biswas, "Design and Development of a Robust ATP Subsystem for the Altair UAV-to-Ground Lasercomm 2.5 Gbps Demonstration," *Proceedings of the SPIE*, vol. 4975, *Free-Space Laser Communication Technologies XV*, G. S. Mecherle, editor, pp. 103–114, July 2003.

[76] A. Biswas, N. Page, J. Neal, D. Zhu, M. Wright, G. G. Ortiz, W. H. Farr, H. Hemmati, "Airborne Optical Communications Demonstrator Design and Preflight Test Results," *Proceedings of the SPIE*, vol. 5712, *Free-Space Laser Communication Technologies XVII*, G. S. Mecherle, editor, pp. 205–216, 2005.

[77] S. Lee, "Pointing Accuracy Improvement Using Model-Based Noise Reduction Method," *Proceedings of the SPIE*, vol. 4635, *Free-Space Laser Communication Technologies XIV*, G. Stephen Mecherle, editor, pp. 65–71, 2002.

[78] C. Chen, J. W. Alexander, H. Hemmati, S. Monacos, T. Y. Yan, S. Lee, J. R. Lesh, and S. Zingales, "System Requirements for a Deep Space Optical Transceiver," *SPIE Proceedings*, vol. 3615, *Free Space Laser Communication Technologies XI*, G. Stephen Mecherle, editor, pp. 142–152, 1999.

[79] J. W. Alexander, S. Lee, and C.-C. Chen, "Pointing and Tracking Concepts for Deep-Space Missions," *Proceedings of the SPIE*, vol. 3615, *Free Space Laser Communication Technologies XI*, G. Stephen Mecherle, editor, pp. 230–249, 1999.

[80] R. E. Cummings, V. Gruev, and M. A. Ghani, "VLSI Implementation of Motion Centroid Localization for Autonomous Navigation," *Advances in Neural Information Processing Systems*, vol. 10, pp. 685–691, 1998.

[81] S. Kraemer, R. Downes, R. Katsanis, M. Crenshaw, M. McGrath, and R. Robinson, "STIS Target Acquisition," *HST Calibration Workshop*, Space Telescope Science Institute, Baltimore, Maryland, pp. 39–46, 1998.

[82] J. A. Cox, "Point-Source Location Using Hexagonal Detector Arrays," *Optical Engineering*, vol. 26, no. 1, pp. 69–74, 1987.

[83] J. A. Cox, "Advantages of Hexagonal Detectors and Variable Focus for Point-Source Sensors," *Optical Engineering*, vol. 28, no. 11, pp. 1145–1150, 1989.

[84] K. M. Iftekharuddin and M. A. Karim, "Acquisition by Staring Focal Plane Arrays: Pixel Geometry Effects," *Optical Engineering*, vol. 32, no. 11, pp. 2649–2656, 1993.

[85] I. E. Abdou, "Effect of Signal Truncation on Centroid Location Error Estimation," *Optical Engineering*, vol. 35, no. 4, 1221–1222, 1996.

[86] S. P. Monacos, R. K. Lam, A. A. Portillo, D. Q. Zhu, and G. G. Ortiz, "Design of an Event-Drive, Random-Access, Windowing CCD-Based Camera," *The Interplanetary Network Progress Report* 42–155, *July–September 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–24, November 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[87] *3-V 10-Bit 27 MSPS Area CCD Sensor Signal Processor, TLV987*, Texas Instruments, Dallas, Texas, September 1999.

[88] S. P. Monacos, A. A. Portillo, W. Liu, J. W. Alexander, and G. G. Ortiz, "A High Frame CCD Camera with Region-of-Interest Capability," *2001 IEEE Aerospace Conference Proceedings*, vol. 3, Big Sky, Montana, pp. 1513–1522, March 10–17, 2001.

[89] A. Talukder, J.-M Morookian, S. Monacos, R. Lam, and C. Labaw, A. Bond, "Fast Non-Invasive Eyetracking and Eye-Gaze Determination for Biomedical and Remote Monitoring Applications," *Proceedings of the SPIE*, vol. 5437, *Optical Pattern Recognition XV*, D. P. Casasent, T.-H. Chao, editors**,** pp. 179–190, April 2004.

[90] *680x500 Pixel CCD Image Sensor TC237*, Texas Instruments, Dallas, Texas, June 1996.

[91] G. G. Ortiz, A. Portillo, S. Lee, and J. Ceniceros, "Functional Demonstration of Accelerometer-Assisted Beacon Tracking," *Proceedings of the SPIE*, vol. 4272, *Free-Space Laser Communication Technologies XIII*, G. S. Mecherle, editor, pp. 112–117, 2001.

[92] S. Lee, G. G. Ortiz, W. Liu, and V. Garkanian, "Increasing Tracking Bandwidth for Deep-Space Optical Communications Using Linear Accelerometers," *The Interplanetary Network Progress Report* 42–155, *July–September 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, November 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[93] T. Tolker-Nielsen and G. Oppenhauser, "In-Orbit Result of an Operational Optical Intersatellite Link between ARTEMIS and SPOT4, SILEX," *Proceedings of the SPIE*, vol. 4635, *Free-Space Laser Communication Technologies XIV*, pp. 1–15, 2002.

[94] M. Kayton and W. R. Fried, *Avionics Navigation Systems*, John Wiley & Sons Inc., New York, New York, 1997.

[95] *The Hubble Space Telescope: Optical Assembly*, web site of Goddard Space Flight Center, Greenbelt, Maryland, site accessed, August 20, 2005. http://hubble.nasa.gov/technology/optics.php

[96] M. C. Algrain and J. Quin, "Accelerometer Based Line-of-Sight Stabilization Approach for Pointing and Tracking Systems," *Second IEEE Conference on Control Applications*, Vancouver, British Columbia, Canada, pp. 159–163, September 1993.

[97] H.-H. Chen, S.-C. Lee, and D. D. DeBra, "Gyroscope Free Strapdown Inertial Measurement Unit by Six Linear Accelerometers," *Journal of Guidance, Control, and Dynamics*, vol. 17, pp. 286–290, 1994.

[98] J. G. T. Ribeiro, J. T. P. Castro, and J. L. F. Freire, "Problems in Analogue Double Integration to Determine Displacements from Acceleration Data," *Proceedings of the 15th International Modal Analysis Conference*, Orlando, Florida, pp. 930–934, 1997.

[99] J. G. T. Ribeiro, J. L. F. Freire, and J. T. P. Castro, "Some Comments on Digital Integration to Measure Displacements Using Accelerometers," *Proceedings of the 17th International Modal Analysis Conference*, Orlando, Florida, pp. 554–559, 1999.

[100] J. M. Ceniceros, C. D. Jeppesen, and G. G. Ortiz, "Vibration Platform Testbed for Deep Space Acquisition, Tracking and Pointing," *Proceedings of the SPIE*, vol. 4272, *Free-Space Laser Communication Technologies XIII*, G. S. Mecherle, editor, pp. 209–218, 2001.

[101] M. Wittig, L. Van Holtz, D. El L. Tunbridge, and H. C. Vermeulen, "In-Orbit Measurements of Microaccelerations of ESA's Communications Satellite OLYMPUS," *Proceedings of the SPIE*, vol. 1218, *Free-Space Laser Communication Technologies II*, pp. 205–214, 1990.

[102] G. Ledebuhr, J. Kordas, I. Lewis, M. Richardson, G. Cameron, W. White, D. Dobie, W. Strubar, T. Tassinari, D. Sawyer, M. Shannon, L. Pleasance, A. Lieber, P. Trost, D. Doll, and M. Grote, "HiRes Camera and LIDAR Ranging System for the Clementine Mission," *Proceedings of the SPIE*, vol. 2472, pp. 62–81, 1995.

[103] *Semiconductor Devices, General Specification for*, MIL STD MIL-PRF-19500, Department of Defense, Columbus, Ohio, internet address checked February 8, 2005. www.dscc.dla.mil

[104] Telcordia Technologies, Inc., Piscataway, New Jersey, web site accessed July 20, 2005. www.telcordia.com

[105] NASA Electronic Parts and Packaging Program, www.nepp.nasa.gov, internet address checked February 8, 2005.

[106] C. Barnes, "The effects of radiation on Opto-Electronic Devices," *Proceedings of the SPIE*, vol. 721, pp. 18–34, 1986.

[107] G. C. Messenger and M. S. Ash, *The Effects of Radiation on Electronic System*s, Van Nostrand, New York, 1992.

[108] M. E. Fritz, "Photonics Space Experiment on-Orbit Results," *Proceedings of the SPIE*, vol. 2811, pp. 106–115, 1996.

[109] H. Lischka, H. Henschel, H. Kohn, O. Lennartz, W. Schmidt, "Radiation Effects in Light Emitting Diodes, Laser Diodes, Photodiodes and Optocouplers," *Proceedings of RADECS 93: Second European Conference on Radiation and its Effects on Components and Systems*, Fraunhofer INT, Euskirchen, Germany, pp. 226–231, 1993.

[110] A. H. Johnson, "Proton Displacement Damage in Light Emitting and Laser Doses," presented at *RADECS 2000* (Louvain-la-Neuve, Belgium), Fraunhofer INT, Euskirchen, Germany, September 2000.

[111] G. M. Williams, M. A. Putnam, and E. J. Fiebele, "Space Radiation Effects on Erbium-Doped Fibers," *Proceedings of the SPIE*, vol. 2811, pp. 30–37, 1996.

[112] H. Henschel, O. Koehn, and H. Schmidt, "Radiation-Induced Loss of Optical Fibers at 1300-nm and 1500-nm Wavelength," *Proceedings of the SPIE*, vol. 2811, pp. 68–76, 1996.

[113] H. Lischka, H. Henschel, O. Köhn, W. Lennartz, "Gamma and Neutron Irradiation of Optoelectronic Devices," *Proceedings of Radiation and Its Effects on Components and Systems, 1993, RADECS 93: Third European Conference on Radiation and its Effects on Components and Systems*, Fraunhofer INT, Euskirchen, Germany, pp. 560–563, 1995.

[114] J. Bogaerts, B. Dierickx, G. Meynants, and D. Uwaerts, "Total Dose and Displacement Damage Effects in a Radiation-Hardened CMOS APS," *IEEE Transactions on Electronic Devices*, vol. 50, pp. 84–90, January 2003.

[115] G. P. Summers, E. A. Burke, M. A. Xapsos, C. J. Dale, P. W. Marshall, and E. L. Petersen, "Displacement Damage in GaAs Structures," *IEEE Transactions on Nuclear Science*, vol. 35, no. 6, pp. 1221–1226, December 1988.

[116] B. H. Rose and C. E. Barnes, "Proton Damage Effects on Light Emitting Diodes," *Journal of Applied Physics*, vol. 53, pp. 1772–1780, 1982.

[117] H. Rose, M. S. Hopkins, and R.A. Fields, "Characterization and Control of Gamma and Proton Radiation Effects on the Performance of Nd: YAG and Nd: YLF Lasers," *IEEE Journal of Quantum Electronics*, vol. 31, pp. 1593–1602, 1995.

[118] Sira Electro-Optics Limited, South Hill Chislehurst, England, internet address checked February 8, 2005.
http://reat.space.qinetiq.com/Reat/reat_dev/REAT_245tn2_draft.pdf

[119] G. L. Jacobs *Evaluation of Fiber Optic Emitters and Detectors*, Document 08-012**,** Goddard Space Flight Center, Greenbelt, Maryland, May 16, 1986.

[120] C. R. Webster, S. P. Sander, R. Beer, R. D. May, R. A. Knollenberg, D. M. Hunter, and J. Ballard, "Tunable Diode Laser IR Spectrometer for in-situ Measurements of the Gas Phase Composition and Particle Size Distribution of Titan's Atmosphere," *Applied Optics*, vol. 29, pp. 907–917, 1990.

[121] R. Craig, B. Li, and B. Chan, "Laser Qualification for the SILEX Program," *Proceedings of the SPIE*, vol. 2123, pp. 238–242, 1994.

[122] A. S. El-Dinary, T. D. Cole, M. T. Boies, R. A. Reitner, and D. Rodriguez, "Testing and Space Qualification of the NEAR Laser Rangefinder," *Proceedings of the SPIE*, vol. 2748, pp. 140–150, 1996.

[123] D. R. Maack, "Reliability Methodology for Fiber Optic Components," *Optics and Photonics News*, vol. 13, no. 5, pp. 34–39, May 2002.

[124] M. W. Wright, D. Franzen, H. Hemmati, H. Becker, M. Sandor, "Qualification and Reliability Testing of Commercial High-Power Fiber-Coupled Semiconductor Lasers for Space Applications," *Optical Engineering*, vol. 44, no. 5, Paper 054204, 2005.

[125] R. H. Czichy, "Optical Design and Technologies for Space Instrumentation," *Proceedings of the SPIE*, vol. 2210, pp. 420–433, 1994.

[126] D.-R. Schmitt, "Degradation Effects of Optical Components in the Low Orbit," *Proceedings of the SPIE*, vol. 2210, pp. 449–465, 1994.

[127] M. I. Anapol, "Silicon Carbide Lightweight Telescopes for Advanced Space Application," *Proceedings of the SPIE*, vol. 2210, pp. 373–382, 1994.

Chapter 6
Earth Terminal Architectures

Keith E. Wilson, Abhijit Biswas, Andrew A. Gray, Victor A. Vilnrotter,
Chi-Wung Lau, Meera Srinivasan, and William H. Farr

6.1 Introduction

Keith E. Wilson

As Solar System exploration missions evolve from flybys to orbiters and landers, high resolution instruments will enable us to establish a virtual and real presence on other worlds. The data rates needed to support advanced high resolution instruments will increase orders of magnitude—to hundreds of megabits per second.

Higher carrier frequencies, larger antennas on spacecraft, larger antennas on the ground, increased receiver sensitivities, and higher efficiency modulation and coding schemes are all strategies that increase the signal-to-noise ratio (SNR) of the link. Yet within this, the trade space is constrained by the size, mass, and power limitations of spaceborne systems. Increasing the transmitted power requires that more power be generated on the spacecraft; a challenge for spacecraft that rely on solar energy to explore the outer planets where the solar intensity is low. Increasing the onboard antenna diameter, D, reduces the required transmitted power. However, higher carrier frequencies impose increasingly more stringent demands on the antenna pointing system to reduce pointing losses and maintain the often-required $\lambda/10D$ beam pointing accuracy, and they increase the losses due to weather. To reduce the impact of these demands on the onboard systems, the approach taken for deep space communications at radio frequency (RF) carrier frequencies has been to emphasize three areas of performance enhancements for closing the link. These

are receiver gain, receiver sensitivity, and coding and modulation efficiency. This paradigm is unchanged for deep space optical links.

Section 6.1.1 discusses the Earth-based optical receiver, the search for low-cost large aperture single telescopes, and the deployment strategies to mitigate the effects of cloud cover. This section also addresses high-power laser beam uplink transmission safety and established strategies to mitigate atmospheric scintillation effects on the uplink beam. It concludes with a series of JPL concept validation experiments performed in support of deep space optical communications.

The required receiver gain for the link can also be achieved by an array of smaller aperture telescopes with equivalent area to that of a single large telescope. An array has the advantage of tapping into established technology for building smaller telescopes, thereby expanding the number of potential suppliers. Also, using a large number of smaller array elements allows economies of scale to be realized. Stahl et al. have developed a parametric cost model for ground-based telescopes using multi-variable statistical analysis [1]. Their results, comparing primary mirror and not mount costs, show the cost advantage of an array of smaller telescopes. Whether this advantage is retained in the presence of operational conditions (including mount calibrations and pointing and alignment of the array elements) is yet to be determined. In Section 6.1.2 the array approach is discussed, and its performance characteristics are compared to those of an equivalent-area single aperture. The results show that the array is a viable alternative to the single aperture.

At the receiver, the downlink signal received by the telescope is focused onto a photosensitive surface that converts the detected signal intensity to an electric current. The efficiency of this statistical process of photon detection defines the quantum efficiency of the detector. Silicon and indium-gallium-arsenide (InGaAs) detectors are the optical communications detectors of choice because of their low noise characteristics are crucial for detecting the weak deep space downlinks. Yet because of its wider band-gap, the quantum efficiency of silicon detectors is low at near-infrared wavelengths. To compensate for this, developers have explored strategies to reduce the band gap of silicon thereby increasing its absorption at infrared wavelengths [2]. A plethora of detector architectures exists within the silicon-based technology. These range from hybrid photomultiplier tubes (PMTs) with silicon surfaces to linear and Geiger-mode avalanche photodiode detectors (APDs) and focal plane arrays and are discussed in Section 6.2. Some of these devices such as the Geiger-mode APDs and hybrid PMTs can operate in the photon-counting mode. Low signal level detection is key for the signal-starved deep space links. High-sensitivity low-noise-figure detectors not only enhance the link performance but also can provide adequate link margin in the presence of higher atmospheric attenuation; a feature that can affect the number of ground stations needed to achieve the required availability.

While telescopes, optical trains, and detectors constitute the front end of the receiver chain, the back end electronics for the deep space receiver are uniquely designed to match the higher order pulse-position modulation formats of the deep space optical link. Section 6.3 describes the advanced time synchronization and post detection filtering techniques required to demodulate the nanosecond pulse streams from the deep space link.

6.1.1 Single-Station Downlink Reception and Uplink Transmission

Keith E. Wilson

6.1.1.1 Introduction. This section addresses the transmitters and receivers for the deep space optical communications. In Section 6.1.1.2 we discuss the evolution of the JPL strategy for achieving the large 10-m receivers that have been shown to be the appropriate aperture size to support deep space optical links [3]. In Section 6.1.1.3, we describe the global required deployment of the Optical Deep Space Network (ODSN) stations to provide the needed coverage and the options on network concepts to mitigate the effects of cloud cover and provide the required availability. We also discuss the application of adaptive-optics technologies to reduce the effects of sky background on the optical link at small Sun angles. Airborne receivers located above the clouds so they can provide an assured link for low-rate engineering and science data on an emergency basis are discussed in Section 6.1.1.5 along with the results of space-based receiver studies. The most recent of these baselined the technologies of the James Webb Space Telescope (JWST) to realize significant cost savings on the development of the large space-based optical receiver.

For the uplink, the laser beam serves both as a reference beacon to signal ground station availability for links with short round-trip light times (RTLTs) and as a low-rate command uplink channel. In Section 6.1.1.6, we discuss the uplink laser requirements for both the command and beacon links. Atmospheric turbulence causes scintillation and wavefront tilt as the laser beam propagates through the atmosphere. Although aperture averaging and large aperture detectors mitigate these atmospheric effects of the direct detection downlink receiver, in the absence of clouds and aerosols, the uplink beam is most affected by clear-air atmospheric turbulence. The effect of atmospheric turbulence on the uplink and the multi-beam strategies to mitigate these effects are described in Section 6.1.1.7.

Ground-to-space laser beam propagation from United States territory is regulated by two Federal government agencies. The Federal Aviation Administration (FAA) regulates beam propagation in navigable air space. The United States Strategic Command (USSTRATCOM of the U.S. Air Force) Laser Clearinghouse (LCH) regulates propagation of laser beams into space for all Department of Defense (DoD) related programs. In Section 6.1.1.10, we describe a JPL three-tiered safety system for safe ground-to-space laser beam

propagation. In Section 6.1.1.11, we discuss a series of experiments conducted at JPL to validate deep space optical communications link concepts. In Section 6.1.1.11, we describe the 1968 laser link from Table Mountain Observatory to the Surveyor-VII camera on the Moon. This demonstration qualitatively showed the effects of atmospheric conditions and "seeing" on laser beam propagation. The 1992 Galileo Optical Experiment (GOPEX) described in Section 6.1.1.11.1 demonstrated a six–million-kilometer link, and quantified the effects of scintillation on single-uplink beam propagation. In addition, GOPEX demonstrated the benefits of ground-station site diversity.

The transmitted beam intensity is inversely proportional to the square of the beam divergence, and adaptive optics can allow the propagation of diffraction-limited beams through the atmosphere. In Section 6.1.1.11.2, we describe the 1994 joint JPL and Starfire Optical Range (SOR), Albuquerque, New Mexico demonstration of an atmosphere-compensated laser uplink to the Apollo 15 retro-reflectors on the Moon. Lessons on beam propagation learned from GOPEX were implemented in the Ground-to-Orbiter Lasercom Demonstration with the National Space Development Agency of Japan's (NASDA's) ETS-VI spacecraft where the first scintillation mitigation using multi-beam uplinks was demonstrated. This is discussed in Section 6.1.1.11.3. The Missile Defense Agency's lasercom terminal manufactured by AstroTerra Corp. of San Diego, California, was integrated onto the Space Technology Research Vehicle (STRV-2) instrument platform at JPL and launched on the Air Force TSX-5 spacecraft. Although a link was never established between the spacecraft and ground terminals, invaluable lessons were learned both about the design and materials selection for space-based terminals and about the design of transportable ground terminals.

6.1.1.2 Deep-Space Optical Ground Receivers. The 1994 Ground-based Advanced Technology Study (GBATS) report explored two deployment options and quantified the number of ground antennas needed to support a high-availability optical link in the presence of cloud cover [4]. The study evaluated two global subnets: (i) The Clustered Optical Subnet (COS) that consisted of groups of three receivers near the current DSN locations, and (ii) the Linearly Dispersed Optical Subnet (LDOS) receivers consisting of six to eight stations approximately equally separated in longitude. To reduce the overall cost of deploying six to nine Deep Space Optical Receiving Antenna (DSORA) stations globally, several designs were considered. Among them were instruments that are non-diffraction-limited at optical wavelengths (photon-buckets) [5]. Millimeter and submillimeter telescope receivers (such as the California Institute of Technology Sub-millimeter Observatory's (CSO) 10.4-m Leighton's dish, shown in Fig. 6-1, and the 3-m Kolner Observatory for Sub-Millimeter Astronomy (KOSMA) telescope at Zermatt, Switzerland, Fig. 6-2)

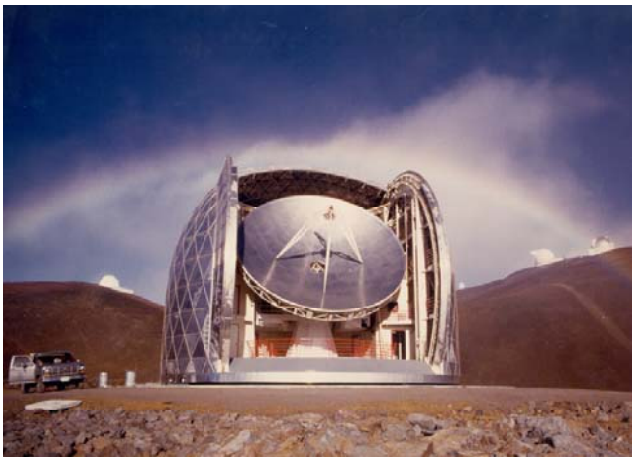


Fig. 6-1. Leighton Dish California Institute of Technology (Caltech) Sub-Millimeter Observatory (CSO), Mauna Kea, Hawaii.

Fig. 6-2. KOSMA 3-m sub-millimeter telescope in Zermatt, Switzerland, built by Vertex Antennentechnik.

were explored [6,7]. The design approach of the low cost 10-m Hobby-Ebberly telescope (Mount Fowlkes, Fort Davis, Texas) was also considered [8].

The CSO dish consists of 84 nominally 1-m hexagonal panels and is modular in design and easily assembled. The f/0.45 parabolic primary panels are made of aluminum front and back face sheets bonded to an aluminum honeycomb core for stiffness. As manufactured at the California Institute of Technology, the antenna had a surface figure of approximately 15 μm root mean square (RMS) Although low-cost and capable of supporting high

nighttime links in the absence of sky background, the surface roughness of the CSO could not support daytime operation at small Sun angles.

Strategies explored included an integral sunshade that consisted of a series of 1.11-m side hexagonal tubes approximately 12 m long integrated to the hexagonally segmented primary to prevent direct sunlight from being incident on the primary [9,10,11]. Analysis showed that the sunshade would enable pointing to as low as ~12 deg of the Sun without signal loss. A series of vanes arranged like an asterisk designed into each tube would allow pointing to within 6 deg of the Sun with only 3.6 percent signal loss.

The KOSMA telescope uses actuators to achieve the required panel-to-panel alignment, and it uses highly polished surfaces to achieve the required surface quality. The primary is made of 1–2 mm thick light-weighted aluminum panels that are 0.9 m × 0.8 m. The panels were micro-machined to better than 4 μm RMS surface roughness by the Laboratory for Micro-machining at the University of Bremen, Germany, to serve as a starting point. The modifications considered for the DSORA were first to nickel plate the panels to fill the grooves from the micro machining, and then inscribe the plated surface with micron-sized stress relief patterns to relieve any thermally induced stresses caused by the differential thermal expansion between the nickel and aluminum. An aluminum layer with an SiO_2 overcoat would then be deposited over the nickel to increase the surface reflectivity in the visible from the 50 percent of nickel to approximately 90 percent. The residual panel surface roughness expected by using this process is approximately 200 nm peak-to-valley [12].

A JPL study by Sandusky et al. analyzed the effects of sunlight scattered from the primary mirror into the focal plane in greater detail [13]. Their results showed that a telescope meeting the expected performance of the modified KOSMA antenna could support optical communications at much smaller Sun angles than the DSORA [14]. The Sandusky results were that machined panels with RMS surface roughness $\sigma \sim \lambda/10$ could support optical communications at a 1-deg Sun angle.

Other cost reduction approaches considered to achieve the desired surface quality have included glass fusion technology for making lightweight panels. This is shown in Fig. 6-3 [14]. Alternatively, a spherical rather than a parabolic primary, as was done in the Hobby-Ebberly, can take advantage of the economies of scale. More recently, an array of smaller telescopes of about 1-m diameter each has been considered for the Mars laser communications demonstration project.

6.1.1.3 Mitigating Cloud Cover and Sky Background Effects at the Receiver. Selection of optical deep space network (ODSN) sites to provide both nominal twenty-four hour coverage and weather availability requires the consideration of several factors. Among these are:

1) Longitude of locations to achieve near twenty-four hour coverage

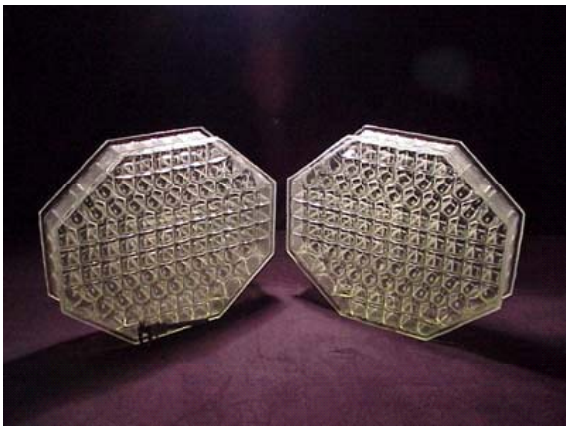


Fig. 6-3. Candidate lightweight glass panels from HEXTEK for making the primary for a 10-m receiver.

2) Altitude and weather statistics of the site to mitigate effects of atmospheric turbulence and sky background.
3) Accessibility and infrastructure support including transportation, housing and telecommunications facilities.
4) Geopolitical: To protect NASA's investment and ensure long-term accessibility and security.
5) Environmental impact of construction at the site. Ignoring these impacts can result in cost overruns and significant schedule delays.
6) Security and access control: To protect the assets and ensure system integrity.
7) Aircraft avoidance: To prevent illumination of aircraft by high power uplink lasers.

Cloud cover data for sites in the United States Southwest show seasonal cloud cover ranging from 12 percent in the summer months to greater than 50 percent in the winter and fall months [15]. Table 6-1 gives the seasonal cloud cover variation during the night at Table Mountain Facility (TMF) observatory over the eleven-year period 1993 to 2003. The data show that the mean cloud cover ranges from 21 percent in the summer to 51 percent in the winter. For an array of n receivers located in independent weather cells in a region of cloud cover probability q, the weather availability p for is given by $p = (1 - q^n)$, when n stations are simultaneously visible to the spacecraft. For a three-station COS configuration located in a region of 20 percent to 50 percent cloud cover probability, the corresponding weather availability decreases from 99 percent to 88 percent. Currently, weather satellite data are being used to develop cloud-free line-of-sight statistics for a range of global sites [16].

The 1994 GBATS study considered a six-, seven-, and eight-station LDOS and a 3 × 3 and 3 × 4 COS located between +40 and –40 degrees latitude [4]. The locations of the 3 × 3 COS and the six-station LDOS shown in Fig. 6-4 were in the vicinity of existing DSN sites at Goldstone (California), Canberra (Australia), and Madrid (Spain) separated by 200 km to 500 km, the nominal size of a weather cell. For this configuration, the size of the weather cell

Table 6-1. Table Mountain, California, cloud statistics for 1993–2003.

Year	**Annual Avg. %Cloudy**	Winter %Cloudy	Spring %Cloudy	Summer %Cloudy	Fall %Cloudy
1993	**27.5**	50.0	36.7	15.4	22.2
1994	**37.6**	56.5	40.5	32.5	17.7
1995	**26.6**	33.3	35.6	16.9	29.3
1996	**31.6**	80.0	39.5	11.1	13.8
1997	**22.1**	16.0	16.0	15.4	39.4
1998	**21.2**	52.6	22.2	8.5	20.0
1999	**27.3**	54.5	26.1	14.3	21.0
2000	**28.3**	57.1	27.6	26.3	27.1
2001	**39.6**	50.0	46.1	29.4	53.9
2002	**39.2**	57.1	12.0	23.5	55.7
2003	**39.1**	56.6	39.7	33.3	33.3
Seasonal average		**51.2**	**31.1**	**20.6**	**30.3**

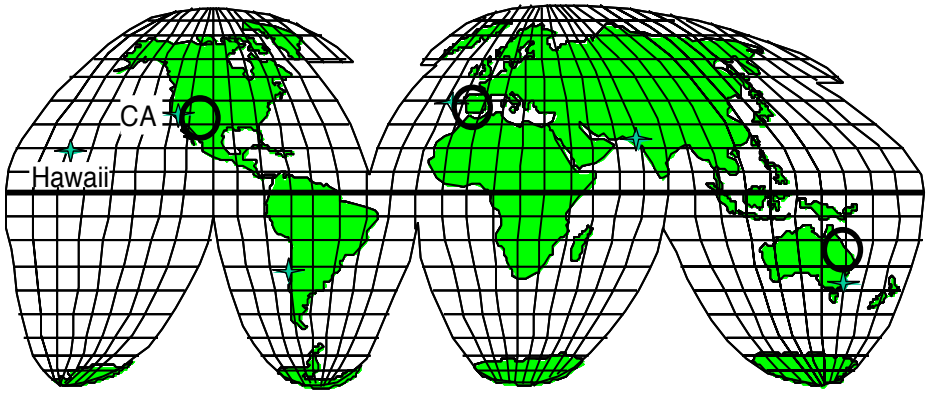


Fig. 6-4. One realization of a COS 3 × 3 and a six-station LDOS deployment for global

establishes the minimum station separation. The maximum separation will be based on the mission operations concepts. In Fig. 6-4, the locations of the six-station LDOS were Mauna Kea (Hawaii), U.S. Southwest, Cerro Pachan (Chile), Calar Alto (Spain), Zairat (Pakistan), and Siding Spring (Australia); all clearly in independent weather cells.

A Mars-to-Earth link offers an excellent example of the impact of the mission's concept of operations on the ground station deployment for weather diversity. The large difference in range between conjunction and opposition (0.37 AU and 2.7 AU) in the Mars-to-Earth link allows a ready comparison of COS and LDOS operational scenarios.

At the two extreme ranges of the link, the footprints of the optical downlink beam of 3-μrad 3-dB beam width (a 30-cm transmitter operating at a 1-μm wavelength) are 166 km at opposition and 1200 km at solar conjunction. A maximum COS station separation of less than 1200 km would allow the downlink beam to simultaneously cover all stations in the region at conjunction. This approach is particularly advantageous for communications from ranges where the light time is so long (for this case, the one-way light time at conjunction is approximately 21 minutes) that threatening clouds at any one station could obscure the line of sight within the time it takes to send a command to the spacecraft. At opposition the beam footprint is small. However, the light time is short, approximately 3 minutes, and could support an operations scenario of re-pointing the beam from one station to the next to communicate through a cloud-free line of sight in patchy cloud sky, thereby increasing each station's availability.

These options do not apply to the LDOS configuration. In all cases station separation exceeds 1200-km, and for all ranges the spacecraft terminal would need to be re-pointed from one station to the next to avoid impending cloud cover. The selection of sites for a global ground network will be based on an optimization of several considerations including those listed above.

6.1.1.4 Daytime Sky Background Effects. Rayleigh scattering of sunlight by the atmosphere is the main source of optical background noise on the daytime optical link. The ability to point at small Sun angles increases the availability of the optical link around solar opposition. Mitigating the effects of sky background on the optical link requires the implementation of spatial and spectral filtering techniques coupled with higher order pulse position modulation (PPM). The optical noise is directly proportional to the slot width and can therefore be reduced by implementing higher order PPM schemes. Yet, the receiver complexity and the susceptibility of the link performance to pulse jitter increases with PPM order. Angstrom-wide Fabry-Perot optical filters further suppress sky background noise. However, when pointing as close as 3 deg to the Sun additional techniques are required. JPL has been exploring the application of adaptive optics methods to reduce the receiver field-of-view

(FOV) and thereby further mitigate the effects of sky background on the optical link. Preliminary analysis has shown that an adaptive optics system with 900 actuators per square meter adaptive optics techniques can realize as much as 6 dB increase in SNR for PPM-16 formats in the presence of high sky background noise [17]. Figure 6-5 illustrates improvement in SNR for different background levels realized by implementing adaptive optics across a 1-m aperture.

6.1.1.5 Earth-Orbiting and Airborne Receivers. Cloud cover, atmospheric turbulence, sky background, and high winds affect the optical availability and quality of the link to a ground receiver. Oceans and geopolitics are constraints on the availability of locations for deploying a ground stations. A space-based receiver offers the optimum availability for a single receiver, and an optical interplanetary crosslink between a geostationary receiver and the probe offers a high bandwidth link with high availability. Past studies of space-based receivers [18,19] have shown that the payload and launch costs drive the cost of the space–based receiver to approximately twice that of the 3 × 3 COS ground-based receiver network. The high payload cost was due to the pre-development technology needed to build and deploy large apertures in space. A 1998 study took advantage of the James Webb Space Telescope deployable apertures technology to achieve significant reductions in the payload costs. The result showed that the cost of the 7-m space-based receiver station was comparable to and within cost uncertainties of a 3 × 3 COS (10-m apertures) [20].

Concepts for airborne receivers are based on 2.7-m apertures such as the Stratospheric Observatory for Infrared Astronomy (SOFIA) telescope [21]. With advances in station keeping, balloon-borne receivers have been recently added to the list of approaches to mitigate atmospheric effects. Key advantages of this approach are: (i) The receivers are stationed above the cloud cover and the scintillation effects of the troposphere, (ii) they are above most of the atmosphere, and background sky scatter when pointing close to the Sun is minimal, (iii) deployment costs are low, (iv) they are serviceable, (v) they can be deployed globally over oceans and avoid geopolitical issues. Yet, because of their smaller apertures airborne receivers have been considered only to expand availability for the transmission of low-rate science and critical engineering data. This, however, is rapidly changing with the development of efficient photon-counting IR detectors at the infrared communications wavelength.

6.1.1.6 Uplink Beacon and Command. For bi-directional optical communications to a deep space probe out to Mars ranges, the uplink will not only serve as a carrier for transmitting commands to the spacecraft but also as a beacon to signal to the spacecraft identifying the ground station location with a cloud-free line of sight. Although hybrid concepts that use an RF uplink and an

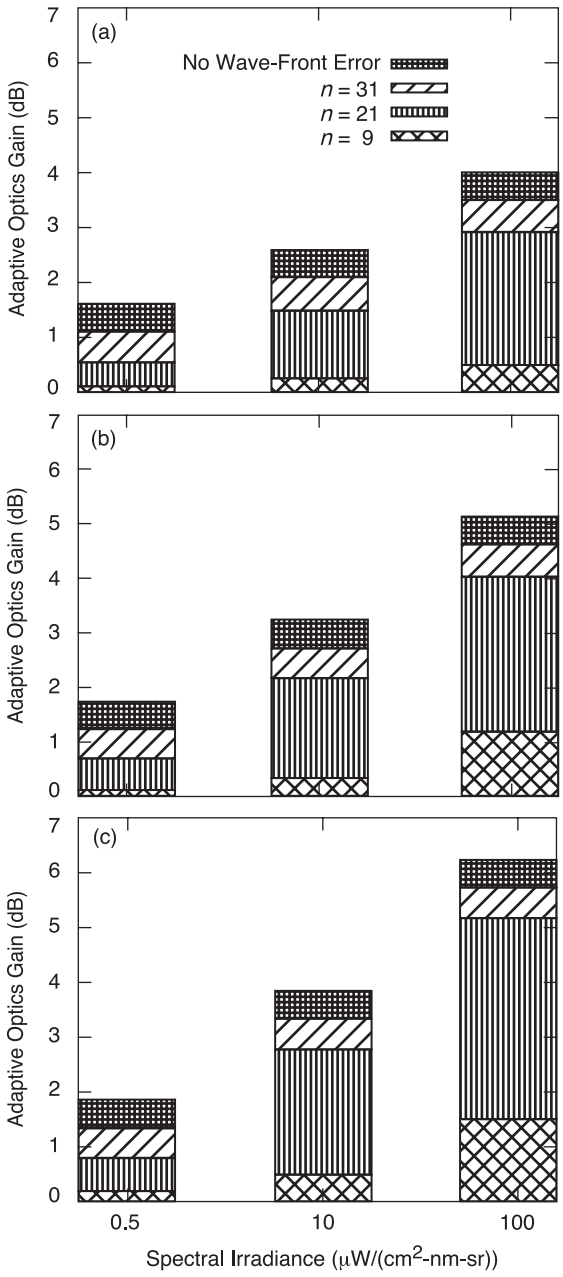


Fig. 6-5. Improvement in SNR for different background levels realized by implementing adaptive optics across a 1-m aperture: (a) 256-PPM, r_0 = 7 cm, probability of bit error (PBE) = 0.01, (b) 64-PPM, r_0 = 7 cm, PBE = 0.01, and (c) 16-PPM, r_0= 7 cm, PBE = 0.01.

optical downlink are an alternative approach, with such an approach, there is still a need for precision pointing of the narrow optical beam from the spacecraft to the ground. Proposed approaches for pointing the downlink from Mars, use an uplink beacon.

The required beacon power is determined from the requirements of the onboard acquisition and tracking sensors. For an onboard sensor requiring N signal photoelectrons per charged-couple device (CCD) frame, the uplink power is given by Eq. (6.1-1):

$$P_T = \frac{\frac{hc}{\lambda}}{G_T\left(\frac{\lambda}{4\pi R}\right)^2 G_R \eta_{QE}\eta_{atmos}\eta_{AT}\eta_{trans}} NF \qquad (6.1\text{-}1)$$

where h is Planck's constant; λ is the optical wavelength; F is the frame rate; and η_{QE}, η_{atmos}, η_{ATP}, and η_{trans} are the quantum efficiency, the atmospheric transmission, the optical transmission of the acquisition and tracking optical train, and the uplink transmitter efficiency, respectively.

An estimate of the required uplink power has been calculated for two scenarios, for a 30-cm receiver aperture at the spacecraft with the requirement that the downlink beam be stabilized to 1-μrad and 0.3-μrad, respectively [22]. Assuming a noise spectral density of the Olympus spacecraft, the required number of signal photoelectrons per CCD frame is 7,500 and 33,000 for the 1-μrad and 0.3-μrad stabilization requirements, respectively. The required uplink power over the link range 0.38 AU to 2.7 AU is given in Fig. 6-6 for these two cases, where $\lambda = 1064$ nm, $F = 10$ Hz, $\eta_{QE} = 0.4$, $\eta_{atmos} = 0.7$, $\eta_{ATP} = 0.25$, $\eta_{trans} = 0.5$, $G_T = 2.33 \times 10^{10}$ (5-cm uplink beam aperture), and $G_R = 8.37 \times 10^{11}$ (30-cm effective receiver aperture).

Strategies that use inertial sensors to measure spacecraft vibrations between camera frames allow for precision pointing at slower camera frame rates. This is an approach being considered at the farthest ends of the Mars link [23]. For ranges beyond Mars, plans call for a beacon-less approach that includes the use of an advanced system of well-calibrated and characterized onboard star trackers for precise pointing.

The optimum uplink transmitter gain is dictated by the atmospheric seeing conditions of the site. A plot of daytime atmospheric seeing measurements, θ_s, made at the Optical Communications Telescope Laboratory (OCTL) is shown in Fig. 6-7. Atmospheric seeing is a measure of the atmospheric turbulence and changes with time of day, atmospheric conditions, and season. The data in the figure were based on measurements using the star Polaris at a 94-deg Sun-angle. The flux incident on a CCD science camera at the OCTL focal plane was

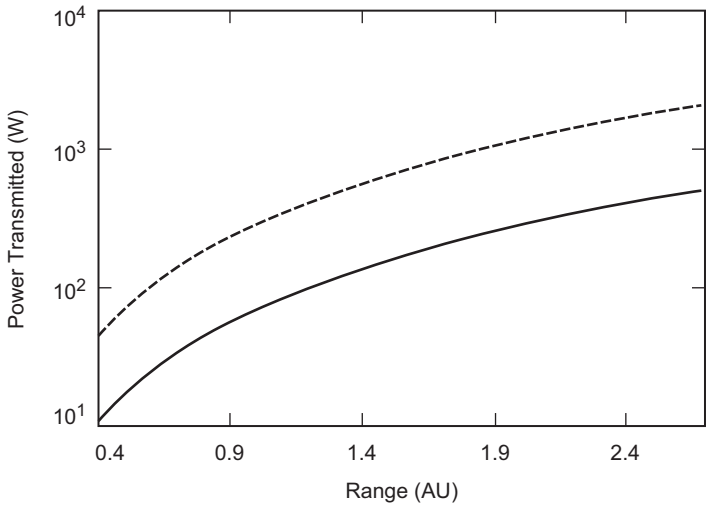


Fig. 6-6. Plot of required beacon power as a function of range for 7500 (solid) and 33,000 (dashed) signal photons per CCD frame.

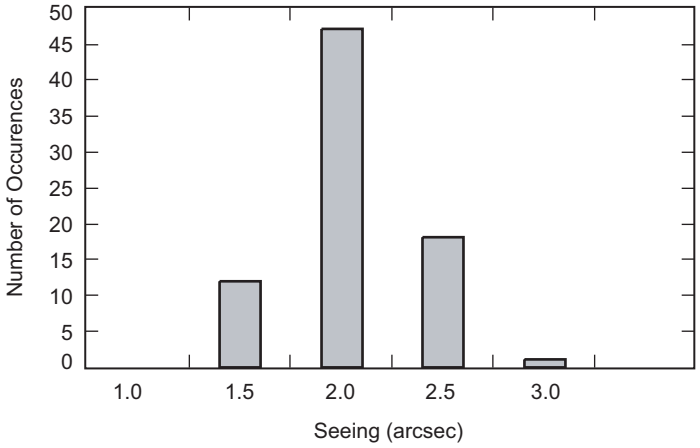


Fig. 6-7. Daytime atmospheric seeing measurements made through an 810-nm filter. Measurements were made in March 2004 using the 1-m OCTL telescope.

measured over a period of several seconds, and the full-width-at-half-maximum (FWHM) of the image calculated. Images were taken through a 10-nm wide optical filter centered at 810 nm. The mean seeing was 1.8 ±0.16 arc seconds.

Figure 6-8 shows the coherence length at 1064 nm. The results presented were corrected for the 34-deg zenith angle of Polaris as seen from the OCTL.

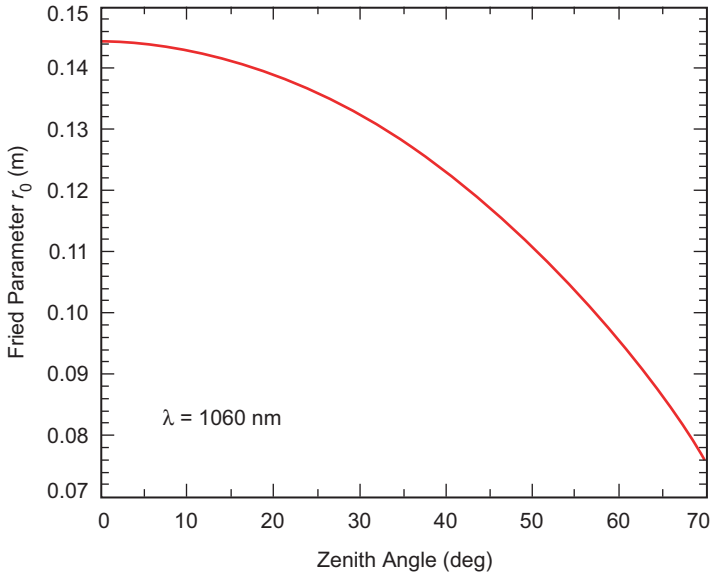


Fig. 6-8. Fried coherence length for range of zenith angles. Calculations are based on data taken at the JPL OCTL, Wrightwood, California.

Communications with deep space probes is highly asymmetric. Data rates for uplink commands are typically a few kilobits per second while downlinks from Mars will be in the tens to hundreds of megabits per second. Beam propagation strategies for the uplink optical command are the same as those for beacon transmission. At the low uplink data rates, the average optical power required for communications is much less than that required for the uplink beacon. However, because the higher pulse position modulation formats afford greater suppression of in-band optical noise, communications design considerations tend towards higher PPM orders. Yet, the peak power requirements will depend on the modulation format with the higher PPM orders requiring peak powers well in excess of those required for the beacon laser. The selection of a modulation format will be based on a systems trade of the modulation format required for background against the availability of lasers with the requisite average and peak power. For low-rate uplinks, on-off keying (OOK) or binary PPM formats are more consistent with the beacon laser requirements. In these modulation formats the peak-to-average power ratio is a factor of two.

The number of photons per bit required to close the link depend on receiver efficiencies, coding gains, and other factors. Modulating the beacon uplink in an OOK or binary PPM format is an operational scenario that would support the required beacon signal and a low rate communications uplink. Figure 6-9 plots the bits per photon received as a function of range for 20 W of optical

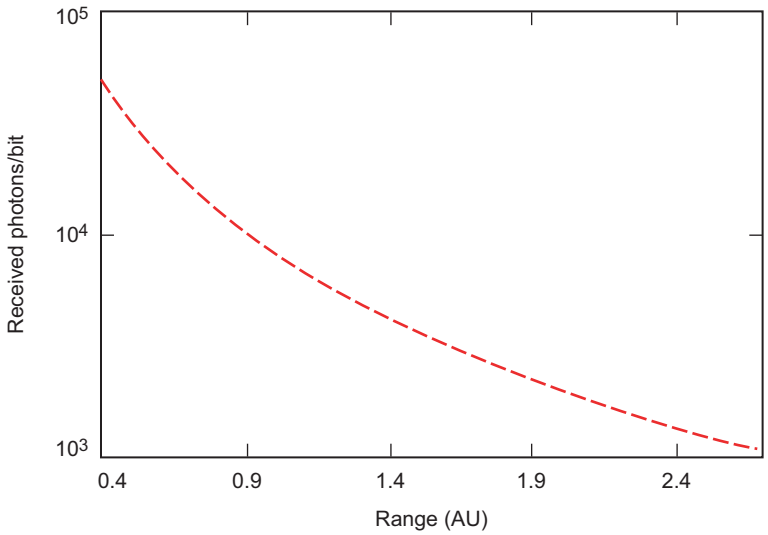


Fig. 6-9. Photons per bit received versus range for a 5-cm transmitter aperture emitting 20 W of power at 1064 nm.

power at 1064 nm. The beam is transmitted through a 5-cm aperture, corresponding to 20-μrad full-width-at-half-maximum beam divergence, and received by a 0.3-m aperture. The optical losses in the receiver and transmitter are 3 dB, and atmospheric transmission is 80 percent. Pointing losses correspond to 30 percent off-peak combined pointing bias and jitter. The data show that the received photons per bit ranges from 5.1×10^4 at 0.4 AU to 1.1×10^3 at 2.7 AU. The details of the link analysis are given in Table 6-2 for the 1-AU range.

Although a full-beacon power modulation is modeled above, it should be noted that other alternative approaches to integrating the command uplink and the beacon (such as wavelength or polarization diversity) are equally viable. In these approaches the command beam is shifted in wavelength or is orthogonal in polarization to the beacon and is then separated at the receiver by dichroic or polarizing beam splitters.

Over the years high power laser sources have been developed by industry for a variety of DoD applications. Boeing, Northrop Grumman, IPG Photonics, and other manufacturers build high power lasers in the 1000-nm to 1500-nm region that are useful for communications and beacon uplinks [24,25,26]. Unconstrained by size, mass, and power requirements of spaceborne systems, high power ground-based lasers satisfying kilobit per second uplink command links to probes in the Solar System are within the capabilities of the current state-of-the-art.

Table 6-2. Link analysis table for 10-bps uplink to spacecraft at 1 AU (5 percent of the transmitted power, i.e., 20 W, is diverted for the uplink command).

		dB
Gains		
Transmitted power, W	400	26.02
Effective transmitter aperture, m	5.00^{-2}	
Transmitter gain	2.18^{10}	103.38
Receiver aperture, m	3.00^{-1}	
Receiver gain	7.85^{11}	118.95
Losses		
Transmitter optical system	0.5	–3.01
Pointing loss		–1.55
Atmospheric transmission	8.00^{-1}	–0.97
Range, m	1.50^{11}	
Free space loss	3.19^{-37}	–364.97
Receiver optical system	0.5	–3.01
Loss to acquisition/tracking	0.05	–13.01
Signal Reception		
Received power, W	1.53^{-14}	–138.17
Energy/photon, J	1.87^{-19}	–187.29
Received number of photons/s	8.17^{4}	49.12
Bit rate	10	
Detection		
Detected photons/bit	8.17^{3}	

6.1.1.7 Techniques for Mitigating Atmospheric Effects. Scintillation and wavefront tip/tilt induces signal fades in an optical beam propagating through clear air turbulence. In this section we explore the impact of atmospheric turbulence and discuss mitigation strategies. On the downlink, the source of the disturbance is close to the receiver aperture, and using receiver apertures that are many atmospheric coherence cell sizes averages the scintillation and mitigates the effects of atmospheric turbulence. Discussions of aperture averaging effect are given by Fried and by Churnside [27,28]. Churnside's approximation for the aperture averaging is given in Eq. (6.1-2), where the aperture averaging factor A, is defined as the ratio of the intensity variance for an aperture D to that for a point detector $(\sigma_I(D)^2/\sigma_I(0)^2)$ for weak turbulence conditions. Churnside has shown that under weak turbulence conditions the aperture-averaging factor for large apertures can be approximated by

Eq. (6.1-2) for a plane wave A_{pw} and by Eq. (6.1-3) for a spherical wave A_{sw}. The wavenumber is k, and L is the optical path through the atmosphere.

$$A_{pw} = 0.932\left(\frac{kD^2}{4L}\right)^{-7/6} \tag{6.1-2}$$

$$A_{sw} = \left(1 + 0.214\left\{\frac{kD^2}{4L}\right\}^{7/6}\right)^{-1} \tag{6.1-3}$$

Figure 6-10 shows the aperture averaging factor for both cases. The downlink from deep space is in the far field of the transmitter and corresponds to the plane wave case. The results show that for a 10-m class receiving aperture the intensity variance is reduced by approximately five orders of magnitude.

On the uplink the wavefront disturbance is close to the source, and the effects on the far-field intensity distribution are more severe. Aperture averaging is impractical, so alternative mitigation strategies are required. Adaptive-optics wavefront correction and multi-beam propagation are uplink strategies used to mitigate atmospheric turbulence effects.

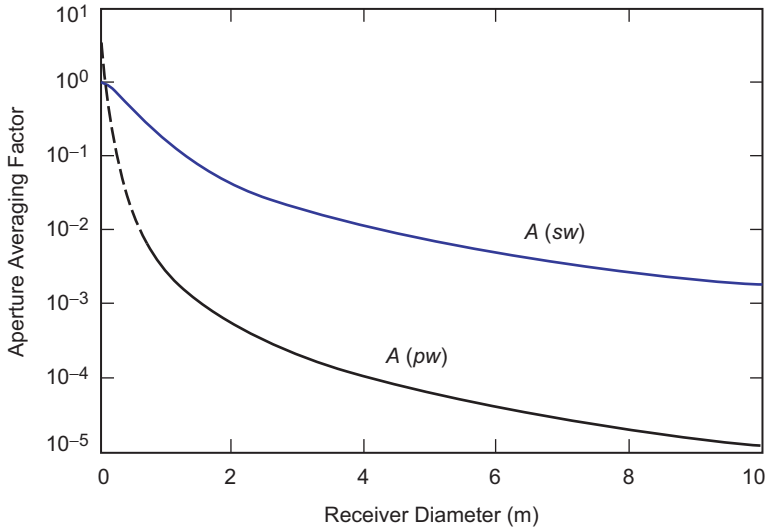


Fig. 6-10. Aperture averaging factor (the ratio of the intensity variance for an aperture *D* to that for a point detector versus receiver diameter). The upper trace is for a spherical wave and the lower for a plane wave, which is more indicative of optical links from deep space.

6.1.1.8 Adaptive Optics. The literature is replete with references on adaptive optics wavefront correction [29,30,31]. Rayleigh backscatter from copper vapor lasers focused 15 km to 18 km into the atmosphere have been used to correct higher order wavefront aberrations [32]. Sum frequency generation, using the 1064-nm and 1320-nm outputs from neodymium: yttrium aluminum garnet (Nd:YAG) lasers, and dye laser technologies have been used to generate a 589-nm probe beam for exciting the 90-km high mesospheric sodium layer [33,34]. Yet, laser guide stars allow for correction of higher order aberrations. Correcting lower order tip/tilt aberration requires the use of an exo-atmospheric beacon such as a natural guide star.

Figure 6-5 shows the gain that could be realized by adaptive optics in the deep-space channel. Yet, realizing this gain requires correction of both lower-order and higher-order aberrations. For deep-space communications where the spacecraft moves across the star field during interplanetary cruise and is in orbit around a planet during its mission, there is a paucity of stars bright enough for the lower-order tip/tilt correction. The planets are extended sources, and tip/tilt correction would require the use of a correlation tracker. Yet, for uplink beam propagation that uses the light from a planet or the downlink from the spacecraft as a reference beacon, the large point-ahead angles, given by Eq. (6.1-4), can exceed the isoplanatic angle θ_0, given by Eq. (6.1-5), and the uplink beam with such a reference can experience tilt errors greater than the beam divergence.

$$\theta_{pa} = 2\frac{V_{s/Earth}}{c} \tag{6.1-4}$$

$V_{s/Earth}$ is the relative velocity between the spacecraft and Earth

$$\theta_0 = \left[2.914\left(\frac{2\pi}{\lambda}\right)^2 \left(\sec(\xi)\right)^{8/3} \int_h dh C_n^2(h) h^{5/3}\right]^{-3/5} \tag{6.1-5}$$

for an estimated 5-km mean turbulence height the isoplanatic angle ranges from 7 μrad at zenith to 0.5 μrad at an 80-deg zenith angle.

Equation (6.1-6) gives the RMS atmospheric turbulence-induced wavefront tip/tilt across an aperture *D*. From the wavelength dependence of r_0 given in Chapter 3 and the r_0 dependence of the RMS tilt error in Eq. (6.1-6), the wavefront tilt is seen to be independent of wavelength. Figure 6-11 shows a plot of the RMS wavefront tilt as a function of zenith angle for three apertures of 5 cm, 10 cm, and 1 m. These apertures correspond to 3-dB beam widths of 22 μrad (4.4 arc-sec), 11 μrad (2.2 arc-sec) and 1.1 μrad (0.22 arc-sec), respectively. The RMS tilt error increases with zenith angle and is proportional to $D^{-1/6}$ [35]. Thus, the larger the aperture, the smaller the atmospheric tip/tilt.

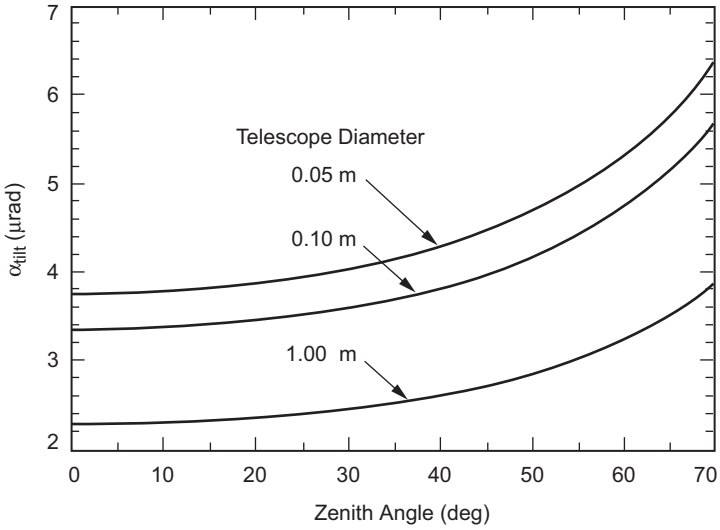


Fig. 6-11. RMS wavefront tilt as a function of zenith angle for three telescope apertures.

$$\sigma_{tilt} = 0.43\left(\frac{\lambda}{D}\right)\left(\frac{D}{r_0}\right)^{5/6} \tag{6.1-6}$$

Although atmospheric tip/tilt affects the uplink beam pointing, the key metric of the link is how the atmosphere-induced beam jitter (dynamic tip/tilt) impacts the beam intensity on target. Specifically, what fraction of the far field beam size is due to the tip/tilt. The results in Fig. 6-11 show that the tip/tilt increases with decreasing transmitter aperture. The figure also shows that the ratio of the tip/tilt to beam size for these cases ranges from 17 percent to 30 percent of the beam size for the 5-cm beam, to 30 percent to 53 percent for the 10-cm beam; increasing with zenith angle. For the 1-m beam, the tip/tilt ranges from 2 to 3 times the beam size.

Currently, tip/tilt correction is achieved by using an exo-atmospheric beacon, such as a natural guide star or a nearby planet. For optical communications, the downlink from the spacecraft can serve as an exo-atmospheric beacon. The degree of correction achievable when using the downlink as a guide star for the uplink beam depends on the angle between the received and uplinked beams, i.e., the point-ahead angle. Figure 6-12 is a plot of the RMS wavefront tilt normalized to the 3-dB diffraction limited beam width of a 1-m telescope transmitting a 1064-nm beam. The results show that for point-ahead angles of less than 10 μrad, the RMS tilt error is less than 10 percent of the beam width for zenith angles less than 48 deg. For

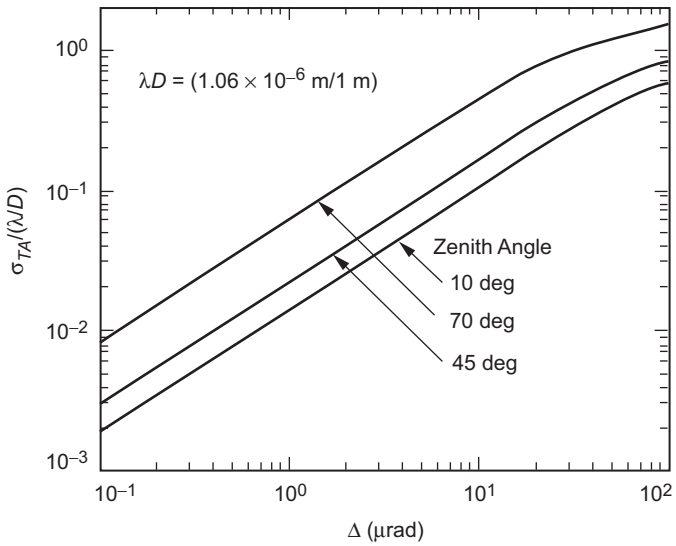


Fig. 6-12. Normalized RMS wavefront tilt as a function of point-ahead angle for a 1-m telescope.

communications to geosychronous Earth orbit (GEO) and lunar spacecraft, where point-ahead angles are less than 10 μrad, the downlink can serve as an exo-atmospheric beacon to stabilize the uplink beam pointing from a 1-m aperture.

For a deep-space uplink, the point-ahead angles are on the order of hundreds of microradians. Figure 6-12 plots the normalized wavefront tilt as a function of point-ahead angle. If a planet such as Mars or the downlink from a spacecraft in deep space were used as an exo-atmospheric beacon, the residual normalized wave front tilt error would be on the order of the beam width and will degrade with increasing zenith angle.

Although the above results show that conventional adaptive optics techniques will not support uplinks to deep space probes, there are techniques under development that could mitigate the challenge. One such technique is the polychromatic laser guide star [36]. This proposed polychromatic guide star approach begins with two-photon absorption of the optical pump by the mesospheric sodium layer, and the subsequent relaxation to the ground state by a variety of fluorescence paths ranging from the UV to the near-IR. The approach uses the observed dispersion of the UV visible through IR components of the image to determine and compensate for atmospheric tip/tilt.

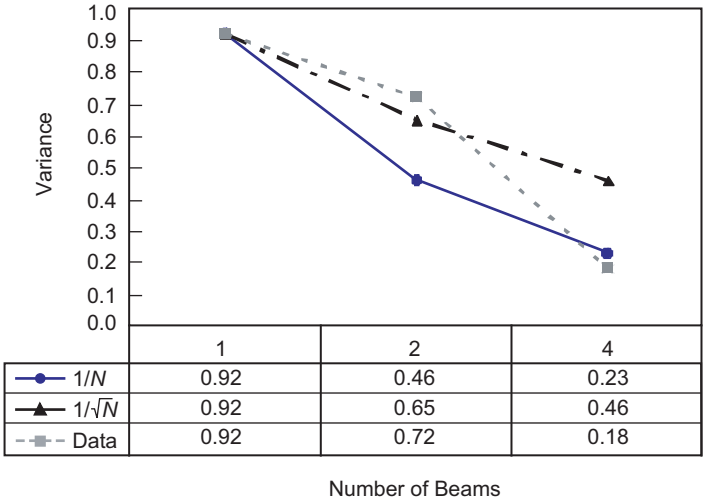
6.1.1.9 Multiple-Beam Propagation. An alternative method to mitigate scintillation on the uplink is to transmit multiple beams through different atmospheric spatially coherent cells. The uplink beam can be either from

different sources or from a single beam separated into multiple beams. When originating from a single source, it is desirable that the beams not be spectrally coherent to avoid interference effects when they are recombined in the far field. This can be done by introducing a path length delay among the beams that is greater than the coherence length of the source.

When the uplink beamwidth is much larger than the atmospheric seeing, the atmosphere has a negligible contribution to the beam divergence. Yet, atmospheric seeing depends on site location, atmospheric conditions, and zenith angle. For a site such as TMF with 2-arcsecond seeing, turbulence-induced beam spreading would increase the effective beamwidth of a 20-μrad transmitter beam by approximately 10 percent. To maintain the same beam intensity at the spacecraft, this would require a 20-percent increase in transmitted power.

Depending on the atmospheric conditions and the spacecraft tracking scenario, it may be necessary to transmit 4, 8, or 16 beams to reduce the intensity variance to the required level. Experimenters have reported a variety of observations when propagating any number N of beams. Reported intensity variance mitigation ranged from no effect to mitigation inversely proportional to $1/\sqrt{N}$ and to $1/N$ [37,38,39]. Results from a limited data set in the JPL Ground-to-Orbiter Lasercomm Demonstration is shown in Fig. 6-13. Although inconclusive, the data shows that the normalized variance in the intensity is approximately proportional to $1/\sqrt{N}$ for two beams, and to $1/N$ for four beams.

Based on these experimental results, we have plotted in Fig. 6-14 the normalized intensity variance given by Eq. (6.1-7) as a function of zenith angle



	1	2	4
1/*N*	0.92	0.46	0.23
$1/\sqrt{N}$	0.92	0.65	0.46
Data	0.92	0.72	0.18

Fig. 6-13. Comparison of normalized intensity variance data for 1-, 2-, and 4-beam transmissions compared to $1/N$ and $1/\sqrt{N}$.

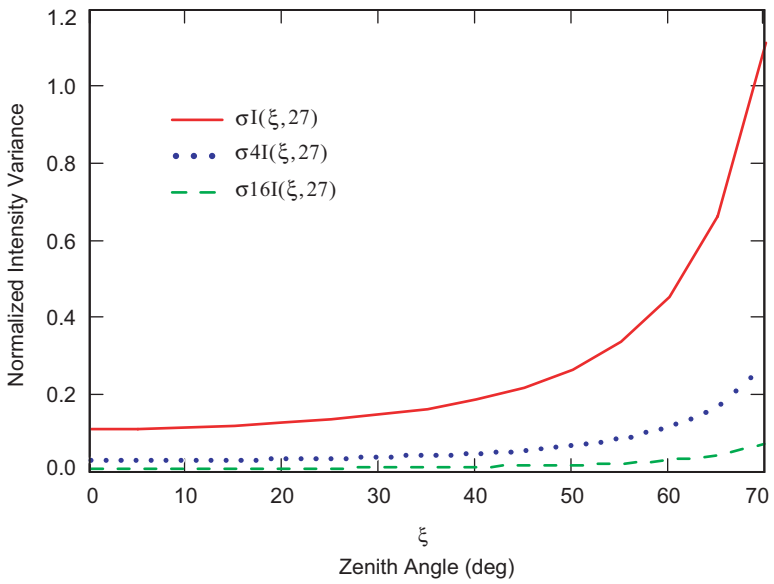


Fig. 6-14. Normalized intensity variance versus zenith angle for 1-, 4-, and 16-beam uplinks (wind speed is 27 m/s).

for 1, 4 and 16 beams [40]. These results were used in Eq. (6.1-8) to calculate the probability $F(\sigma^2,\zeta)$ that the signal fades depths are less than ζ for normalized intensity variance σ^2. Figure 6-14 plots the results for the normalized intensity variance corresponding to 1-, 4-, and 16-beam uplinks and fade depths less than 3 dB [41].

$$\sigma^2_{I/I_0} = \left(\exp\left[2.24k^{7/6}\sec(\xi)^{11/6}\int_{h_1}^{h_2} C_n^2(h,V)h^{5/6}dh\right]-1\right) \quad (6.1\text{-}7)$$

$$F(\sigma^2,\zeta) = 0.5\left[1+erf\left(\frac{-0.23\zeta+0.5\sigma^2}{\sqrt{2\sigma^2}}\right)\right] \quad (6.1\text{-}8)$$

In the final analysis, the uplink beacon transmitter design will be based on an optimization of the trade space of available uplink power, the site's atmospheric seeing conditions and transmitter gain, and the spacecraft's acquisition pointing and tracking strategy.

6.1.1.10 Safe Laser Beam Propagation into Space. Propagation of high-powered laser beams from the ground to space can put piloted aircraft and sensitive space assets at risk. JPL has developed a three-tiered system for safe ground-to-space laser beam propagation. The concept is shown in Fig. 6-15

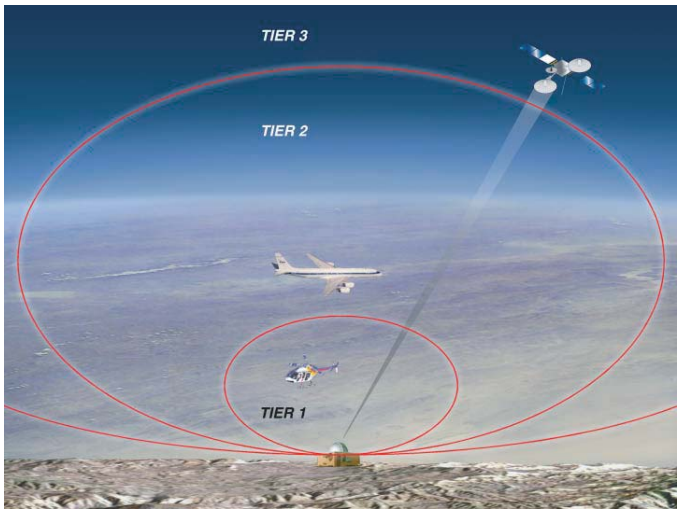


Fig. 6-15. Three-tiered system for safe laser beam propagation into space.

[42]. The objective is to demonstrate an electronic detection system that will eliminate the need for ground observers and lead the way to the implementation of autonomously operating ground stations.

Tier-1 extends from the telescope aperture to a range of approximately 4 km, and it consists of a pair of high-sensitivity long wave infrared (LWIR) cameras—the first wide field with 35 deg × 46 deg FOV, and the second narrow field with approximately 9 deg × 12 deg—for detecting small, low, and fast-moving aircraft approaching the beam. The system shown in Fig. 6-16 was developed for JPL by Image Labs International (ILI) Corporation of Bozeman, Montana.

Tier-2 overlaps with the outer range of Tier-1 and extends to the edge of space. Aircraft in Tier-2 are detected by radar. Two different radar schemes are implemented in this tier. The first, a radar display interface (RDI) is applicable to areas with controlled airspace and uses a radar feed from the controlling agency—the FAA for the United States and its territories and the corresponding air traffic control agencies for foreign stations—to alert laser beam operators to aircraft that are at risk. Such a system is used at the Air Force Research Laboratory's (AFRL's) Starfire Optical Range and the Air Force Maui Space Surveillance System (MSSS). Aircraft detected by the Air Traffic Control (ATC) radar systems in Honolulu CERAP (Combined Center Approach Control) that are in a prescribed area around MSSS are displayed at the MSSS on a direct feed from the ATC. A picture of such a display centered on Maui is shown in Fig. 6-17. In the figure the status of the RF link to the FAA is shown here on the display as "up." The exclusion zone and the laser beam pointing direction are also shown.

Fig. 6-16. Image Labs Tier-1 aircraft avoidance system consists of two long wave infrared (LWIR) cameras (wide-field 35 deg × 46 deg and narrow-field 12 deg × 9 deg).

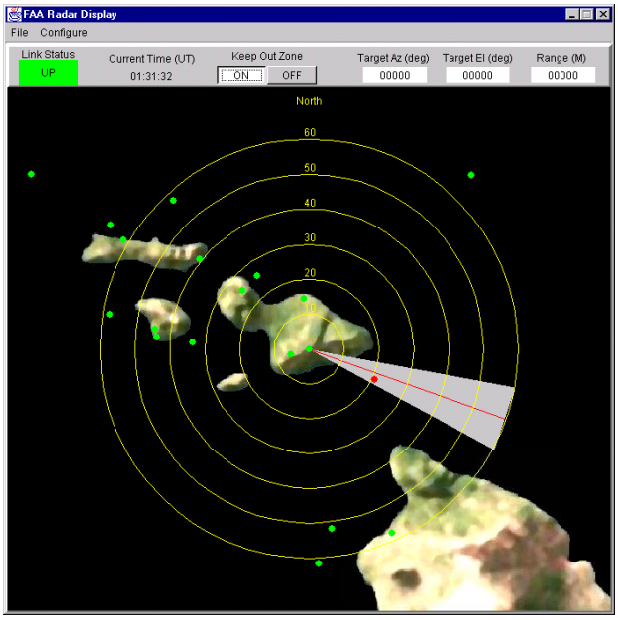


Fig. 6-17. RDI system developed by Boeing Technical Services for the AMOS. The system shows aircraft positions superimposed on a topographic map of the Hawaiian Islands. (The picture is courtesy of Capt. J. Snodgrass, Maui Space Surveillance Center, USAF.)

The second radar approach is that used during the Ground to Orbit Lasercomm Demonstration (GOLD) project [43]. An X-band radar system bore sighted with the telescope was used to detect aircraft at risk. The 60-mrad radar beam is much wider than the nominal 0.03-mrad laser beam. The radar detects aircraft at risk of entering the laser beam and triggers the laser beam shutter to interrupt lasing before the aircraft enters the laser beam. Depending on the location of the ground station, operating such a radar system may require frequency allocation by regulatory agencies and coordination with other users. The station may be assigned an operating frequency and bandwidth, and allowed to transmit up to a maximum radiated power. The performance specifications of the Primus-40 radar system that will be installed at JPL's OCTL are given in Table 6-3. A picture of the system is shown in Fig. 6-18.

The ability of the Primus-40 system to detect aircraft can be estimated from the manufacturer's specification and the radar Eq. (6.1-9) [45].

$$R_{\max} = \left[\frac{P_T G A_e \sigma}{(4\pi)^2 S_{\min}} \right]^{1/4} \tag{6.1-9}$$

In Eq. (6.1-9), PT is the power transmitted, G is the transmitter antenna gain, A_e is the effective receiver area, σ is the target cross-section, and $S_{\min}$ is the minimum detectable signal. The radar cross section is a function of the radar wavelength and the aspect angle of the target with respect to the incident beam. Reference [45] gives nominal cross sections at microwave frequencies for various objects at normal incidence. These are: 1 m^2 for a small single-engine craft, 40 m^2 for a large jet airliner, and 100 m^2 for a large jumbo jet. It should be noted that the cross section, and hence the radar return signal, varies with aspect angle. Notwithstanding this, the above cross sections do provide a guideline for bounding the radar detection capabilities.

Table 6-3. Primus-40 radar system parameters [44]

Parameter	Value
Average power, W	3
Peak power, kW	7
Pulse width, μs	1
Repetition rate, Hz	121
Operating frequency, MHz	9345
Bandwidth, kHz	375
Antenna diameter, m	0.45
Minimum detectable signal, dBm	–108

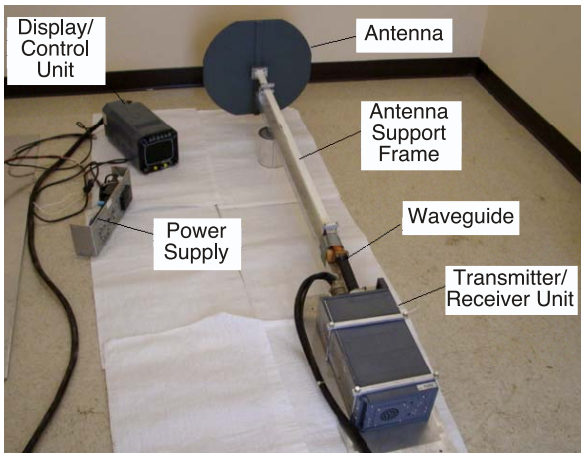


Fig. 6-18. Honeywell Primus-40 weather radar used for aircraft detection during the 1996 GOLD optical communications experiments.

From the target cross sections above and the data from Table 6-3, the maximum detection ranges for these targets are 26 km, 65 km, and 82 km, respectively. For mean sea level service ceilings of 7.6 km (20,000 ft), 12.2 km (40,000 ft), 13.7 km (45,000 ft) for these aircraft, the Primus-40 will detect these aircraft when operating at their service ceilings at ranges of 15 km, 29 km, and 34 km, respectively. The estimated radar returns for the three types of aircraft are 9, 14, and 15 dB above the minimum detectable signal, $S_{\min}$.

Tier-3 covers transmission in space and extends over the range covered by Earth-orbiting satellites. At this time, coordination of laser transmission with the LCH is not required for civilian agencies. For sites required to coordinate their transmissions, the procedure begins with the registration of the site and the laser. Key information includes:

- Laser site geodetics (latitude, longitude, and altitude)
- Laser output power peak and average
- Laser wavelength
- Beam divergence
- Operating aperture diameter
- Jitter angle

Based on the laser characteristics, the LCH would either issue a waiver stating that no coordination is needed for transmission or require that all laser transmission be coordinated. If coordination were required, the station would submit the details of the laser transmission for each operation, identifying the target satellites and pointing directions and times. The LCH would then respond with a listing of intervals when laser beam propagation is permitted. Civilian

agencies, such as NASA, although not required to coordinate with the LCH, may opt to do so for high-power laser beam transmission.

6.1.1.11 Concept Validation Experiments Supporting Future Deep-Space Optical Links. Ground-to-space optical beam propagation experiments began shortly after the invention of the laser, and several of the key challenges were identified in the early beam propagation experiments. Among these were: (i) reliably pointing narrow laser beams at spacecraft, (ii) effects of atmospheric scintillation on laser beam propagation, and (iii) effects of cloud cover. Experiments performed by JPL over the years have highlighted the challenges of propagating laser beams through the atmosphere. In some cases, these experiments have also demonstrated strategies to mitigate these deleterious effects. the difficulties of the challenge have been highlighted, and strategies to mitigate the effects have been demonstrated. In this section we discuss the JPL experiments from the Surveyor VII experiments through the GOPEX, Compensated Earth-Moon-Earth Retro-Reflected Laser Link (CEMERLL), and GOLD demonstrations, their results, and their contribution to the body of knowledge of laser communications.

In May 1962, Smullin and Fiocco demonstrated the first detection of a laser beam reflected from the lunar surface. They coupled a 50 joule per pulse ruby laser to a 30-cm telescope and transmitted the laser beam to the Albategnius, Copernicus Tycho, Longomontanaus regions of the Moon dark during the first quarter. The beam reflected from the lunar surface was detected by a photomultiplier at the Cassegrain focus of a 1.2-m telescope [46]. This lunar bounce experiment was followed by a series of retro-reflecting experiments to Earth-orbiting satellites. Yet, the effects of atmospheric seeing on a deep space optical link was not observed until 1968 when experimenters at Table Mountain (Wrightwood, California) and at Kitt Peak (Arizona) transmitted laser beams to the Surveyor VII lunar lander [47].

A 2.25-W continuous wave (CW) Ar:ion laser coupled through the full 0.6-m TMO[1] telescope was shuttered 30 s on and 30 s off. Losses (4.2 dB) in the telescope optical train resulted in 0.9-W optical power being transmitted in the 2–3 arc second atmosphere-seeing-limited beam. The 2-day long experiment was conducted during 08:15–09:26 and 10:30–11:10 Greenwich mean time (GMT) and on January 20, and during 08:30–8:41 GMT on January 21. The laser beams were detected by the vidicon on the spacecraft, and the images relayed back to Earth in real time. Figure 6-19 shows the picture of the uplink laser beams from the two stations in the non-Sun-illuminated area of the planet. The record shows that the detected uplink was sporadic. The TMF

[1] The Table Mountain Observatory (TMO) was renamed Table Mountain Facility (TMF) in the mid 1990s. Documents of different dates reflect these two terms for the same location.

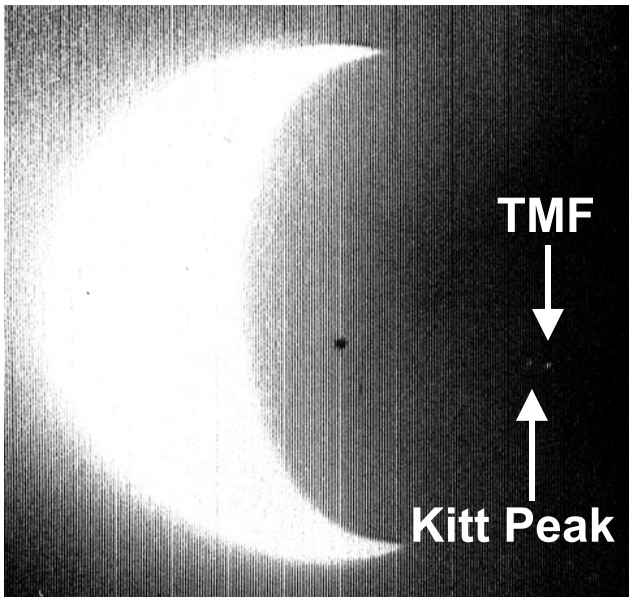


Fig. 6-19. Laser beams transmitted from Kitt Peak and Table Mountain detected by Surveyor VII vidicon camera on the Moon.

records of the experiment reported that the link was severely affected by the wind and seeing conditions at the site.

6.1.1.11.1 Galileo Optical Experiment (GOPEX). With the development of high power lasers and sensitive detectors, laser ranging experiments to Earth-orbiting satellites and to lunar retro-reflectors became more prevalent. Yet, the low percentage of returns from space-based retro-reflectors provided little experimental data to differentiate between tracking and scintillation-induced signal fades. What had not been demonstrated was reliable open-loop pointing of a narrow laser beam to a deep space probe. By 1992, there was enough uncertainty in precision pointing of laser beams to space that the opportunity provided by Galileo's VEEGA (Venus, Earth, Earth, Gravity Assist) trajectory (Fig. 6-20) was recognized as a unique chance to demonstrate an optical communications link to a spacecraft in deep space.

GOPEX was the first demonstration of open loop pointing of a laser beam from Earth to a deep space probe [48]. Uplink beams were transmitted from ground stations at the JPL Table Mountain Facility, Wrightwood, California and the Air Force Starfire Optical Range, Albuquerque, New Mexico to the Galileo spacecraft beginning at a range of 600,000 km out to 6,000,000 km [49]. The SOR provided cloud cover site diversity for the uplink. The experiment was conducted over the eight-day period December 9 through

December 16, 1992 during the second Earth Gravity assist segment of Galileo's trajectory, Fig. 6-20. The spacecraft's solid state imaging camera, Fig. 6-21, was used to detect the doubled Nd:YAG (532-nm) uplink laser transmission. The experiment's objectives were to:

- Demonstrate laser beam transmission to a spacecraft at deep-space distances.
- Verify laser-beam pointing strategies applicable to an optical uplink based solely on spacecraft ephemeris predicts.
- Validate models developed to predict the performance of the optical link.

The 90-deg Sun–Earth–spacecraft phase angle during the second Galileo flyby allowed nighttime laser transmission from the ground, providing excellent

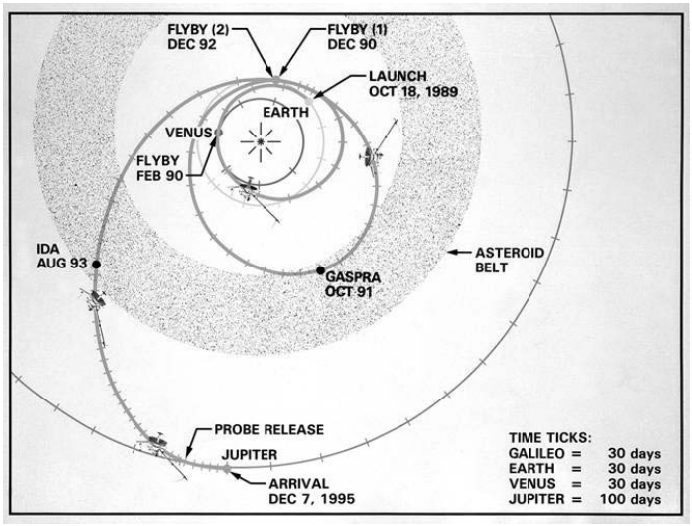


Fig. 6-20. The GOPEX demonstration was performed on the second Earth flyby of Galileo's VEEGA trajectory.

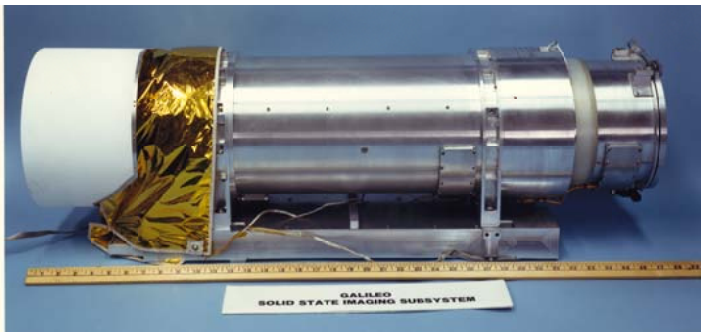


Fig. 6-21. Galileo solid-state imaging camera.

contrast of the uplink laser against a dark Earth background. The experiment was conducted between the hours of 3:00 a.m. and 6:00 a.m. Pacific Standard Time. Camera shutter exposure times were held to less than 800 ms to prevent saturation of the signal pixels by blooming of pixels imaging the sunlit part of the Earth.

The spacecraft instrument platform with the solid-state imaging (SSI) camera was scanned across the Earth, parallel to the Earth's terminator during each exposure, Fig. 6-22(a) and the pulsed uplink laser beams were imaged as a series of evenly spaced bright dots in the frame, Fig. 6-22(b). The laser pulse repetition rates from the two uplink facilities were set at 10 Hz for the SOR and 15 Hz for TMF.

The laser transmitters at both sites were frequency-doubled Nd:YAG lasers emitting at 532 nm coupled to a telescope through the coudé focus. The characteristics of both the SOR and TMF transmitters are given in Table 6-4.

At TMF, the telescope was the 0.6-m equatorial-mount astronomical telescope that had been used in 1968 to perform the laser transmission to the Surveyor VII spacecraft on the Moon. However, unlike in the Surveyor VII experiment, the laser beam illuminated a subaperture of the 0.6-m telescope. The telescope is f/36 at the coudé focus, and the appropriate beam-forming lens set was inserted into the coudé optical train to achieve the required laser beam divergence. From the coudé optical bench, the laser beam was reflected from a turning flat and through two turning mirrors before being incident on the 0.2-m secondary mirror. The beam then illuminated a 12-cm subaperture on the

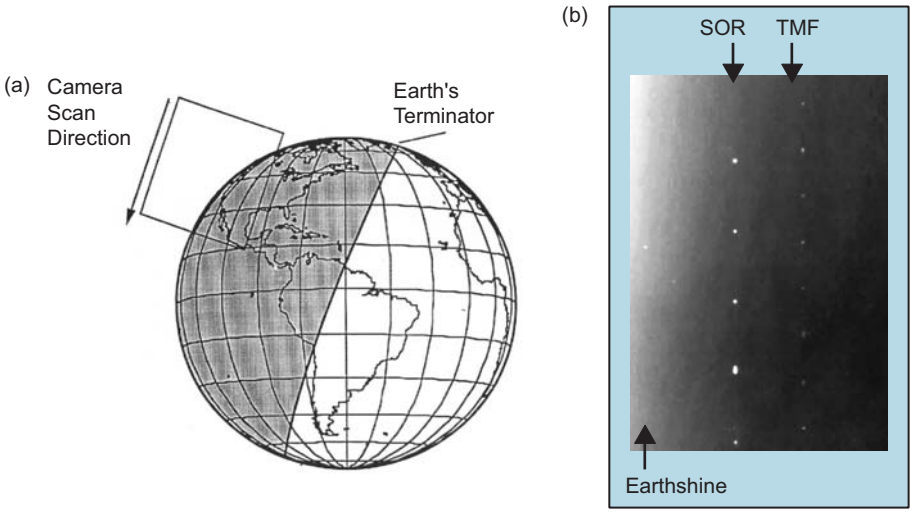


Fig. 6-22. SSI camera: (a) schematic showing how Galileo's SSI camera was scanned parallel to the Earth's terminator to capture SOR and TMF laser transmissions and (b) an SSI camera frame showing both the SOR and TMF laser beam transmissions.

Table 6-4. GOPEX laser transmitter characteristics.

Item	Table Mountain Observatory	Starfire Optical Range
Wavelength, nm	532	532
Pulse energy, mJ	250	350
Repetition rate, Hz	15–30	10
Pulse width, ns	12	15
Beam divergence, μrad		
Days 1–4	110	80
Days 6–8	60	40
Telescope mirror diameter		
Primary, m	0.6	1.5
Secondary, m	0.2	0.1
Optical train transmission	60 percent	43 percent

telescope primary. Subaperture illumination eliminated the power loss of a full-aperture illuminated system and allowed the propagation of the wide beam divergences needed to compensate for pointing errors and atmospheric beam wander effects. The full 0.6-m aperture of the telescope was used to collect light from the reference stars that were used to point the telescope to the spacecraft.

Two beam-forming lens sets, one for 110-μrad divergence and another for 60-μrad divergence, were designed to bring the laser beam to a focus at 1.3 km with the telescope focused at infinity. These wide-uplink beam divergences were several times atmospheric seeing, and thereby mitigated the effects of atmospheric tilt on pointing.

At the SOR, the flat-top laser beam was transmitted through the full 1.5-m telescope aperture with less than 10-percent loss from the secondary mirror obscuration. At both facilities, the requisite beam divergence was achieved by focusing the laser in the atmosphere. A thin-film-plate polarizer served as the aperture-sharing element, and coupled the laser output to the coudé optical train while allowing reference stars to be observed by the CCD camera positioned in the orthogonal leg of the optical train. The required laser beam divergence was achieved by focusing the outgoing laser beam at ranges of 40 km and 20 km, corresponding to 40-μrad and 80-μrad beam divergence, respectively.

Blind-pointing the telescope to Galileo was accomplished by implementing the following sequence of steps:

1) Point the telescope to a reference star located within 0.5 deg and northwest of the spacecraft.

2) Reset the telescope's right ascension (RA) and declination (Dec) readings to match the star's catalogued coordinates.
3) Two minutes and thirty seconds before transmission, center the star in the eyepiece at the telescope's coudé focus.
4) Ten seconds before transmission, point the telescope to Galileo and set the telescope to sidereal tracking for the next thirteen seconds.

This procedure was repeated during the 3-minute to 6-minute intervals between the laser transmissions to mitigate the effects of atmospheric refraction on the uplink beam pointing. Because the telescope calibration was performed just before transmission, the pointing errors introduced by mount sag were small [50]. In contrast, mount calibration at the SOR was performed only once per night, just prior to the night's experiment. Over the 8-day period, six reference stars of magnitudes 6 to 10 were used to point the telescopes to Galileo.

A summary of the detected laser detections over the 6-day GOPEX demonstration is given in Table 6-5. The uplink was detected on 56 of the 156 frames returned from the spacecraft. The uplink was detected on 82 percent (45 of the 67) of the frames with exposure times greater than 400 ms, and on only one of the 89 frames with exposures of less than 400 ms. Post-demonstration analysis showed that the loss in detection on these frames was due to an unanticipated bias in the scan platform direction. The loss of signal on the remaining 12 frames was due to a variety of causes including denial of laser beam transmission by the LCH, poor RF downlink, denial of opportunity by the Galileo Project office, and attenuation due to cirrus clouds. The large number of frames that were received with images of the uplink signal validated both the pointing strategies and the selection of uplink beam divergences.

Inclement weather, aborted transmissions, and restrictions imposed by regulatory agencies and by the Project Galileo team, temporary signal-to-noise anomalies on the RF downlink data transmission, and an unexpected camera-pointing bias error accounted for the loss of data on the remaining frames. The weather at TMF and SOR in the last 4 days of the demonstration validated the need for ground-station site diversity in optical communications. A Pacific storm dropped snow at TMF on the fourth day of the experiment precluding transmission from that location. The storm moved eastward and precluded transmission from the SOR for the remainder of the demonstration.

Analysis of the pulse strengths in the GOPEX frames showed the effects of scintillation fades. Calculations of scintillation variance from these and other GOPEX data showed that for both sites the data fit the strong turbulence theoretical models on all but the second day [51]. Figure 6-23 shows a 3-D image of one of the GOPEX frames. The fluctuations in the detected signal caused by atmospheric scintillation at TMF are clearly shown, and signal strength variations would clearly affect the quality of a command uplink to a

Table 6-5. Summary of laser signals detected during GOPEX.

Day	Shutter Speed (ms)	Frames Received	Frames with Detections
1	133	9 of 10	0
	200	24 of 25	0
	400	19 of 20	11[b]
	800	5 of 5	5
2	200	5 of 5	0
	267	15 of 15	0
	533	15 of 15	13
	800	5 of 5	5
3	200	5 of 5	0
	267	10 of 10	0
	533	5 of 5	5[c]
4	200	3 of 3	1[b,c]
	267	4 of 4	0
	533	3 of 3	1[c]
5		No GOPEX transmissions	
6[a]	133	3 of 3	0
	267	6 of 6	0
	533	3 of 3	3
7[a]	200	3 of 3	0
	400	3 of 4	3
	800	3 of 3	3
8[a]	267	2 of 2	0
	533	4 of 4	4
	800	2 of 2	2

[a] Adverse weather at SOR precluded laser transmission.

[b] Detection on some of the frames was based on one to three possible spots in the frame.

[c] Adverse weather at TMF precluded laser transmission; cloudy at SOR.

deep space probe. Subsequent JPL experiments would explore strategies to mitigate the effects of scintillation on the optical uplink.

6.1.1.11.2 Compensated Earth-Moon-Earth Retro-Reflected Laser Link (CEMERLL). The objective of the 1994 CEMERLL experiment was to evaluate the use of adaptive optics techniques in mitigating the effects of atmospheric turbulence on the uplink laser transmission [52]. The experiment was conducted jointly by JPL and the Air Force Research Laboratory's SOR

from March through September 1994 and transmitted both atmosphere-compensated and uncompensated laser beams to the retro-reflectors on the Moon [52].The target was the 300-element lunar Apollo-15 corner cube array located near Hadley Rille [53,54] (see Fig. 6-24).

The experiment was conducted during the first and last quarter phases of the Moon when the retro-reflectors were on the dark side of the lunar landscape [55]. This prevented saturation of the detectors by the sunlit lunar surface. The

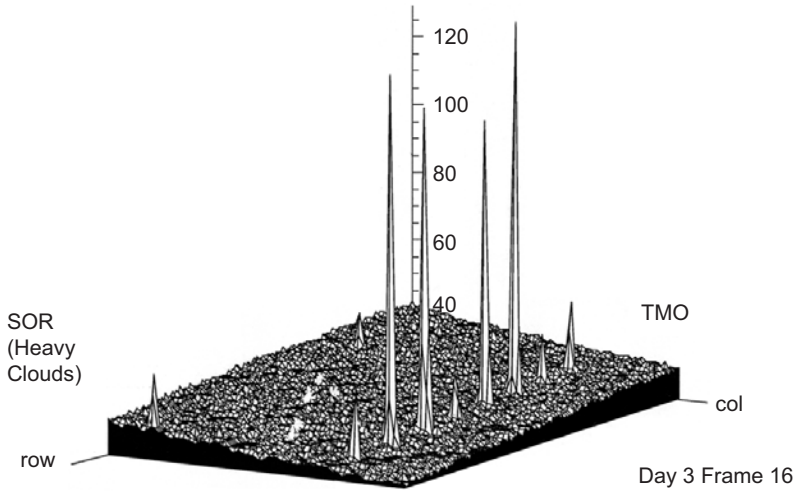


Fig. 6-23. Plot of received GOPEX uplink signal strength over a 400-ms interval.

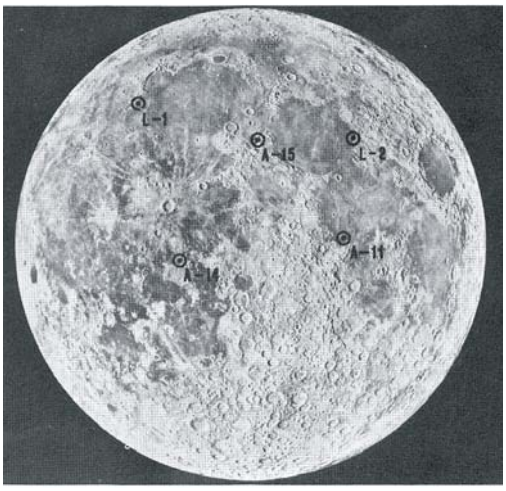


Fig. 6-24. Locations of Apollo 11, 14, and 15 and Lunakod 1 and 2 retroflectors on the moon [1].

retro-reflected laser pulses were detected at the 3.5-m SOR telescope over the three nights September 27, 28, and 29. In excess of 100 retro-reflected photons per pulse were detected when the beam was compensated. No returns were detected with the uncompensated beam.

The Rayleigh backscatter from a copper vapor laser (CVL) focused 10 km above the observatory created the laser guide star (LGS) for adaptive optics wavefront correction. Because the angular separation between the retro-reflectors and the closest exo-atmospheric guide star was well beyond the 12-μrad isoplanatic angle, tip/tilt compensation was not feasible. As seen from Earth, the closest stars passing by the edge of the Moon were approximately 3 mrad from the retro-reflectors. A Shack-Hartman wavefront sensor at the transmitter facility detected the wavefront aberrations of the CVL beam, and the wavefront reconstructor generated the commands to the deformable mirror (DM).

The Q-switched 1064-nm Nd:YAG scoring beam emitted 15-ns pulses at a 20-Hz pulse repetition rate. The laser uplink was coupled through the telescope's coudé optical path. The energy per pulse transmitted from the telescope to the Moon was 340 mJ. The scoring beam was made incident on the DM, pre-distorting the beam's wavefront so that it was well corrected at 10 km above the telescope. Under typical conditions, the seeing-limited divergence of the 1064-nm laser was typically 17 μrad. With the higher order adaptive optics loop closed, the uplink beam divergence decreased to less than 3 μrad.

The receiver was a Rockwell Science Center solid-state photomultiplier (SSPM) detector in the focal plane of the 3.5-m SOR telescope was the receiver. The detector quantum efficiency was 6–8 percent at 1.06 μm, and its 800 × 800 μm active area corresponded to a 70-μrad FOV at the f/5.6 Nasmyth focus. The detector was integrated to a high-bandwidth trans-impedance amplifier, and the assembly was enclosed in an Infrared Laboratories liquid-helium dewar cooled to 11 K. A 10-nm and a 1-nm optical filter were inserted in front of the detector to suppress the infrared signature from the dark side of the Moon.

The round trip light time between Earth and the lunar array is approximately 2.7 s; a value that continuously changes as the Moon transits the sky. The detector electronics were range gated to protect the SSPM from saturating. The sequence for initiating the detection window began with a trigger signal generated by the outgoing Nd:YAG laser at the 1.5-m telescope. The SOR calculated the expected time of arrival of the return signal at the 3.5-m telescope was calculated from the ephemeris predicts and the timing mark generated by the outgoing laser pulse. A time interval counter kept track of the delay at the 1.5-m facility. At the appropriate time, the counter sent a trigger signal to the scalar averager at the 3.5-m facility to initiate the detection window on a dedicated line between the facilities of known transmission delay. The detector trigger delay was adjusted to accommodate for the finite transit

time of the electronic signal between the 1.5-m and 3.5-m facilities. In addition, the delay was continuously adjusted to correct for the changing RTLT to the retro-reflectors as the Moon transited the sky.

Figure 6-25 is a scalar averager display of the return signals in an 82-ms interval. The vertical scale is 2 counts/division. The detection threshold was set at 210 mV, corresponding to 30 photons. This reduced the probability of false alarms while still retaining the ability to detect retro-reflected returns. The figure clearly shows the returns occurring 50 ms apart, an interval corresponding to the 20-Hz laser pulse repetition rate.

Of the more than 200,000 pulses transmitted, approximately 30 returns were detected all over the three-day period September 27 through September 29. Transmission strategies included (i) direct pointing and (ii) scanning the beam in 1-μrad increments over a 15-μrad × 15-μrad space around the location of the retro-reflectors. Although the highest percentage of returns (0.1 percent) was observed when scanning on September 28, this approach did not yield the same degree of success on September 29. The low and erratic frequency of returns was attributed to the uncompensated wavefront tip/tilt in the transmitted beam.

6.1.1.11.3 Ground to Orbit Lasercomm Demonstration (GOLD). The ETS-VI satellite was launched by Japan's NASDA on August 28, 1994 as a test bed to assess a variety of telecommunications scenarios over a wide range of frequencies. The satellite payload included the laser communications equipment (LCE) to demonstrate space-to-ground laser communications [56]. Failure of the apogee kick motor prevented geostationary positioning and left the satellite in a geo-transfer orbit with the spacecraft visible from ground facilities around the globe. NASDA transferred the satellite into a 3-day sub-recurrent orbit in November 1994 to facilitate the performance of communication experiments around the globe. Figure 6-26 shows the ground track of the satellite. The GOLD experiments with TMF were conducted during the section of the apogee pass when the satellite was over Central America [57,58].

The experiments were conducted from November 1995 to May 1996, except for a one-month break in January 1996 when the spacecraft went into eclipse and lasercom activities were shut down to conserve power. The experiment objectives were to:

- Demonstrate two-way spatial acquisition/tracking of laser beams with a spacecraft
- Accomplish one-way and two-way optical data transfer to a spacecraft and measure bit error rates
- Accumulate 10 elapsed hours of transmission/reception experience and 30 GB over a 6 month period

- Compare downlink atmospheric transmission losses with similar data from the Table Mountain Facility's Atmospheric Visibility Monitoring (AVM) observatory

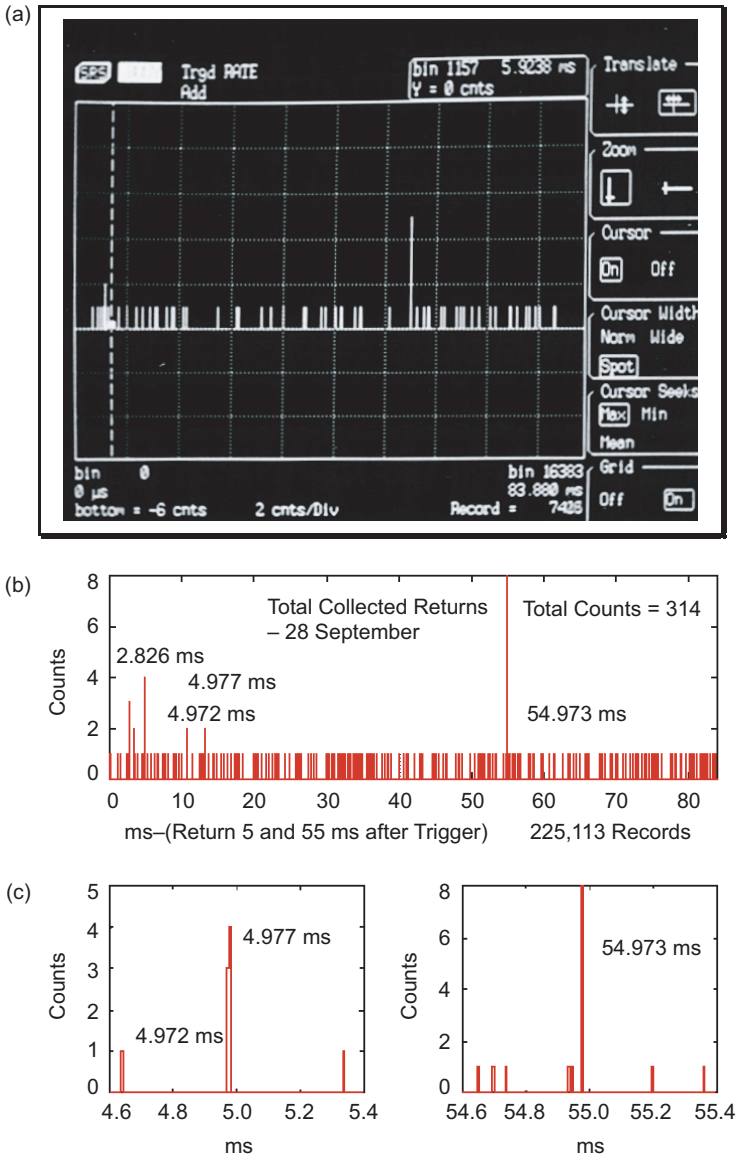


Fig. 6-25. (a) photo of scalar average output from September 28 transmission showing "hits" approximately 5 and 55 ms after trigger, (b) typical laser returns window covers 82 ms with 16,000 bins 5.12-µs wide, and (c) expanded views showing two areas of interest.

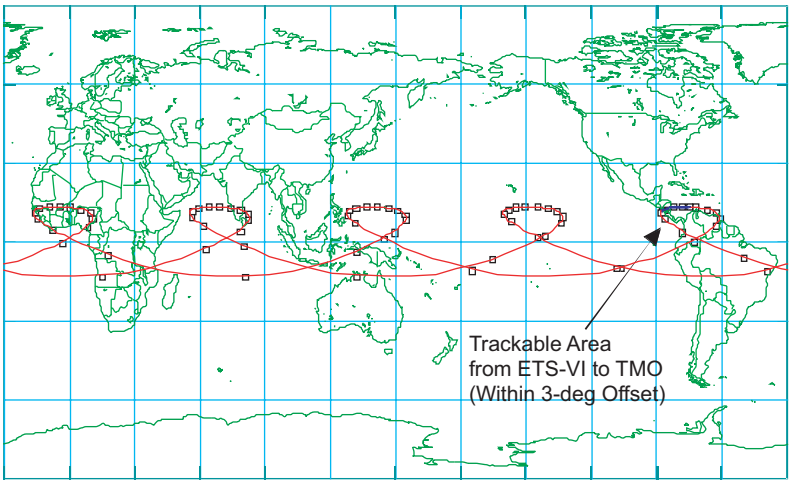


Fig. 6-26. Ground track of ETS-VI satellite showing trackable area from TMF.

The experiment was highly successful and accomplished of the following firsts:

- Bi-directional optical link between a ground station and a spacecraft at geostationary ranges.
- Demonstration of mitigation of atmospheric scintillation by using a multi-beam uplink.
- Demonstration of regeneration of an optical uplink signal to a spacecraft.
- Demonstration of the return of real data on an optical space-to-ground link.

Designed for a GEO-to-ground link, the LCE terminal was body-fixed to the ETS-VI satellite with limited coarse tracking of the LCE's gimbaled flat mirror. A combination of satellite attitude control and LCE gimbal mirror pointing was required to point the LCE to the ground. An involved and time-consuming process, communications opportunities were limited to apogee passes above the ground stations. Figure 6-27 shows the control region of the satellite attitude control system (ACS) and the LCE gimbal and the elevation and azimuth orientation of TMF in the spacecraft coordinate system. Markers are separated by 1 hour, and the trace shows approximately a 5-hour experiment window. Because the LCE was designed for geostationary operation over Japan, in a geo-transfer orbit, and at the higher latitude TMF ground station, the LCE could not center TMF in its region of control even with the assistance of the satellite ACS.

The ETS-VI satellite had limited onboard memory, and the LCE and satellite ACS control had to be updated in real time to keep TMF in the LCE's FOV. This was done by commands originating at the Communications

Research Laboratory (CRL) in Tokyo, Japan. The satellite ephemeris for pointing the telescopes to the ETS-VI were prepared from NASDA-generated orbit solutions that were sent to JPL twice weekly [59]. Figure 6-28 shows this command and data flow chain during the experiment. The procedure in chronological order was:

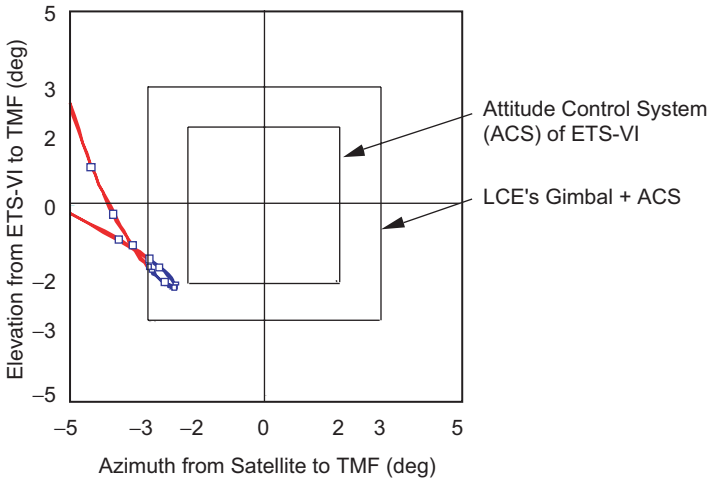


Fig. 6-27. Plot of control region of combined ACS and LCE gimbal for November 26, 1995, pass over TMF, after the first orbit maneuver.

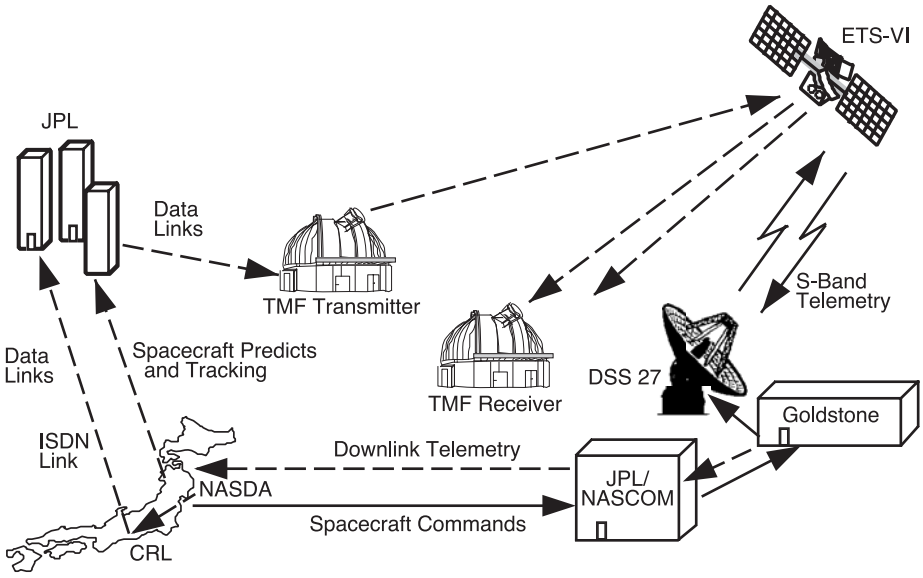


Fig. 6-28. Operations approach for GOLD: TT&C are sent from NASDA to ETS-VI via JPL/NASCOM; DSS 27 at Goldstone commands ETS-VI.

1) CRL generated LCE and spacecraft attitude pointing commands and transmitted them to NASDA.
2) NASDA relayed the commands to the JPL NASCOM and on to Goldstone, California.
3) Deep Space Station 27 (DSS 27) uplinked the commands to the ETS-VI on an S-Band link.
4) DSS 27 received the satellite telemetry and data and routed it to NASDA.
5) NASDA transferred the attitude control system (ACS) and the LCE coarse pointing and other related GOLD data to the CRL.
6) CRL immediately processed ACS and LCE data to confirm that the commands were positioning the LCE to point to TMF. This process was iterated incrementally until the combination of satellite attitude and LCE pointing placed TMF in the LCE FOV [60].
7) CRL transmitted the CCD data to TMF on a TMF-to-CRL Integrated Services Data Network (ISDN) link to allow TMF operators at TMF to validate the acquisition of the uplink laser beam.

The acquisition of the uplink was further confirmed by the direct optical downlink from the satellite.

GOLD Transmission Record. The GOLD transmission record is shown in Fig. 6-29 and Fig. 6-30. There were 22 opportunities to transmit to the satellite in each of the two GOLD demonstration phases. Phase-1 transmissions extended from October 30 to January 13 when the satellite went into eclipse and NASDA precluded all but critical satellite support to conserve battery power. In Phase 1, there were 12 two-way transmissions, 1 one-way transmission, 5 cancellations due to bad weather, 2 cancellations due to hardware failures, and 2 non-detections. Phase-2 transmissions began on March 21, 1996, shortly after the satellite had emerged from the apogee eclipse and continued to May 26, 1996. During this phase, there were 12 successful two-way

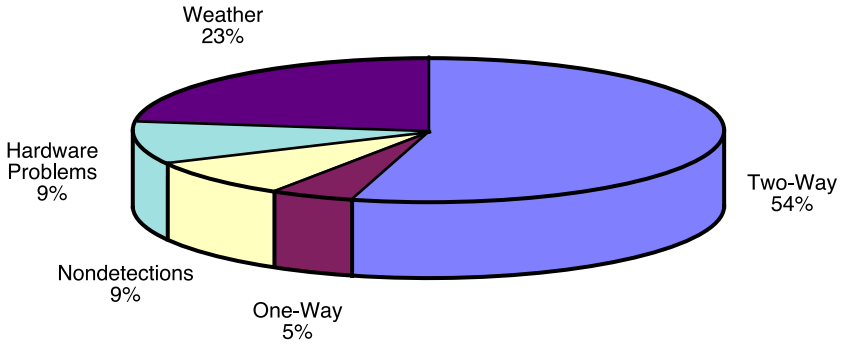


Fig. 6-29. Pie chart of the GOLD Phase-1 transmission record from October 30, 1995, to January 13, 1996.

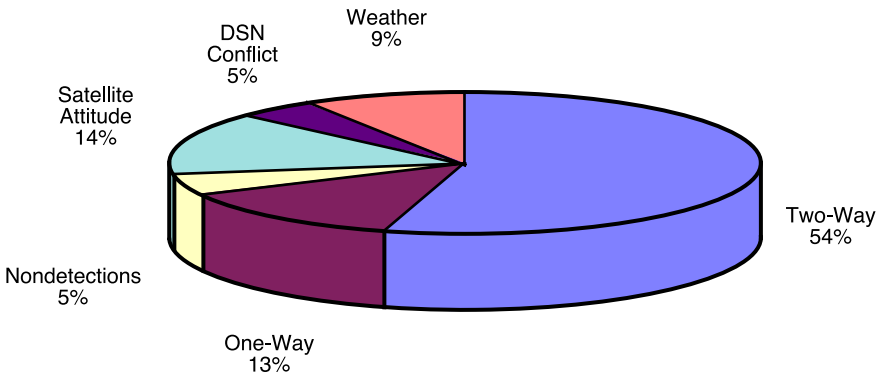


Fig. 6-30. Pie chart of the GOLD Phase-2 transmission record from March 21, 1996, to May 26, 1996.

transmissions, 3 one-way transmissions, 1 non-detection, 2 cancellations due to bad weather, and 3 cancellations due to satellite attitude-control problems.

GOLD Uplink. The TMF transmitter and receiver for GOLD were designed and developed over the 4-month period from July to November 1995. Figure 6-31 (a) is a schematic of the transmitter optical train. The transmitter telescope was the 0.6-m that had previously been used for Surveyor-7 and GOPEX demonstrations. The Ar-ion laser seen in Fig. 6-31 (b) was operated on the 514.5-nm line with 13-W continuous wave (CW) output, and was coupled to the telescope through its coudé path. The laser beam was modulated at 50-percent duty cycle to provide the required discriminant for the LCE tracking loop. The modulator consisted of four potassium dideuterium phosphate (KD*P) crystals and a polarizer. A 6000-bit error rate tester (BERT)—was used to generate a basic data pattern of 0 to 1 V square wave modulation that was amplified to the modulator's half-wave voltage in the driver, with zero volts corresponding to maximum transmission through the modulator. After modulation, the beam was incident on a concave/convex lens pair that set the beam divergence from the telescope; nominally set at 20 μrad.

A multi-beam uplink approach was used to mitigate the effects of atmospheric turbulence [61]. Here, the laser output was divided into two or four beams, and the beams kept spatially and temporally incoherent. Spatial incoherence was accomplished by separating the beams at the telescope aperture by a distance greater than the coherence cell of the atmosphere. Temporal incoherence was achieved by delaying the path length difference between the beams so that it exceeded the coherence length of the laser.

For the two-beam transmission, the beam splitter separated the beam into two equal parts; one that went through a 25-cm optical delay line with a path length difference greater than the laser's coherence length ($\lambda^2/\lambda\Delta \sim 10\,\mathrm{cm}$).

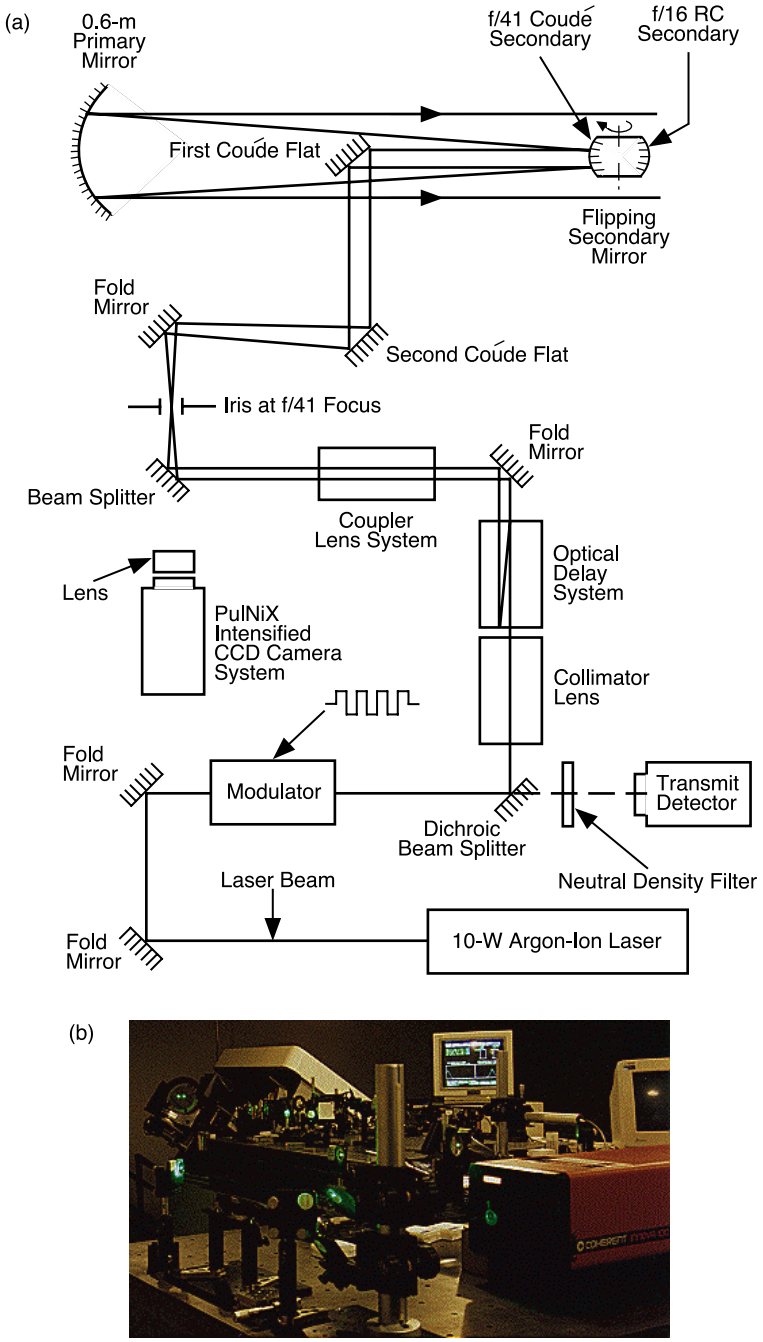


Fig. 6-31. GOLD: (a) schematic of the uplink optical train and (b) photograph of the transmitter optical train, showing the argon-ion laser Conoptics modulator and, to the far left, the M5 coudé mirror.

Both beams were reflected from a high power dichroic beam splitter and were brought to a focus at the iris located at the f/41 focus of the telescope. From there, the beams diverged and were reflected by the third coudé flat and into the telescope. The beams were made incident on opposite sides of the 0.6-m telescope's primary mirror; a distance greater than the size of an atmospheric coherence cell. See Figs. 6-32(a) two-beam and 6-32(b) four-beam transmissions. The use of spatial and temporal incoherent beams resulted in a more uniform and constant intensity profile at the satellite.

Figures 6-33, 6-34, and 6-35 show the signal strength detected by the LCE for the one-, two-, and four-beam uplink transmissions along with histograms showing the frequency of detection of a given signal strength [61]. These data clearly show the advantage of multi-beam uplink beacon. For the single-beam uplink, the median signal strength detected was 1.38 V, and the mean was 1.35 V. The data show surges of 5.7 dB and fades –12.3 dB from the mean. The histogram shows a distribution with the most frequent signal detected at 0.08 V, almost zero. The two–beam uplink showed a significant improvement in surge reduction but marginal improvement in reducing fade depth. The data did show a significant reduction in fade frequency, however. The median received signal strength for the two-beam uplink was 2.4 V. The mean was 2.19 V. Signal surges were 1.9 dB above the mean, and signal fades were –10.4 dB from the mean. For the four-beam uplink, the median signal strength was 3.08 V, and the mean was 2.93 V. Signal surges were 1 dB above the mean, and fades were –4.2 dB below the mean; a significant reduction in fade depth and fade frequency over two-beam propagation.

GOLD Downlink. The LCE aluminum gallium arsenide (AlGaAs) laser on board the ETS-VI transmitted 14-mW, average power at 830-nm current modulated. The 1.024-Mbps downlink signal was Manchester coded and

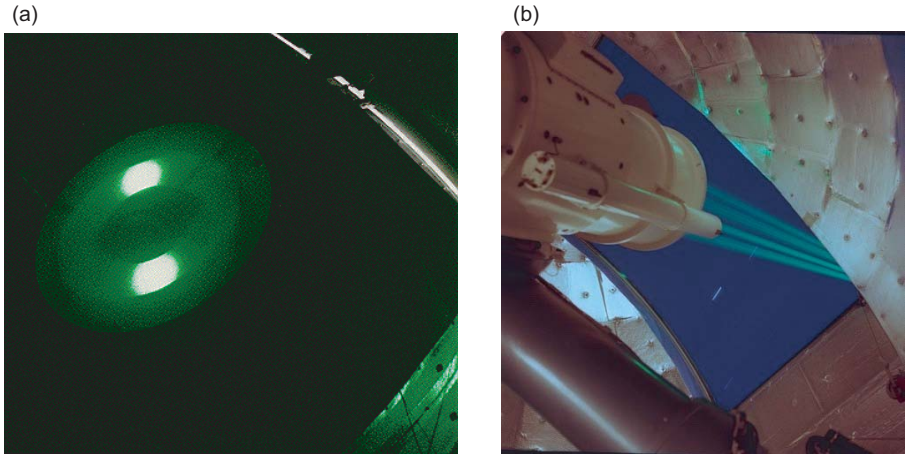


Fig. 6-32. GOLD uplink: (a) two-beam laser output projected onto a scoring board in the dome of the 0.6-m telescope and (b) four beams transmitted through the dome.

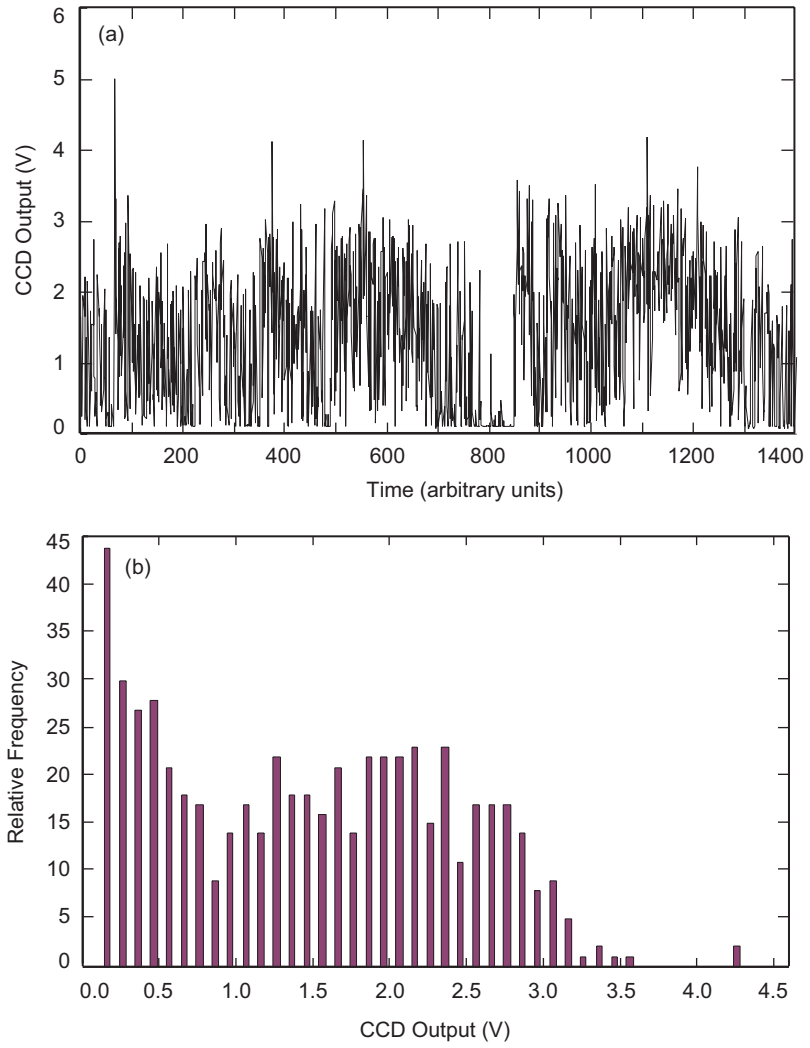


Fig. 6-33. (a) uplink signal strength detected by LCE CCD camera with single beam transmitted to ETS-VI spacecraft and (b) histogram of signal strength detected for single-beam uplink.

transmitted through a 7.5-cm aperture telescope. A schematic of the receiver optical train is shown in Fig. 6-36 (a). The receiving telescope was a 1.2-m bent Cassegrain located approximately 60 m from the 0.6-m transmitter [62]. The downlink signal reflected from the telescope was reflected onto the optical bench attached to a flange at the telescope Cassegrain focus. See Fig. 6-36 (b). The satellite downlink was reflected from a dielectric mirror and through a notch filter to separate the signal from the solar illuminated satellite image. The signal was then made incident on the APD detector. Light from the Sun-

illuminated satellite was evenly split between the tracking and seeing cameras. A red-filter was introduced into the optical path to suppress the sky background during daytime experiments.

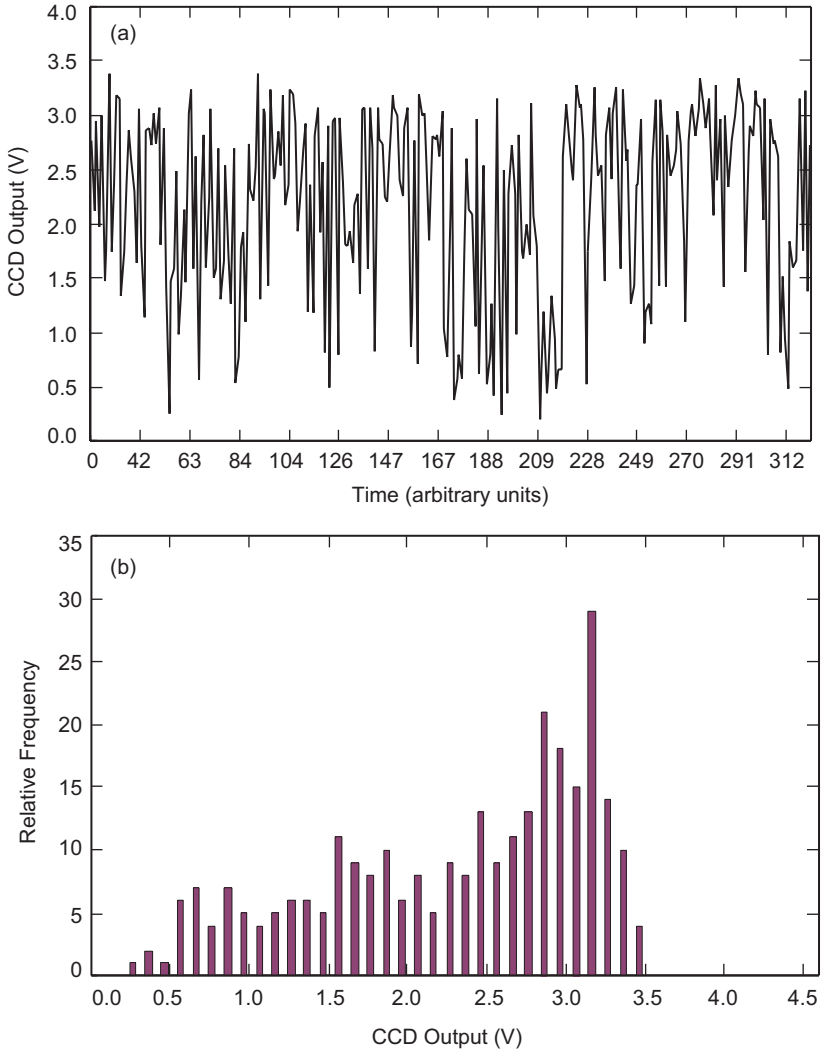


Fig. 6-34. (a) signal strength detected by CCD for two-beam uplink and (b) histogram of detected signal for two-beam uplink.

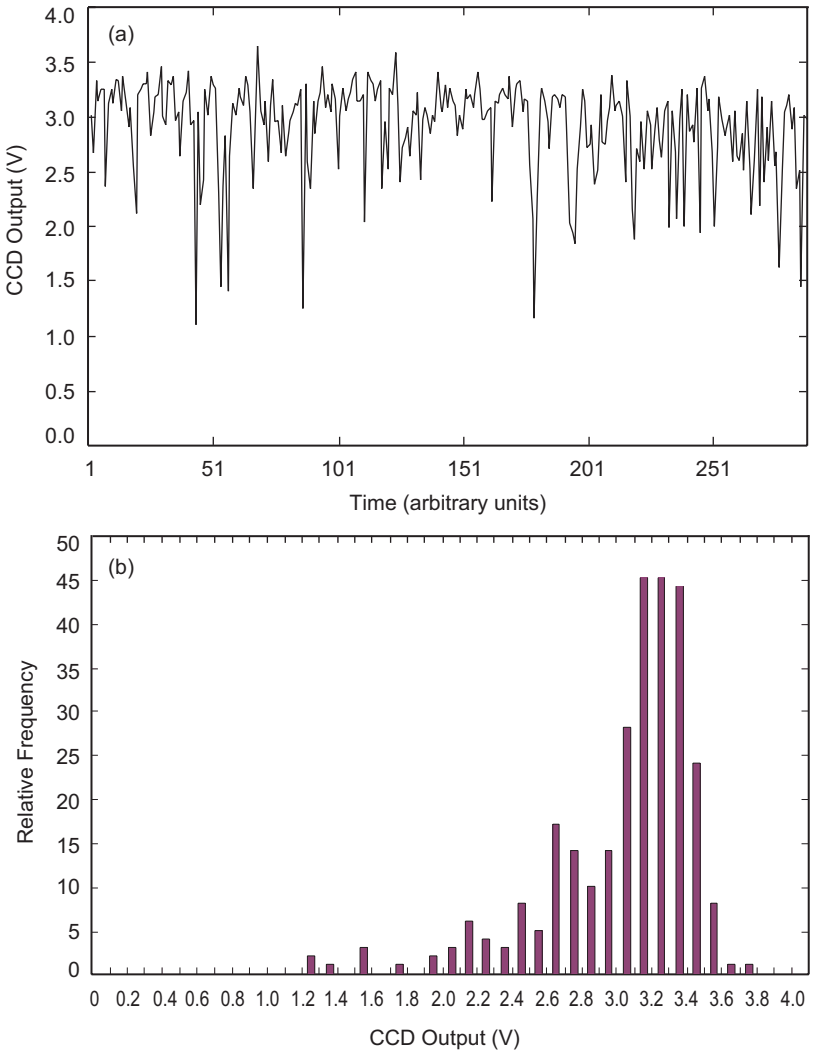


Fig. 6-35. (a) signal strength detected by CCD for four-beam uplink and (b) histogram of detected signal for four-beam uplink.

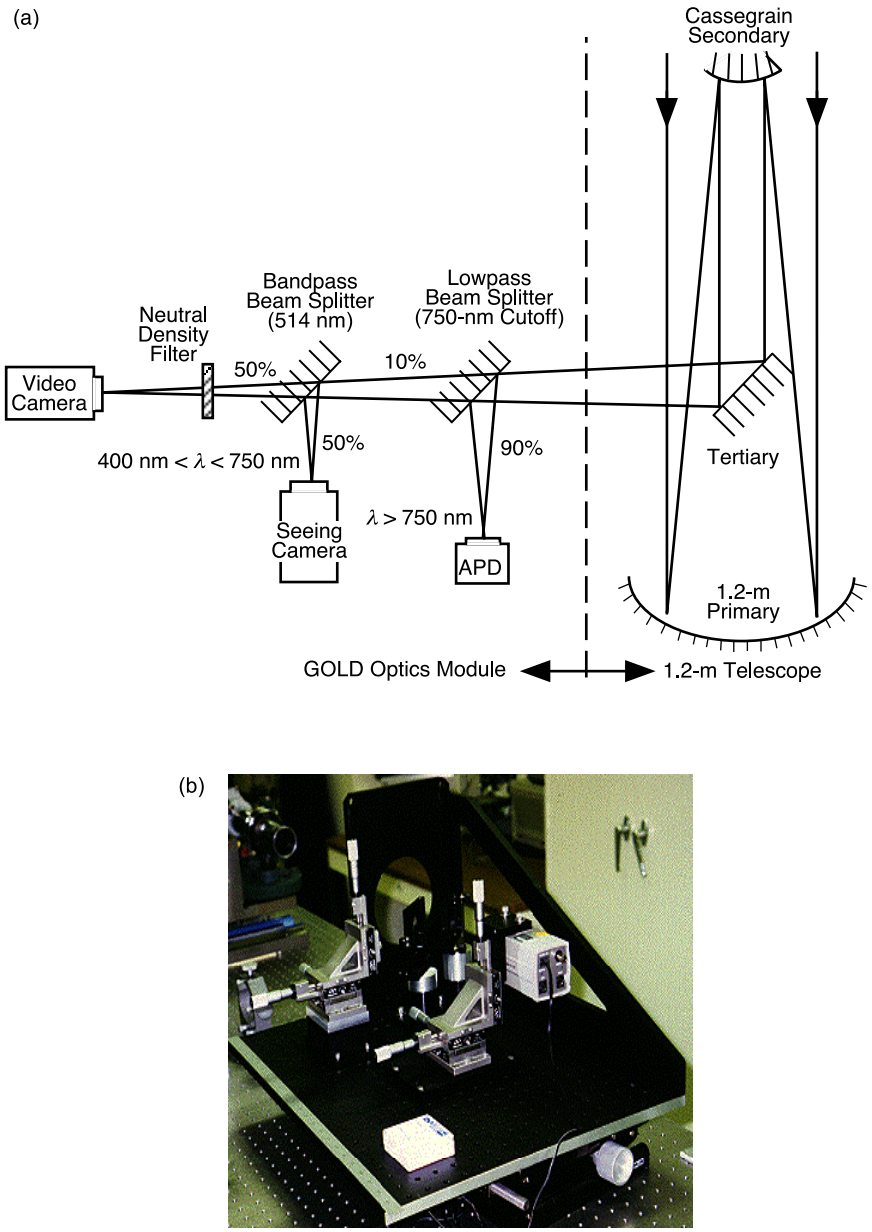


Fig. 6-36. Receiver optical train: (a) schematic of the 1.2-m receiver telescope at TMF, showing tracking, seeing, and receiver detectors, and (b) photograph of the optical train, with detectors and cameras, that was attached to the Cassegrain focus flange of the 1.2-m telescope.

Figure 6-37 shows a block diagram of the GOLD receiver. The signal detected by the APD was amplified, and the output simultaneously sent to a Tektronix 802-A digital oscilloscope (that recorded the downlink signal waveform) and to a bit-error rate tester for the PN code downlink. The BERT's 1-Hz output was recorded on a PC. A bit synchronizer and a PCM recorder were used to recover the downlink telemetry [63].

The three data types received at the ground are shown in Figs. 6-38, 6-39, and 6-40. Figure 6-38 is a representative sample of the PN-coded 1.024-Mbps data stream that was used to measure the performance of the optical link. The 128-kbps spacecraft instrument data stream repeated eight times to fill the 1.024-Mbps downlink channel, as shown in Fig. 6-39. Figure 6-40 shows the 1-MHz square wave data stream that was regenerated on the spacecraft and retransmitted to the ground. The dropouts in the regenerated signal are most likely caused by atmosphere-induced signal fades on the uplink. In addition to the square-wave pattern, a 1.024-Mbps Manchester-coded 511-bit-long PN sequence was transmitted on a two-beam uplink to the satellite. The recorded atmospheric seeing during the experiment was 2 arc-sec, and the measured BER ranged from 3×10^{-4} to 2.5×10^{-3} over the forty-minute measurement interval.

6.1.1.12 Conclusion. JPL has performed a series of free space-optical communications demonstrations and has developed strategies both for safe ground-to-space laser beam propagation and for mitigating the effects of atmospheric scintillation.

To meet the needs of continuous coverage of the deep-space optical communications mission, we have studied a variety of single-aperture ground receiver options to support the link. Recognizing the need for site diversity to mitigate the effects of cloud cover, we have explored approaches to reduce the cost of large 10-m class single apertures that range from the RF Leighton dishes to lightweight glasses. While some of these approaches show promise, the challenge is to reduce scattering and to achieve the requisite performance during the daytime at small Sun angles. Our preliminary adaptive-optics

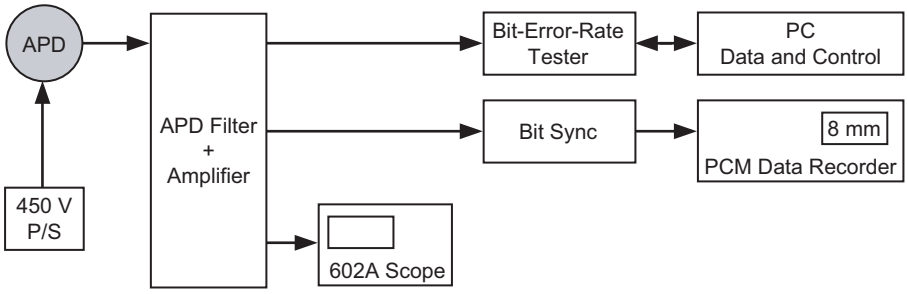


Fig. 6-37. GOLD receiver block diagram.

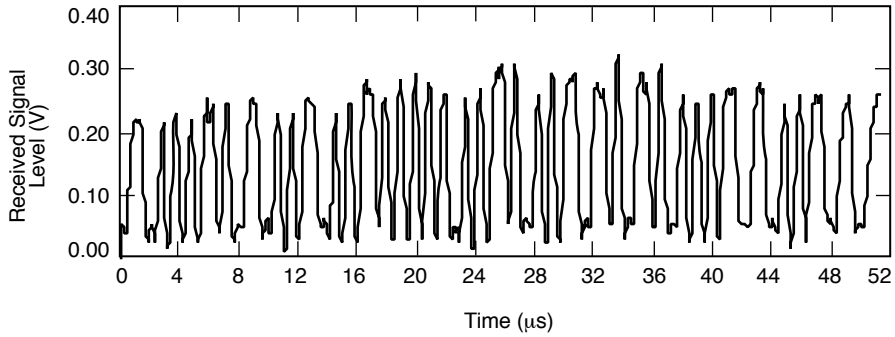


Fig. 6-38. Sample of 1.024-Mbps Manchester-coded PN sequence downlink telemetry from the LCE showing random bit flips.

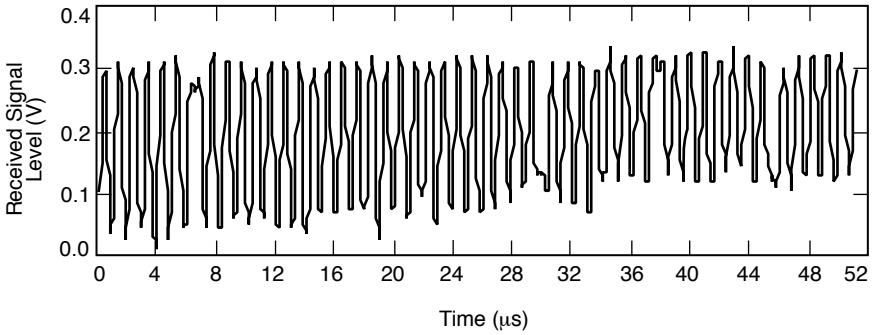


Fig. 6-39. Downlinked satellite telemetry at 128 kbps. The data show bit flips in multiples of eight consistent with 8X repetition of the bit pattern.

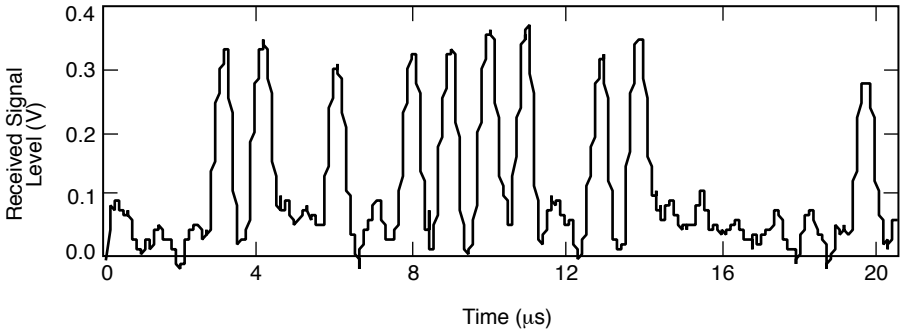


Fig. 6-40. Section of a 1-MHz square-wave regenerated from a 1-MHz uplink to the satellite.

analysis suggests that an actuator density of 900 actuators per square meter can enable FOV reduction at the deep-space receiver and background suppression, both of which can increase the capability to allow pointing at small Sun angles.

On the uplink, JPL has demonstrated the benefit of multi-beam propagation in the 1995–1996 GOLD demonstration. Yet, although this approach mitigates

uplink scintillation and supports blind pointing to the deep-space probe, its atmosphere-limited beam divergence restricts the range of application to spacecraft at Mars for reasonable uplink laser powers. In the CEMERLL demonstration, JPL and the AFRL showed that using adaptive-optics techniques on the uplink will require tip/tilt in addition to higher-order wave front correction. Yet, because the point-ahead angles for deep-space missions exceed the isoplanatic angle, using the downlink for uplink tip/tilt correction is not feasible. Polychromatic laser guide star techniques offer the possibility of tip/.tilt correction across the sky without the need for a natural guide star. This technology is currently under development.

6.1.2 Optical-Array Receivers for Deep-Space Communication

Victor A. Vilnrotter, Chi-Wung Lau, and Meera Srinivasan

6.1.2.1 Introduction. Earth-based reception of deep-space optical signals using an array of relatively small telescopes together with high-speed digital signal processing is a viable alternative to large-aperture telescopes for receiving deep-space optical signals. Large-aperture telescopes are costly to build and operate, and inherently suffer from single-point failure in case of malfunction, thus jeopardizing precious data. Performance of a properly designed array tends to degrade gracefully in case of element failures, even without replacement, but the array approach also provides the option to switch in spare telescopes in case of failure, without a significant increase in cost.

In addition to a favorable cost and risk trade-off, the optical-array receiver approach has advantages in terms of implementation complexity and performance for several key communications functions that need to be clarified and evaluated. These characteristics can best be explained in terms of an array-receiver model that emphasizes the communications aspects of the optical array. We begin by developing the underlying concepts governing the behavior of optical arrays. Next, similarities and differences between single-aperture and array telescopes designed for reception of deep-space telemetry are explored, followed by a detailed investigation of communications performance. We conclude with a comparison of array receivers with the more conventional large-aperture optical receiver under realistic operating conditions.

6.1.2.2 The Optical-Array Receiver Concept. The essential difference between a single-aperture optical communications receiver and an optical-array receiver is that a single aperture focuses all of the light energy it collects onto the surface of an optical detector before detection, whereas an array receiver focuses portions of the total collected energy onto separate detectors, optically detects each fractional energy component, and then combines the electrical signal from the array of detector outputs to form the observable, or “decision statistic” used to decode the telemetry signal. A single-aperture receiver need

not be constructed from a single monolithic glass lens or reflector element; large modern telescopes generally are constructed from hexagonal segments with each surface placed near its neighbors and monitored to form a single parabolic surface. If the image of a point source is formed using every segment of the surface and then detected, that collection of segments would be considered a single-aperture receiver. However, if the focal spot of each segment were separated from the rest and detected with a separate detector element or focal-plane detector array, that would constitute an array receiver in the framework of our definition. Note that a focal-plane array in a single telescope might be construed as an optical-array receiver under this definition, but the intent is for there to be multiple apertures. This construction serves as a convenient conceptual vehicle to illustrate the similarities and differences between single-aperture and multi-aperture array receivers.

A conceptual block diagram of an optical array receiver suitable for deep-space telemetry reception is shown in Fig. 6-41, intended to identify the key components required for array receiver operation. The most conspicuous feature of an optical array receiver is a large number of small- to medium-sized telescopes, with the individual apertures and number of telescopes designed to make up the desired total collecting area. This array of telescopes is envisioned to be fully computer controlled via the user interface and predict-driven to achieve rough pointing and tracking of the desired spacecraft. Fine-pointing and tracking functions then take over to keep each telescope pointed towards the source despite imperfect pointing predicts, telescope drive errors, and vibration caused by light wind.

The optical signal collected by each telescope is focused onto a detector array located in the focal plane, designated as the focal-plane array (FPA). Despite atmospheric turbulence degrading, the coherence of the received signal fields and interfering background radiation entering the receiver along with the signal within the passband of the optical pre-detection filter, the FPA and associated digital signal-processing electronics extract real-time pointing information, keeping the centroid of the turbulent time-varying signal centered over the FPA. This is accomplished in two steps: first, large accumulated pointing errors are fed back to the telescope drive assembly, repointing the entire instrument; and second, small excursions of the signal distribution from the center of the array are corrected via a fast-response tip-tilt mirror that responds to real-time pointing updates generated by the FPA electronics. In addition, the short-term average signal energy over each detector element is measured and used to implement various adaptive background-suppression algorithms to improve or optimize the communications performance of the entire array.

The electrical signals generated at each telescope by the FPA signal-processing assembly are collected at a central array signal-processing station, where the final operations necessary for data decoding are performed. The

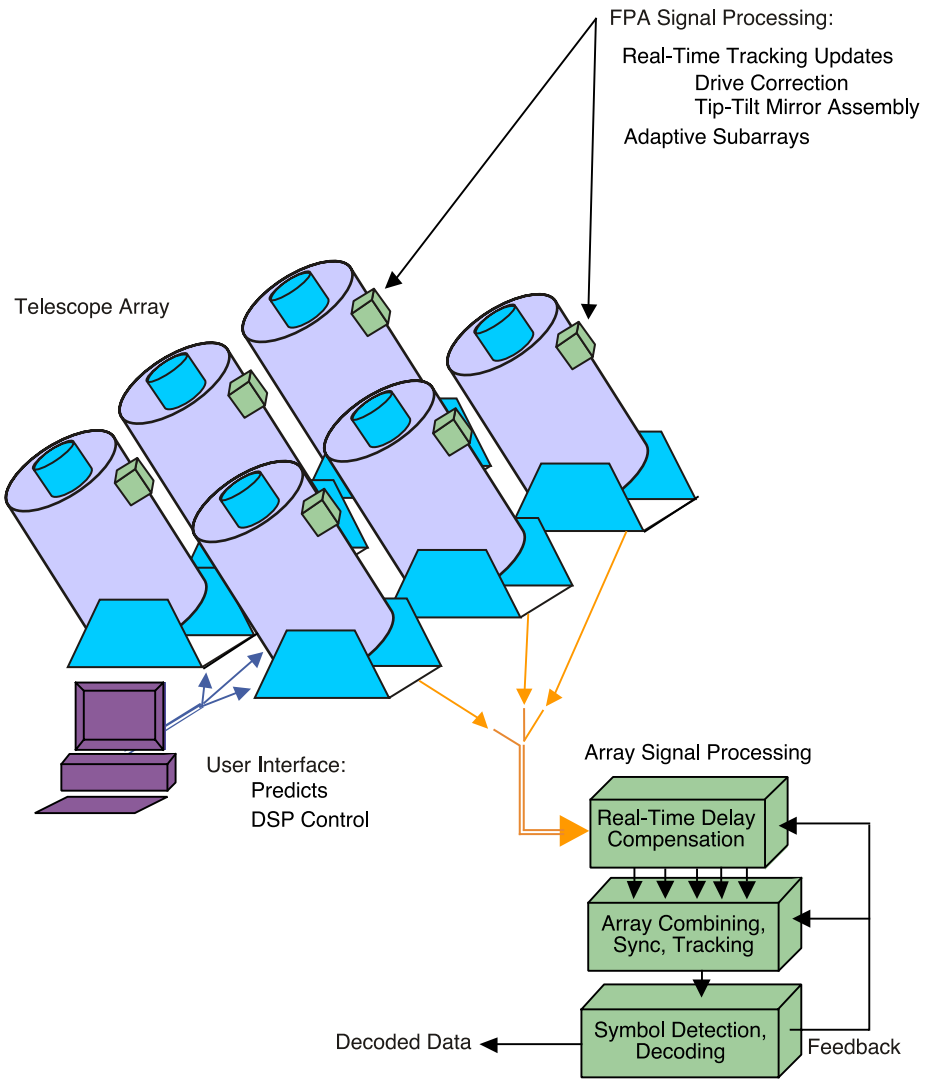


Fig. 6-41. Conceptual design of an optical array receiver, illustrating key functions required for array reception: the telescope array; focal-plane pre-processing of the signal at each telescope; and the additional array processing needed to combine the pre-processed signals and decode the received data.

relative time delay between signals from various telescopes is measured and removed, effectively aligning each signal stream in time before further processing is carried out. The delay-compensated signals then are combined in a manner to optimize array performance, and further processed to affect symbol synchronization, frame synchronization, Doppler compensation, and maintain lock between the combined signal and the receiver time frame. At this point,

the transmitted channel symbols are detected, and the decoding operation begins. The detected channel symbols now could be handed over to a hard-decision decoder, or various forms of soft-decision decoding could be implemented through the use of additional side information along with the detected symbols. The channel symbols detected by the array receiver also may be used to aid the delay compensation, array combining, and synchronization operations.

6.1.2.3 Aperture-Plane Expansions. It is common practice in the optical communications literature to expand random fields at the telescope aperture into spatial modes that are closely related to diffraction-limited fields of view, clustered to form a complete orthonormal set of functions that essentially samples the field of view (FOV) of the receiver. This concept is illustrated in Fig. 6-42, where a cluster of diffraction-limited FOVs (denoted Ω_{dl}) is shown collecting light energy from both signal and background radiation with slightly different angles of arrival. If it can be shown that each element of the set of spatial modes satisfies an integral equation whose kernel is the field coherence function over the aperture, then each sample function of the field can be expanded into an orthonormal series with uncorrelated coefficients, also known as a Karhounen-Loeve expansion [64].

When applying this model to the array receiver, it is more appropriate to view the received turbulence-degraded optical fields as "frozen" over time scales of milliseconds, the characteristic time scale of turbulence, during which time a great many optical data symbols are received. It is advantageous to view

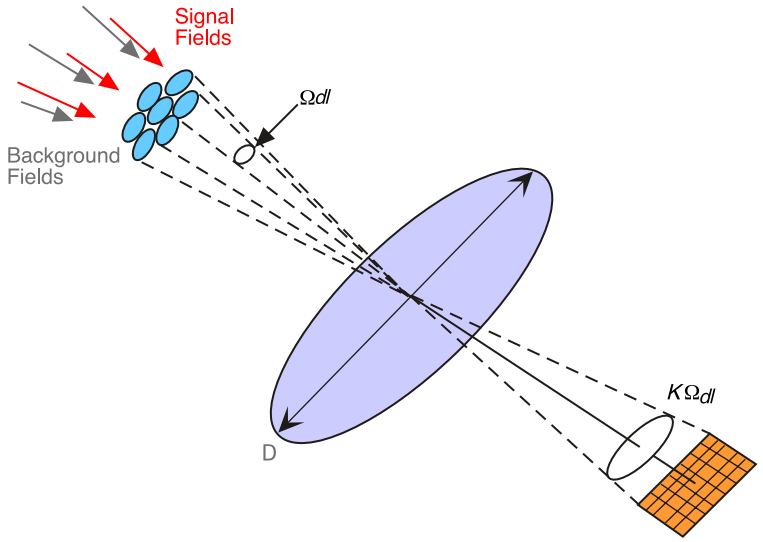


Fig. 6-42. Plane-wave expansion of aperture-plane fields: diffraction-limited FOV interpretation.

each sample function of the received field as a frozen realization of a two-dimensional field, representing a certain epoch in time over which the stationary model applies, as a deterministic two-dimensional function defined over the aperture plane of the receiver. This enables the application of well-known sampling theory to the fields, resulting in a model that clarifies the array receiver concept and contributes to the development of a useful engineering model.

Plane-wave expansions are especially useful for characterizing the effects of background radiation. In Fig. 6-42, both signal and background power are shown entering the receiver through each diffraction-limited FOV. If the diameter of the receiver is smaller than the coherence length of the signal field (a condition that can be met in the vacuum of space or even on the ground with very small apertures), then all of the signal energy is concentrated into a diffraction-limited point-spread function (PSF) in the focal plane. In effect, the signal appears to originate from a point source an infinite distance from the receiver. For the case of receiving spatially coherent signal fields, the signal power collected by a diffraction-limited telescope under ideal conditions is therefore proportional to the collecting area. The model for background radiation is somewhat more complicated, since background radiation is an extended source with respect to the narrow FOV of typical optical receivers. Surprisingly, the background power collected by a diffraction-limited telescope does not depend on the collecting area, but instead is a constant that depends only on the brightness of the source and the bandwidth of the optical filter at the wavelength of interest, λ. This can be shown by examining the units of the spectral radiance function, $N(\lambda)$, used to quantify background radiation: the units of the spectral radiance functions are power (microwatts) per receiver area (square meters), FOV (steradians), and optical filter bandwidth (angstroms or nanometers). The background power collected by a diffraction-limited receiver is given by [65]

$$P_b = N(\lambda)\Delta\lambda\Omega_{dl}A_r$$

where $\Delta\lambda$ is the optical filter bandwidth and A_r is the collecting area. Since the diffraction-limited FOV is inversely related to area, i.e., $\Omega_{dl} \approx \lambda^2 / A_r$, we have

$$P_b^* = \frac{\pi^2}{16} N(\lambda)\Delta\lambda\lambda^2$$

and therefore, the background power collected by a diffraction-limited receiver, P_b^*, is independent of receiver area. Hence, all diffraction-limited receivers of

whatever size collect the same background power when observing a given background distribution, independent of receiver collecting area.

The distorted optical fields resulting from atmospheric turbulence are necessarily band limited, since the "image" of a point source observed through turbulence is of finite extent. This conclusion is based on the observation that any complex field distribution over the aperture plane is the Fourier transform of the complex "image" function generated in the focal plane [66,67], which therefore represents the wavenumber space of the aperture. In order to determine the extent of significant wavenumbers of representative optical field distributions, sample functions generated from Kolmogorov phase screens were analyzed. For the case of 10-cm atmospheric coherence length, or $r_0 = 0.1\,\mathrm{m}$, Fig. 6-43 shows representative examples of the resulting aperture-plane coherence function, and Fig. 6-44 shows its Fourier transform, representing the two-dimensional power spectrum of the field in wavenumber space.

If we can show that the received fields over the aperture are wavenumber-limited two-dimensional functions, then we can invoke a two-dimensional version of the sampling theorem, as described in [67]. This enables the expansion of the received fields using two-dimensional sampling functions, analogous to the time-domain expansion of band-limited time sequences. The canonical sampling functions described in [67] for two-dimensional functions are closely related to the "sinc" function familiar from one-dimensional sampling theory, but functionally depend on the first-order Bessel function

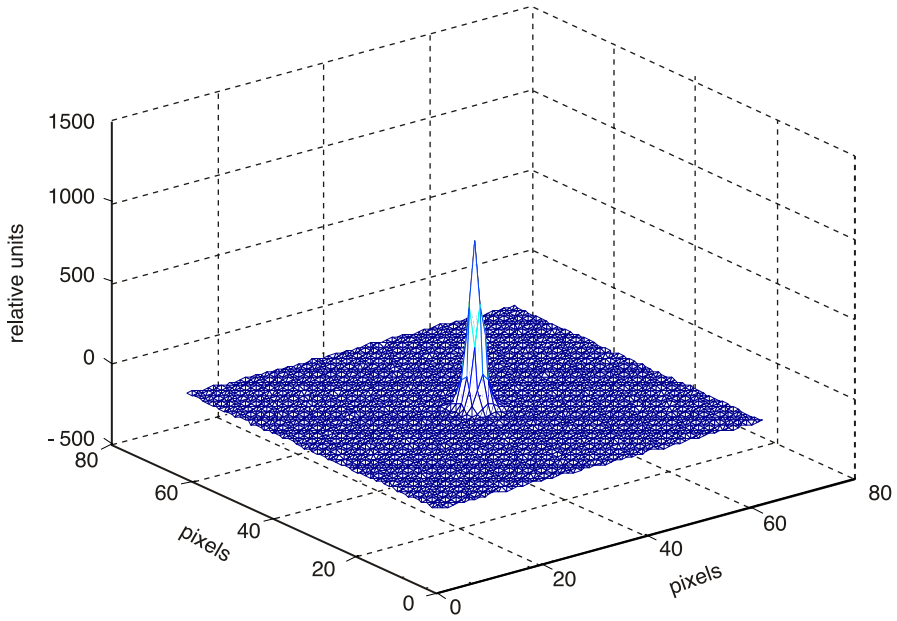


Fig. 6-43. Two-dimensional sample autocorrelation function over the aperture plane: r_0 = 10 cm.

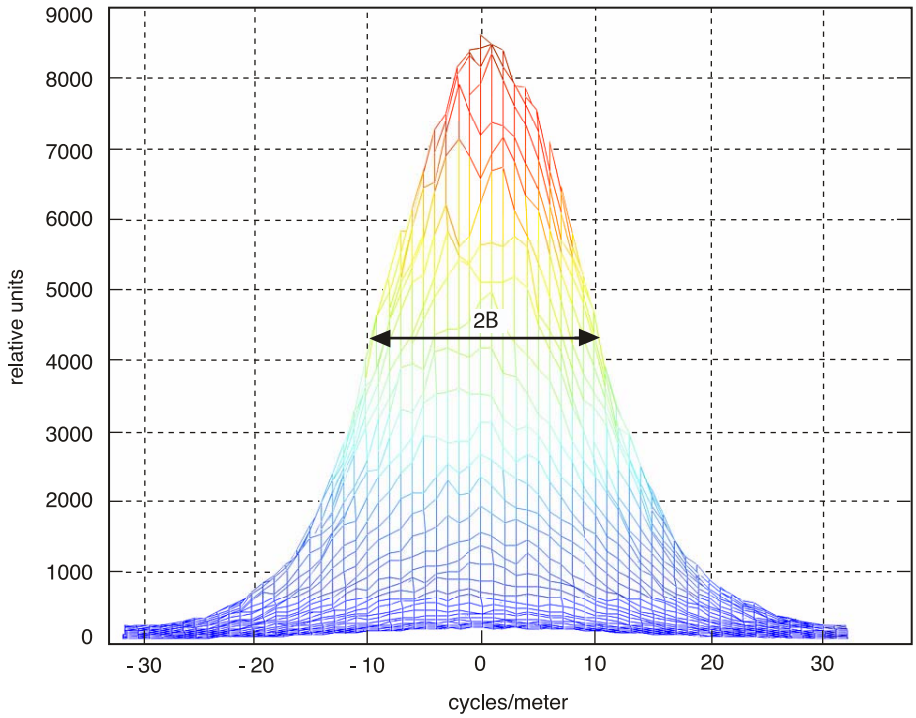


Fig. 6-44. Power spectrum of the sample field in wave-number space: r_0 = 10 cm.

$J_1(r)$, where $r = \sqrt{x^2 + y^2}$ is the radial variable along any direction in the aperture plane. If the two-dimensional optical field is limited in wavenumber space to $B = \omega/2\pi$ radians, where ω represents the angle of arrival measured from the normal to the aperture plane, then the appropriate sampling functions are of the form

$$\varphi(x,y) = C\frac{J_1\left(2\pi B\sqrt{x^2+y^2}\right)}{2\pi B\sqrt{x^2+y^2}} \tag{6.1-10}$$

where the normalization C is chosen to ensure that $\iint dxdy\,\varphi^2(x,y) = 1$.

We observe from Fig. 6-44 that the effective two-sided bandwidth, 2B, of the field is about 20 cycles per meter in wavenumber space at the 3-dB point, implying that the bandwidth is approximately given by $B = 10$ cycles per meter; this implies a sample-function spacing of 0.1 m, or 10 cm, consistent with the assumed coherence length of the field.

The two-dimensional expansion functions defined in Eq. (6.1-10) can be placed on a regular grid of points corresponding to the first zero-crossing

distance of the Bessel function, forming cells of equilateral triangles as shown in Fig. 6-45. The first three functions so placed are orthogonal, but subsequent functions at various distances from a given grid point are not exactly orthogonal; however, no two functions have greater than a few percent overlap, as shown in [68]. In principle, the set of sampling functions needed to represent any realization of a wavenumber-limited optical field over the aperture can be orthogonalized using the Gram-Schmidt procedure [67], resulting in a complete orthonormal set. This complete set of sampling functions then serves as the basis for expanding arbitrary received optical fields over the aperture, with the interpretation that the coefficients are complex samples of the field at the appropriate points in the sampling lattice.

The sampling expansion illustrated in Fig. 6-45 enables the representation of the instantaneous, or frozen, received optical field, $f(x,y)$, in terms of complex samples $\alpha_{i,j}$ over a grid of points defined by the vertices of equilateral triangles, whose separation is determined by the turbulence-induced atmospheric coherence length over the aperture plane:

$$f(x,y) = \sum_{(ij)\in A_{rec}} \alpha_{ij}\varphi_{ij}(x,y) = \sum_{(ij)\in A_1} \alpha_{ij}\varphi_{ij}(x,y) + \cdots + \sum_{(ij)\in A_K} \alpha_{ij}\varphi_{ij}(x,y) \quad (6.1\text{-}11)$$

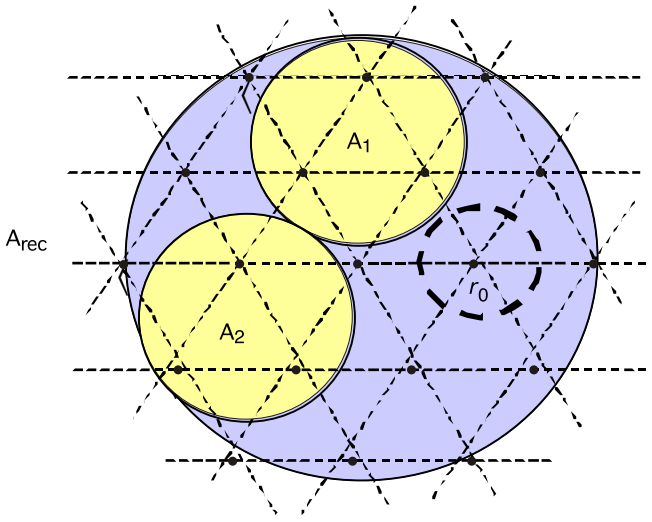


Fig. 6-45. Sampling grid over the receiver aperture defining placement of sampling functions and illustrating the interpretation of the field as samples over the entire aperture and over two subapertures.

Here,

$$\alpha_{ij} = \iint_{A_{rec}} f(x,y)\,\varphi_{ij}(x,y)\,dxdy \tag{6.1-12}$$

and the sampling functions are given by

$$\varphi_{ij}(x,y) = C\frac{J_1\left(2\pi B\sqrt{\left(x-x_i\right)^2+\left(y-y_j\right)^2}\right)}{2\pi B\sqrt{\left(x-x_i\right)^2+\left(y-y_j\right)^2}}$$

where x_i and y_j are the sampling point coordinates. These sampling functions are approximately orthonormal, i.e.,

$$\iint_{A_{rec}} \varphi_{ij}(x,y)dxdy \approx 1$$

and

$$\iint_{A_{rec}} \varphi_{ij}(x,y)\varphi_{kl}(x,y)dxdy \approx 0, \quad i \neq k, j \neq l$$

As shown in Fig. 6-45, the coefficients α_{ij} represent samples of the field over the (*i, j*)-th grid point. Due to the circular symmetry and unimodal nature of the two-dimensional sampling functions defined in Eq. (6.1-10), these samples may be interpreted as effective coherent areas of radius r_0 centered over each grid point shown by the dashed circle in Fig. 6-45. Note that we have not taken into account boundary conditions at the edge of the aperture; nor have we formally applied the Gram-Schmidt procedure to fully orthogonalize the basis functions. Therefore, this representation is somewhat approximate, but it is sufficiently accurate to motivate the development of optical array receiver theory starting with a single large aperture.

If we take the magnitude squared of the complex received field and integrate it spatially over the entire aperture, we obtain the power of the signal field through the aperture. Applying this operation to both sides of Eq. (6.1-11) yields

$$\begin{aligned}\iint_{A_{rec}} dx\,dy\,|f(x,y)|^2 &= \iint_{A_{rec}} dx\,dy\left|\sum_{(i,j)\in A_{rec}} \alpha_{ij}\,\varphi_{ij}(x,y)\right|^2 \\ &\cong \sum_{(i,j)\in A_{rec}} |\alpha_{ij}|^2\end{aligned} \tag{6.1-13}$$

where the last expression follows from the approximate orthonormality of the sampling functions. Since the left side of Eq. (6.1-13) represents the total power of the signal, so does the right-hand side, with the interpretation that the signal power over the receiver aperture can be built up from the sum of the individual sample powers. Equivalently, the total power through the aperture can be viewed as the sum of powers collected by small effective apertures centered over each grid point. Note that if the field coherence length is equal to or greater than the diameter of the receiver aperture, then a single sampling function suffices to represent the field: in this case, there is only one coefficient in the expansion, and the total signal power through the aperture equals the squared magnitude of the sample representing the entire received optical field.

Summarizing the key features of the array model, we conclude the following:

1) The amount of background power collected by any diffraction-limited receiver aperture is a constant independent of the aperture dimensions, because the diffraction-limited FOV (steradians) is inversely related to the receiver area.

2) The amount of coherent signal power collected by a diffraction-limited receiver is directly proportional to the collecting area. For the case of free-space (not turbulent) reception, the signal is in the form of a plane wave that represents a uniform power density in units of power per unit area. For the turbulent case, the plane wave is broken up into small coherence areas over the collecting aperture, but the total average signal power collected by the receiver remains the same.

Based on the above model, the following general conclusion follows:

3) A necessary condition to ensure that total signal and noise powers collected by an array of subapertures of the same total area as a single large receiver is that each subaperture contain at least one sampling coefficient.

This conclusion follows from the interpretation of the sample coefficients as small coherence areas surrounding the center of each sampling function, with effectively constant (complex) value throughout. This model effectively replaces the continuous-field model with a discrete representation, where now the total power through the aperture is the sum of the squares of the coefficients. In effect, we have replaced integration over a continuous area with a sum of discrete terms, where each coefficient represents power flowing through a small area over the aperture; these equivalent areas are disjoint and cover the entire aperture. Subapertures containing more than one sample of the field can be constructed in a similar manner, as shown in Fig. 6-45, conserving energy since the total power flowing through a given subaperture can be estimated by counting the number of discrete signal-field samples it contains.

Whereas the signal energy per sample collected by a coherence area can be estimated directly from the coefficients, the background power cannot, since it is not coherent over the same area as the turbulence-degraded signal field; the background energy is more easily estimated from the conventional plane-wave decomposition illustrated in Fig. 6-42. However, since the total background power collected by any diffraction-limited receiver is the same, it follows that each coherence area represented by an aperture-plane sample collects P_b^* microwatts of background power. Therefore, according to this model, the total signal power collected by a subaperture containing *K* samples is proportional to $\sum_{k=1}^{K}|\alpha_k|^2$, whereas the total background power is proportional to $K P_b^*$. This conclusion is correct for any subaperture containing at least one sample, from the smallest area with diameter equal to a coherence length to the largest single-aperture receiver containing a great many coherence areas. We conclude that, as long as each subaperture contains at least one signal sample, the total amount of signal and background energy can be built up from the primitive coherence-area samples, and therefore a single large aperture can be constructed as the sum of non-overlapping subapertures with an equivalent collecting area. This model holds over any region on the ground where the turbulence parameters are constant, typically on the order of hundreds of meters or more, and hence over areas much greater than any single-aperture receiver under consideration. Therefore, the centers of the subapertures can be separated by relatively large distances, creating an array from the single large aperture receiver, without affecting the collected signal power and the collected background power. Arrays of receivers therefore are equivalent to a single-aperture receiver of the same collecting area in the sense that both collect the same total signal and background power, as asserted in conclusion (3), stated above.

However, these conclusions do not extend to the case where the diameters of the array elements become smaller than the atmospheric coherence length. The reason is that, as the coherence area is subdivided into smaller apertures while holding the total area constant, creating in effect subsample apertures, the signal power flowing through each subsample aperture remains proportional to the subsample area, thus conserving signal power, but the background noise power flowing into each subsample aperture is a constant independent of collecting area. This means that each subsample aperture collects less signal power while collecting the same background power; therefore, receiver performance suffers. Alternatively, the total signal power collected by a full-sample aperture and the equivalent array of subsample apertures remains constant as the number of array elements increases, but the total noise power increases proportionally with the number of array elements. A direct consequence of this behavior is the observation that in the absence of turbulence, so that the field coherence length is equal to or greater than the collecting area, subdividing the single aperture into subapertures results in

degraded performance when background noise is present. This, of course, is merely a restatement of the fact that coherence areas cannot be subdivided without collecting more noise power than signal power and so incurring a performance loss. We may conclude, therefore, that single-aperture receivers of any diameter can be subdivided into arrays of smaller apertures only if the number of array elements remains smaller than the number of signal samples needed to represent the signal field over the receiver aperture. The above concepts can be restated in terms of the largest single receiver aperture diameter, D, and atmospheric coherence length r_0:

4) The performance of an array of small telescopes is equivalent to that of a single-aperture receiver of the same collecting area, provided the number of array elements obeys the condition $N \le (D/r_0)^2 \equiv N^*$.

Since the only cause of performance degradation is background radiation, it follows that in the absence of background radiation the total number of array elements, N, is not constrained to be less than N^*. Therefore, in situations where the background radiation is not significant—such as might be the case at night with appropriate narrowband optical filtering—any number of elements could be used to construct an array without suffering performance degradation. However, as a practical matter, the constraint on the number of array elements is not a serious impediment to array design, as the following example illustrates.

Example. It is generally accepted that even under the best possible seeing conditions on the ground, the Fried parameter typically does not exceed 20 cm at an operating wavelength of 1 μm during the day. If a 10-m aperture is needed to communicate from deep space, then the maximum number of elements permissible for an array receiver is $N \le (D/r_0)^2 = (10/0.2)^2 = 2500$. This number is far greater than what is needed to synthesize a 10-m aperture with reasonably sized telescopes; in fact, with 1-m apertures, only 100 telescopes would be needed.

We may conclude from this example that, since turbulence is always present during ground reception, an array of telescopes can be constructed with a reasonable number of elements to synthesize a single large-aperture receiver for communications applications.

6.1.2.4 Array Receiver Performance. In the following analyses, we assume that the optical bandwidth of the receiver is great compared to its electrical bandwidth, so that a multimode assumption can be applied to both the signal and background fields. It has been shown that multimode Gaussian fields with suitably small average modal noise count generate approximately Poisson-distributed random-point processes at the output of an ideal photon-counting detector [69]. This model is reasonable for communications systems operating

even at gigabit-per-second rates, and it justifies the use of the relatively simple Poisson model which, in turn, often leads to mathematically tractable solutions.

6.1.2.4.1 Array Detector Model. Consider an array of detectors consisting of $K \times L$ detector elements, representing K detector elements per telescope (FPA) and L telescopes. Assuming a frozen-atmosphere model, the sample-function density of the array of count observables from a particular focal-plane detector element of a given telescope can be written as $p[N_{mn}(t) \mid \lambda_{mn}(t); 0 \le t < T]$, where $\lambda_{mn}(t)$ and $N_{mn}(t)$ represent the Poisson count intensity and number of counts, respectively, over the mn-th detector element, and where m indexes a detector element in the focal plane of the n-th telescope [69]. This represents the output of a particular element of the array. Note that if the spatial intensity distribution is known, and the location and size of each detector element are also known, then conditioning on the spatial intensity distribution is equivalent to conditioning on the array of intensity components, each of which is still a function of time. Assuming that each detector element observes the sum of a signal field plus multimode Gaussian noise field, the array outputs can be modeled as conditionally independent Poisson processes, conditioned on the average signal intensity over each detector element. The joint conditional sample-function density over the entire array can be expressed in terms of the KL dimensional vector $\mathbf{N}(t)$ as

$$p\left[\mathbf{N}(t) \mid \lambda(t); 0 \le t \le T\right] = \prod_{m=1}^{K} \prod_{n=1}^{L} p\left[N_{mn}(t) \mid \lambda_{mn}(t); 0 \le t < T\right] \qquad (6.1\text{-}14)$$

where $\mathbf{N}(t) \equiv (N_{11}(t), N_{12}(t), \ldots, N_{KL}(t))$. This detection model can be used as a starting point for problems involving hypothesis testing and parameter estimation where the desired information is contained in the intensity distribution, but only the array of count accumulator functions can be observed.

6.1.2.4.2 Hypothesis Testing with Poisson Processes. Consider M-ary pulse-position modulation (PPM) in which one of M intensity functions is received, and the receiver attempts to determine the correct symbol based on observations of the array of count accumulator functions over each of M time slots. It is assumed that the symbol boundaries are known and that the arrival time of each detected photon and the total number of detected photons can be stored for a limited duration of time necessary for processing.

With M-ary PPM modulation, a signal pulse of duration τ seconds is transmitted in one of M consecutive time slots, resulting in a PPM symbol of duration $T = \tau M$ seconds. As shown in [69], the log-likelihood function can be expressed as

$$\Lambda_i(T) = \sum_{m=1}^{K}\sum_{n=1}^{L}\left\{\sum_{w_{j,mn}\in((i-1)\tau,i\tau]} \ln\left(1+\frac{\lambda_{s,mn}\left(w_{j,mn}\right)}{\lambda_b}\right)\right\}$$
$$= \sum_{m=1}^{K}\sum_{n=1}^{L} \ln\left(1+\frac{\lambda_{s,mn}}{\lambda_b}\right) N_{mn}^{(i)} \qquad (6.1\text{-}15)$$

where $w_{j,mn}$ is the occurrence time of the *j*-th photon over the *mn*-th detector element within the *i*-th time slot, $N_{mn}^{(i)}$ is defined as the total number of photons occurring over the *m*-th detector element in the focal plane of the *n*-th telescope during the *i*-th time slot, $\lambda_{s,mn}$ is the count intensity due to signal over the *mn*-th detector element, and λ_b is the count intensity due to background noise. Note that with constant signal intensities, the actual arrival times of photons within each slot do not contribute to the decision; hence, only the total number of detected photons, $N_{mn,}^{(i)}$ matters. Given that we know the intensity over each detector element, the *i*-th log-likelihood function consists of the sum of a logarithmic function of the ratio of signal and background intensities from all detector elements over the *i*-th pulse interval, multiplied by the total number of detected photons; the optimum detection strategy is to select the symbol corresponding to the greatest log-likelihood function.

6.1.2.4.3 Performance of the Optimum Detector-Array Receiver. The probability of a correct decision is the probability that the log-likelihood function associated with the transmitted symbol exceeds all other log-likelihood functions. Thus, when the *q*-th symbol is sent, then a correct decision is made if $\Lambda_q(T) > \Lambda_i(T)$ *for all* $i \neq q$. Denoting the logarithmic functions, or weights, in Eq. (6.1-15) by u_{mn}, the log-likelihood function can be rewritten as

$$\Lambda_i(T) = \sum_{m=1}^{K}\sum_{n=1}^{L} u_{mn}\, N_{mn}^{(i)} \qquad (6.1\text{-}16)$$

In this form, we can see that the log-likelihood function is composed of sums of a random number of weights from each detector element; for example, the *m*-th detector element in the *n*-th telescope contributes an integer number of its own weight to the sum. Note that detectors containing much more background than signal intensity do not contribute significantly to the error probability, since the outputs of these detector elements are multiplied by weights that are close to zero. This observation suggests the following suboptimum decoder concept with greatly simplified structure: list the detector elements from all telescopes

simultaneously, starting with the one containing the most signal energy, followed by every other detector ordered according to decreasing signal intensities. The logarithmic weights are partitioned into two classes: large weights are assigned the value one, while small weights are assigned the value zero. It was shown previously that this simple partitioning achieves near-optimum performance in low-to-moderate-background environments, but with greatly reduced decoder complexity.

The processing to determine which detector elements to use from the array to achieve best performance can be explained as follows: compute the probability of error for the first detector element plus background, then form the sum of signal energies from the first two detector elements (plus background for two detector elements), and so on, until the minimum error probability is reached. The set of detector elements over all telescopes that minimizes the probability of error for the entire array is selected, as this set achieves best performance. However, this straightforward process of performing the optimization is not practical for an array of telescopes, since the output of each detector element must be sent to a central processing assembly, where the computations are performed. While this is conceptually straightforward, the complexity required to achieve this processing with a large number of wideband channels quickly becomes prohibitive.

For the adaptively synthesized array detector, the probability of correct decision given hypothesis q, or H_q, can be obtained by assuming constant signal and background intensities over each time slot, yielding the conditional Poisson densities

$$p_q\left(k \mid H_q\right) = \frac{\left(\lambda_s \tau + \lambda_b \tau\right)^k}{k!} e^{-\left(\lambda_s \tau + \lambda_b \tau\right)}$$

$$p_i\left(k \mid H_q\right) = \frac{\left(\lambda_b \tau\right)^k}{k!} e^{-\lambda_b \tau}$$

Here we interpret $\lambda_s \tau$ as the total average signal count per PPM slot over the selected set of detector elements when a signal is present, and $\lambda_b \tau$ as the total average background count over the same set of detector elements, per PPM slot. As shown in [69], the probability of correct symbol decision is given by

$$P_M(C) = \left\{ \sum_{r=0}^{M-1} \left(\frac{1}{r+1} \right) \binom{M-1}{r} \sum_{k=1}^{\infty} \frac{\left(\lambda_s \tau + \lambda_b \tau\right)^k}{k!} e^{-\left(\lambda_s \tau + \lambda_b \tau\right)} \left[\frac{(\lambda_b \tau)^k}{k!} e^{-\lambda_b \tau} \right]^r \right.$$
$$\left. \times \left[\sum_{j=0}^{k-1} \frac{\left(\lambda_b \tau\right)^j}{j!} e^{-\lambda_b \tau} \right]^{M-1-r} \right\} + M^{-1} e^{-\left(\lambda_s + M\lambda_b\right)\tau} \tag{6.1-17}$$

and the probability of symbol error follows as $P_M(E) = 1 - P_M(C)$.

Performance of the array receiver was evaluated via simulated phase disturbances over each telescope of the array, using Kolmogorov phase screens as described in [70]. A sample field was generated using the distorted phase distribution, resulting in a matrix of complex signal amplitudes over each aperture. For these simulations, an atmospheric correlation length of $r_0 = 10$ cm was assumed. The field intensity generated in the detector plane of each telescope was integrated over the elements of a 128- × 128-pixel detector array, which is assumed to encompass the extent of the signal distribution in the focal plane of each telescope.

The performance of an entire sequence of arrays, starting with one large element representing the single-element receiver implementation and subdividing it into 4, 16, and 64 smaller elements of constant total collecting area, was evaluated in the following manner. Error probability was computed for both ideal turbulence-free conditions and for turbulence characterized by a Fried parameter (or coherence length) of $r_0 = 10$ cm and an outer scale of turbulence corresponding to 64 m. In a typical computation, the 128×128 = 16,384 pixel energies representing the output of a single 4-m aperture were sorted in decreasing order of average signal energy, and *M*-ary PPM symbol-error probabilities were calculated for increasing numbers of detectors, starting with the first detector. Let K_s denote the average number of signal photons per PPM slot collected by the entire array aperture when a signal is present, and let k_b denote the average number of background photons per PPM slot collected by a diffraction-limited detector element (generally made up from several contiguous pixels). We also define the total average number of background photons collected by the entire array aperture as $K_b = (D/r_0)^2 k_b$ where D is the diameter of the large single-aperture receiver. However, we should keep in mind that the adaptively synthesized focal-plane detector array typically rejects most of the background radiation, collecting background photons selectively from those detector elements that contain significant signal energy. Figure 6-46 shows the symbol-error probability for 16-dimensional PPM ($M = 16$) as a function of the number of detector elements used, for the case $K_s = 10$ photons per slot and $k_b = 0.01$ photon per slot per diffraction-limited FOV or,

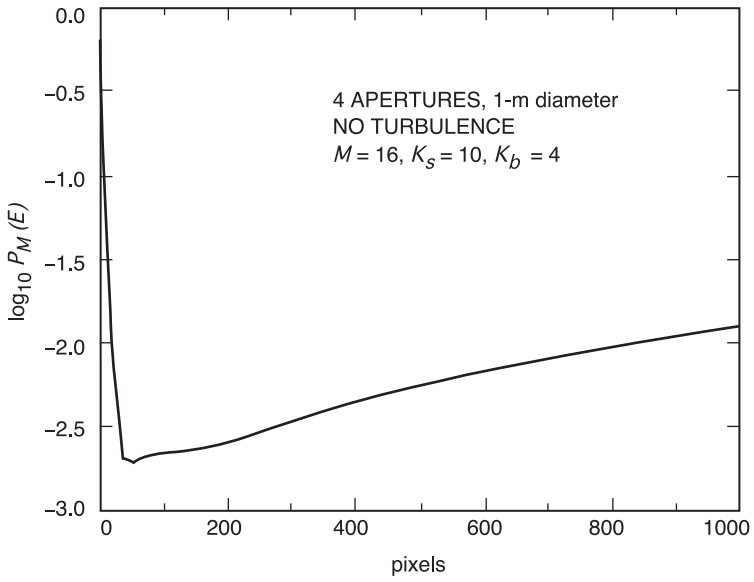


Fig. 6-46. Probability of error as a function of the number of sorted pixels: no turbulence.

equivalently, $K_b = 4$. It can be seen that for this case the smallest error probability of $10^{-2.7} \approx 0.002$ is achieved by assigning unity weight to the first 50 pixels containing the greatest signal intensities and zero to all the rest.

The same approach was used to determine array receiver performance in the presence turbulence, as shown in Fig. 6-47. However, in this case the minimum-error probability is obtained with 1200 pixels sorted according to signal energy, instead of 50 as for the diffraction-limited case. In the presence of noise, this means that performance suffers, because 1200 pixels collect approximately 24 times as much background noise as the 50 sorted pixels that attained the minimum-error probability in the absence of turbulence.

The performance of the optical arrays was evaluated using the procedure described above to determine the error probability of the adaptive "1-0" receiver. Arrays of various sizes, consisting of increasing numbers of telescopes with proportionally smaller apertures to keep the total collecting area constant, were evaluated. Independent samples of Kolmogorov phase screens were generated for each telescope, and the signal intensity distributions in the focal plane were determined for each sample function, using two-dimensional Fourier transforms. The focal plane of each telescope was assumed to contain a focal-plane array, of dimensions consistent with the telescope diameter and the expected level of turbulence.

In a typical simulation, a large telescope is analyzed first, by determining its performance according to the sorting procedure described above. The

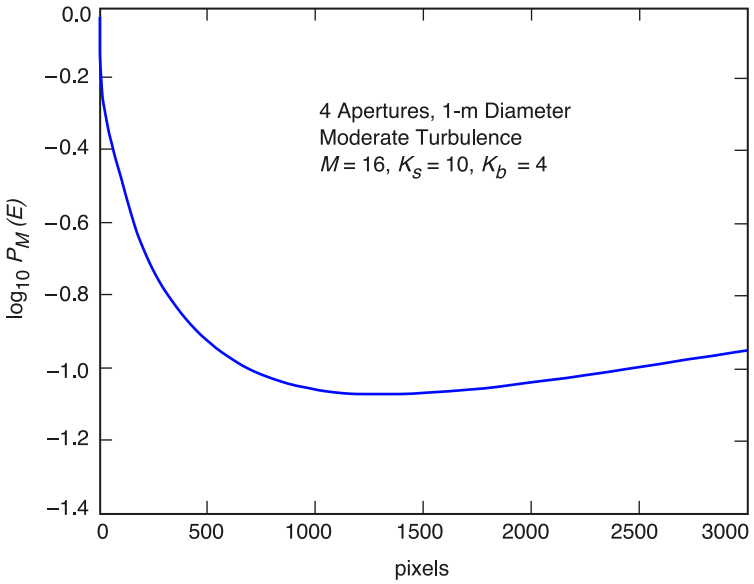


Fig. 6-47. Probability of error as a function of the number of sorted pixels: moderate turbulence, r_0 = 10 cm.

probability of error is calculated for increasing amounts of signal energy passing through the aperture, and distributed in the focal plane according to the two-dimensional Fourier transform of the aperture-field distribution generated using the Kolmogorov phase-screen program. Next, the diameter of each telescope was divided by two, generating four smaller telescopes with the same total area as the previous array, and the performance of the new larger array was computed as before. The process of dividing telescope diameters by a factor of two in order to generate the next larger array was continued four times, generating arrays of 4, 16, and 64 elements from a single large telescope. Different realizations of the signal intensity distribution were generated for each element of the new array; thus, a single telescope used a single phase-screen realization over the aperture plane, whereas an array of $N \times N$ telescopes of the same area as the single telescope used a total of N^2 different phase-screen realizations. The intensity distributions were then scaled such that the total signal energy entering the single large-aperture receiver and the array were equal.

An example of the phase distributions generated for analyzing array performance is shown in Figs. 6-48(a) and 6-48(b). The variation of optical phase over each telescope, when operating in turbulence with a Fried parameter of $r_0 = 0.1$ m (or 10 cm) and an outer scale of turbulence of 64 m, is shown in Fig. 6-48(a); the corresponding focal-plane signal distribution is shown in Fig. 6-48(b). Note that, because the 1-m telescope aperture is large compared to

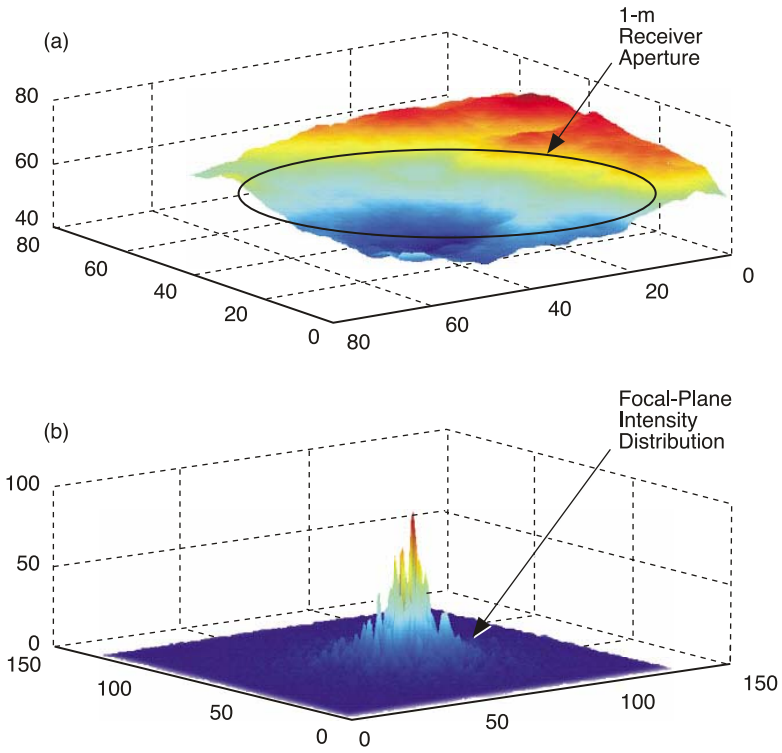


Fig. 6-48. A single 1-m-diameter telescope, 10-cm atmospheric-coherence length: (a) aperture-plane phase distribution and (b) corresponding focal-plane intensity distribution .

the 10-cm Fried parameter, there are a great many spatial modes excited by the signal field, and therefore the PSF is significantly distorted with numerous intensity peaks. This implies that a correspondingly large number of detector elements must be used to collect the signal, inadvertently collecting background radiation as well.

Aperture-plane and focal-plane distributions for two out of the four half-diameter telescopes comprising the first-level array are shown in Fig. 6-49. Note that with the smaller apertures and moderate turbulence as before, the PSF remains peaked and highly concentrated in the focal plane, suggesting that most of the signal energy is now localized in each telescope. However, due to the larger number of telescopes in the array, the same number of modes is observed as with the single large-aperture telescope, and therefore the same amount of background radiation is collected.

The process of dividing each telescope diameter to create larger arrays can be continued indefinitely in principle, yielding second-, third-, and higher-generation arrays, each containing four times as many elements as the previous configuration. However, continuing this process makes sense only as long as

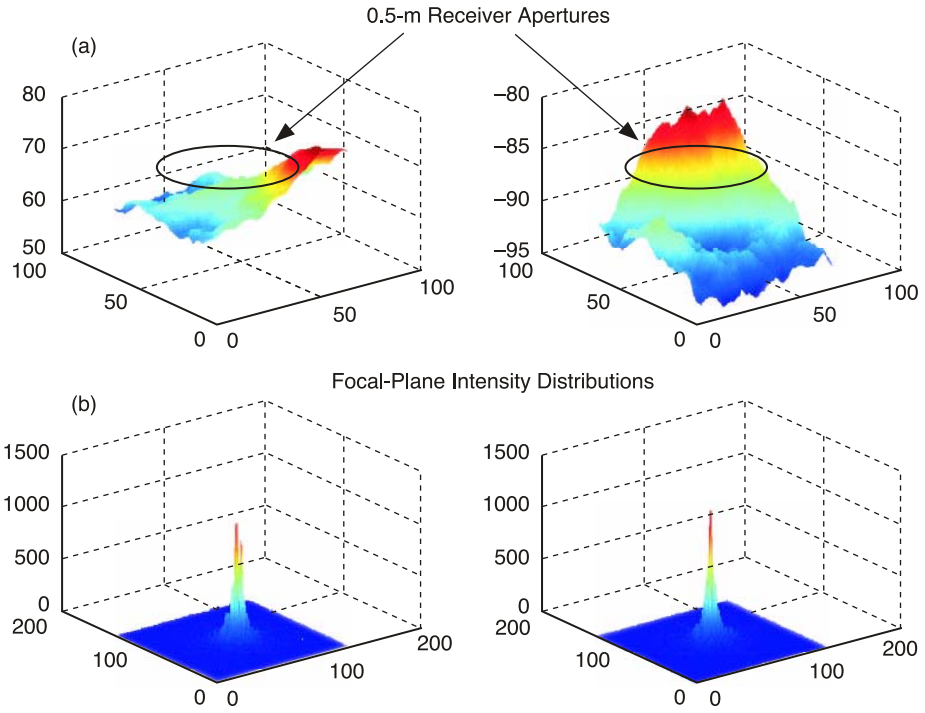


Fig. 6-49. Two telescopes of a larger array, 0.5-m diameter and 10-cm atmospheric-coherence length: (a) aperture-plane phase distributions and (b) corresponding focal-plane intensity distributions.

the telescope diameters exceed the Fried parameter; after that point, further reductions in telescope size lead to degraded performance for the reason described in 6.1.2.3. As the total number of telescopes in the array begins to exceed the total number of signal modes, it is no longer possible to distribute the signal power over more telescopes without incurring additional penalties due to the excess background radiation collected by the array. This behavior is illustrated by the performance curves in Fig. 6-50, which show the performance of a single 2-m aperture, together with arrays of 4, 16, and 64 elements of diameter 1 m, 0.5 m, and 0.25 m, respectively, under two scenarios: first, the performance of the array is determined under negligible-turbulence conditions (as might be the case for a space-borne array); and second, the performance with turbulence of 10-cm coherence length was determined for 16-PPM signals. Background radiation of 0.01 photon per diffraction-limited FOV per slot was assumed for these calculations.

Figure 6-50 shows that with little or no turbulence, such that $r_0 > D$, the single large receiver performs best. However, with significant turbulence, $r_0 << D$, the array performance is comparable to that of a single large receiver.

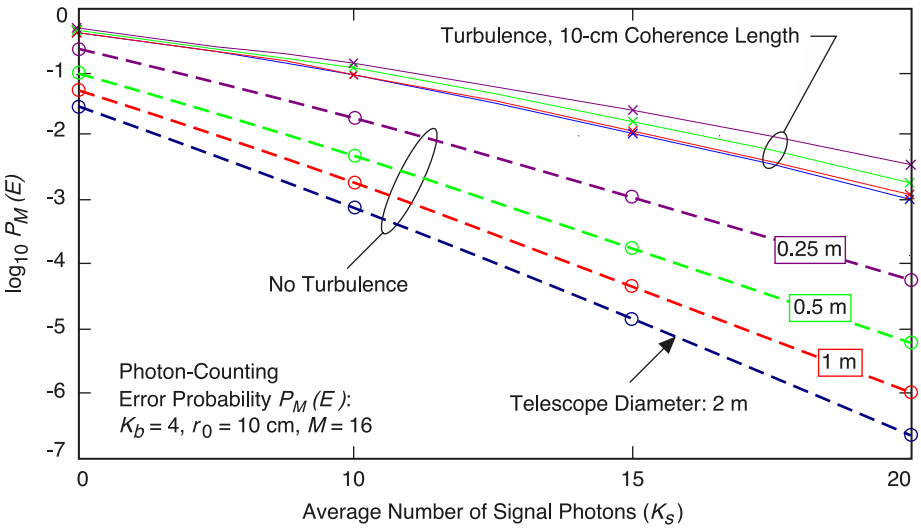


Fig. 6-50. Performance of a single 2-m telescope and an equivalent-aperture array receiver, with and without turbulence.

Note that in the absence of turbulence a single large-aperture receiver performs best because the signal is in the form of an undistorted plane wave, and therefore, a single detector element with diffraction-limited FOV suffices to collect essentially all of the signal energy. At the same time, background radiation entering the receiver from all directions within the FOV can be maximally suppressed, only the minimal amount within the small diffraction-limited cone of angles contributing to background interference within the receiver. Since diffraction-limited FOV and collecting area are inversely related, the amount of background radiation in a diffraction-limited FOV is independent of receiver aperture; this means that N diffraction-limited receivers collect N times as much background energy as a single receiver, regardless of telescope diameter. Therefore, arrays of diffraction-limited receivers collect background energy proportional to the number of elements, not proportional to the total collecting area. The performance of diffraction-limited arrays observing a signal in the absence of turbulence but with moderate background radiation in the diffraction-limited FOV of each telescope is illustrated in Fig. 6-50; note that performance degrades significantly as the number of array elements increases from a single 2-m-diameter telescope to 4 telescopes of 1-m aperture, then to 16, and finally to 64 telescopes of 0.25-m diameter.

However, turbulence is ever present; even in good seeing (corresponding to a Fried parameter of 10 cm or more), the performance of the first-generation array (of four telescopes) is virtually indistinguishable from the performance of the single aperture, as the performance curves corresponding to turbulent conditions indicate. This occurs because in the presence of turbulence the single

telescope must increase its FOV well beyond the diffraction limit to collect sufficient signal energy to minimize the probability of symbol error, and in the process collects more background energy as well. In effect, the single-aperture receiver must observe more than one (typically, a great many) spatial modes, each mode approximated by a diffraction-limited FOV collecting both signal and background energy from slightly different directions. The array of small telescopes observes the same total number of spatial modes, and hence collects the same amount of background energy by the above argument. Since the total collecting area of the array is the same as that of the single receiver, it also collects the same amount of signal energy; therefore, the performance of the array is essentially the same as that of the single-aperture receiver. Small variations in the array performance curves with turbulence are due to slight variations of the random phase distributions generated by the Kolmogorov phase-screen program.

6.1.2.4.4 Extrapolating to Determine the Performance of Large Arrays. In order to determine if there is a "best" array telescope diameter, array performance was evaluated for a given value of total signal energy, background intensity, and turbulence parameter for a large number of arrays. As originally developed, computations for array performance can be carried out in a reasonable amount of time (approximately 1 hour per run) for arrays of no more than 64 elements. This allows analysis of arrays consisting of as many as four different telescope sizes (for example, 2 m, 1 m, 0.5 m, and 0.25 m). It was found that increasing the array size beyond 64 elements, with each array containing its own focal-plane array of detectors, began to incur unacceptably great computational burdens. Therefore, an approach was developed to evaluate the performance of a large array by modifying the input noise parameters of smaller arrays, and then "connecting" the results in order to estimate performance over a much larger range of array diameters.

This technique was first applied to arrays of telescopes with diameters ranging from a single telescope of 4-m diameter to an array of 0.125-m telescopes, creating arrays of 1, 4, 16, 64, 256, and 1024 elements. These arrays were split into two sets: the first set consisted of 4-m, 2-m, 1-m, and 0.5-m telescopes, while the second set consisted of 1-m, 0.5-m, 0.25-m, and 0.125-m telescopes. The noise parameters of the second set were increased by a factor of four to account for the total noise power collected by an array with four times the number of elements, and the performances of the two sets computed using the initial program. Three different test cases were run with an average of 20 signal photons and moderate background noise of 0.01 photon per PPM slot per diffraction-limited FOV. This implies that the average number of background photons collected by the entire array aperture is $K_b = 16$, which represents an extremely high background environment. The results of the individual runs shown in Fig. 6-51 indicate that the two sets could indeed be connected,

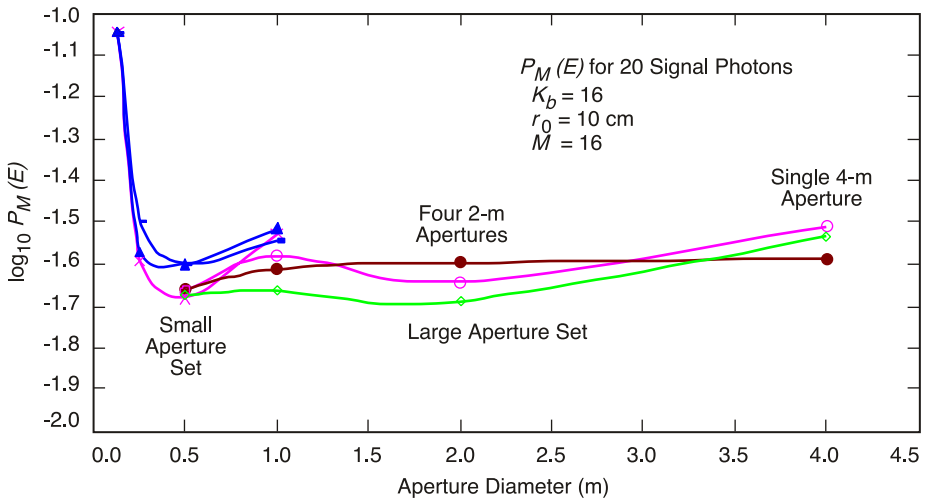


Fig. 6-51. Results of three simulation cases for evaluating the connectivity of small arrays.

resulting in greatly simplified computations. Small differences in the performance of the two sets at aperture diameters of 0.5 meter and 1 meter are attributed to variations in the sample functions of the focal-plane signal energy distributions generated by the phase-screen programs. Note that as the telescope diameters approach the Fried parameter performance deteriorates in all cases. The modal analysis presented in 6.1.2.3 predicted this performance degradation for apertures smaller than the coherence length; however, it is actually observed for somewhat larger collecting areas in Fig. 6-51. This behavior is attributed to the fact that the modal analysis was only approximate, and did not take into account edge effects that start to become significant under these conditions.

Taking the average of the performance curves of Fig. 6-51 at each telescope diameter yields the averaged performance curve of Fig. 6-52, which shows a smooth continuous curve through the entire range of diameters from 4 meters to 0.125 m, supporting the validity of the approach. Fig. 6-52 also shows the hypothetical performance of the diffraction-limited receiver, corresponding to turbulence-free conditions, but observing the same background radiation (a condition that can only be approached in free space for large apertures). Since without turbulence a single field sample suffices to represent the aperture-plane field, and hence only one diffraction-limited unit of background power is collected, the performance of the arrays is always better in the absence of turbulence.

Finally, these results were extended to cover a range of telescope diameters from 8 m to 0.125 m, again using the connected-set approach and averaging a large number of runs to obtain the final results. The results for arrays with the 8-m collecting area are shown in Fig. 6-53. These results are similar to the

previous 4-m case shown in Fig. 6-52, except that performance is uniformly worse for the following reason: with atmospheric coherence length, signal energy, and background intensity held constant for both cases, the total number of signal modes is four times as great for the 8-m equivalent aperture as for the 4-m aperture. This means the larger 8-m array collects four times as much background energy while collecting the same signal energy as the 4-m array,

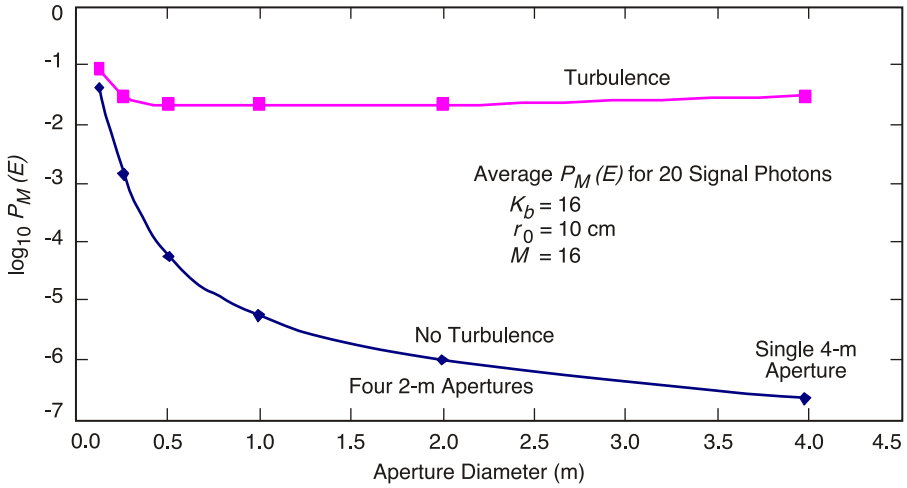


Fig. 6-52. Averaged performance of 4-m-equivalent arrays, 20 signal photons, 0.01 noise photon per slot.

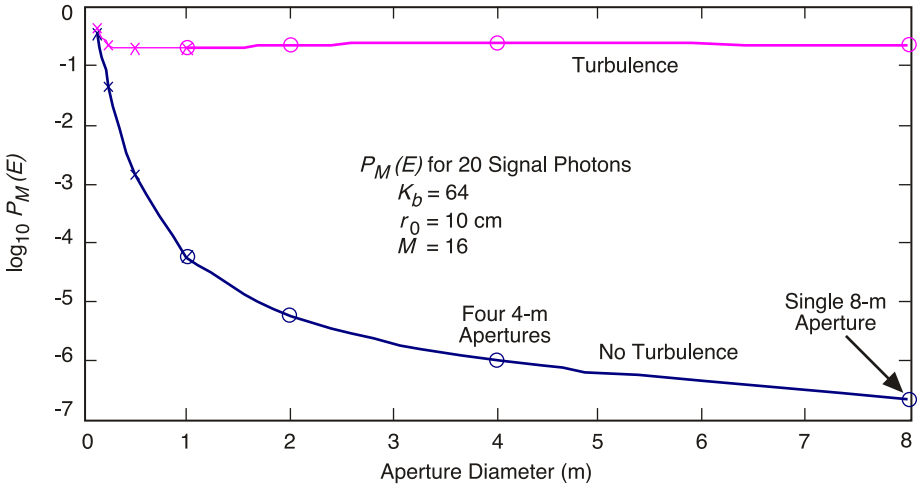


Fig. 6-53. Averaged performance of 8-m-equivalent arrays, 20 signal photons, 64 noise photons per slot. Note the degraded performance relative to the 4-m array, due to increased background noise while keeping the average signal energy the same.

and hence appears to suffer degraded performance. Similar performance would be achieved by the larger array if the signal energy were increased proportionally to the area.

In the above examples, the total number of focal-plane pixels at each telescope was not limited to any specific value a priori; in fact, every attempt was made to make sure that enough detector elements were used to collect all of the signal energy for each sample function of the phase distribution. This approach is not practical, however, since in a physical realization the number of pixels will be limited due to complexity and implementation considerations.

6.1.2.5 Conclusions. The concept of an optical array receiver suitable for deep-space communications has been defined, and the theoretical foundations of optical array receiver operation and performance have been established. The necessary mathematical models for representing aperture-plane and focal-plane fields have been developed, and detection theory has been applied to determine the performance of an optical-array receiver under realistic operating conditions. In particular, the performance of multi-element telescope arrays observing pulse-position modulated optical signals in the presence of atmospheric turbulence and background interference has been investigated, and conditions have been determined under which array performance and single-aperture receiver performance were identical. The performances of various array configurations have been determined and shown to be equivalent to the performance of the large single-aperture receiver, thus demonstrating that the array approach does not incur any losses relative to the single-aperture approach under operating conditions of interest.

Based on the above results, we conclude that the array-receiver concept represents a viable alternative to the more conventional approach of using a single large-aperture telescope to collect the optical signal energy, particularly since high-data-rate reception from the depths of space demands ever increasing collecting areas to achieve the desired performance. Large-aperture telescopes are difficult to construct and maintain, require massive support structures and drive assemblies, and pose the risk of jeopardizing the mission in case of failure. In addition, once constructed, the collecting area of a single telescope cannot be easily expanded to meet demands for increased performance. An array of small telescopes, on the other hand, provides a robust, scalable, parallel receiver architecture that can easily be expanded to meet greater demands with relative ease.

6.2 Photodetectors

6.2.1 Single-Element Detectors

Abhijit Biswas and William H. Farr

6.2.1.1 Deep-Space Detector Requirements and Challenges. This section presents discussions on detectors used for ground-based reception of optical communications transmitted from deep space. Detectors convert received signal photons into an electronic signal that is processed in order to extract information. Laser signals in the near infrared region, specifically at wavelengths close to 1.06 μm or 1.55 μm of the optical spectrum are presumed in the discussion that follows. This choice of wavelengths is dictated by the availability of high peak-to-average power laser transmitters (neodymium: yttrium aluminum garnet [Nd:YAG], neodymium: yttrium vanadium oxide [Nd:YVO_4], neodymium: yttrium lithium fluoride [Nd:YLF] crystal and ytterbium- or erbium-doped fiber amplifier lasers) required for achieving photon-efficient communications [71] over inter-planetary distances. Typically the signals incident on the detector are faint, not only because of the huge distances and associated losses, but also due to constraints on spacecraft power and laser electrical-to-optical-conversion efficiency. Therefore, single photon counting sensitivity combined with high photo-detection efficiency at wavelengths of interest, is highly sought after.

Photo-detection efficiency (PDE) is the product of the probability of absorption and the probability that the primary photoelectrons resulting from absorption are converted to a sensible photoelectron signal at the detector output. For most photon-sensitive detectors, this involves an internal gain or multiplication process. In addition to high PDE, detectors must also possess high bandwidth. This allows a faithful measurement of the time-of-arrival of faint photon pulses that are typically a few nanoseconds in duration. Finally, high performance for detectors demands low dark noise and gain variance or excess noise.

Detector technology drivers are further influenced by the environment in which they must operate. Ground based reception is currently being considered by NASA for near-future technology demonstrations [72] with likely follow-on distributed optical receiving networks deployed on the ground, and/or above the Earth's atmosphere in suborbital or orbital platforms. For receivers located above the atmosphere, coherent communications and near-diffraction limited reception expand the scope of useable detectors. However, the present discussion will be limited to detectors for ground-based receivers.

Deploying ground-based receiving systems adds to the detector challenge. Since the laser beam must propagate through the atmosphere prior to being collected and detected, all the effects described in Chapter 3 gain relevance.

Foremost among these is atmospheric "blurring" [73]. In order to cope with this effect, the detector field-of-view (FOV) must be enlarged. This in turn has the undesirable effect of "viewing" a larger angular region of the sky and proportionately increasing the collected background light scattered from the sky and other celestial sources. Daytime links that are of significant duration for deep space missions must contend with the increased background light. The added background for daytime links of course increases as the angular separation of the spacecraft from the sun decreases. Current thinking demands that optical links be operated at small (2–10 deg) Sun-Earth probe (SEP) angles. Under these circumstances, in addition to the formidable increase in sky background, stray light due to off-axis sunlight scattering into the detector FOV also becomes significant. Note that the penalty associated with background is two-fold. First, signal-to-noise ratio suffers as a consequence of the large additive component of background. Second, the large incident background photon flux that can be orders of magnitude larger than the average signal power can drive the detector into saturation. The latter condition can necessitate spreading the aggregate photon flux over multiple detector elements.

A general observation from the foregoing discussion is that the ability to handle faint signals in the presence of large background is a design requirement. This condition prevails for significant durations when receiving communications from planetary probes. For example, two-thirds of the optical link opportunities to a spacecraft orbiting Mars are in the daytime while for the distant outer planets this is reduced to half. Therefore, the limiting condition for detector selection is the ability to support viable daytime links. Significantly improved performance is expected with the 3–6 order-of-magnitude reduction when night time backgrounds are encountered.

Deep space optical links are rendered viable by reducing the duty cycle of the laser transmitter so that the peak power of the laser pulse can overcome the huge space loss, as well as associated system and channel losses, so that a few signal photons arrive at the detector. This calls for temporally narrow pulses typically 1–10 ns long. Additionally, modulation strategies like pulse position modulation (PPM) [71,72,73] offer temporal discrimination so that precise measurement of the time of arrival of the laser pulse can be used to transmit information. The channel capacity for idealized photon counting can be shown to offer the best performance in terms of bits per photon [74] that can be transmitted, compared to the traditional on-off key (OOK) modulation used by the telecom industry. Even with the penalty for additive background noise and imperfect devices that deviate from idealized photon-counting, the channel capacity for direct detection with photon counting is expected to be superior to alternate schemes. Of course in addition to photon counting sensitivity, a key to making deep space optical links viable is the availability of powerful codes [75] that can approach within 0.5–1.5 dB of capacity.

Therefore, ground-based optical detectors must provide efficient photon counting together with high bandwidth capable of precisely sensing photon arrivals. Additionally, the contributions of thermal or dark noise and gain variance or excess noise that cause device performance to deviate from ideal Poisson detection must be minimized through materials selection and good detector design.

Figure 6-54 provides an example of the data rates achievable from deep space optical links. Data rate is presented as a function of the background photon flux in photons/second incident on the detector. In this example, incident average signal photon fluxes (λ_s) of 1×10^6 and 5×10^6 photons/second were assumed. For each value of λ_s data rates were estimated for PDEs of 0.3 and 1.0. Various combinations of system and channel parameters can give rise to the operating points shown in Fig. 6-54. A PPM-order of 128 with the slot width allowed to vary in discrete steps and with a fixed code rate of 0.5 was chosen for deriving the results plotted in Fig. 6-54. In addition a 3-dB link margin and a 2-dB loss to account for the electronic and hardware implementation were assumed. The data rates flatten out with reduced background because of the constraints imposed by the combination of the assumptions made. These constraints are realistic since relaxing them in order to overcome the data rate limit will entail added complexity to the laser transmitter. Typical mass, power, and complexity limitations on a state-of-the-

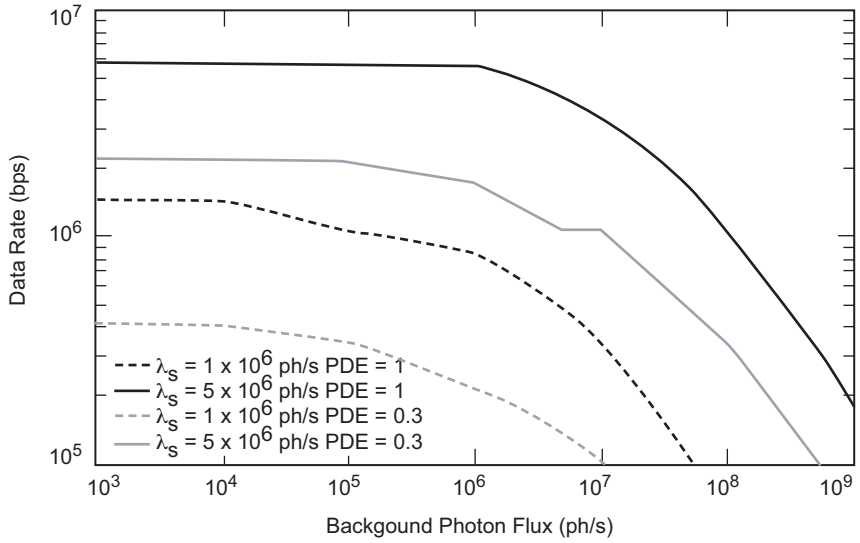


Fig. 6-54. Data rate versus average background photon flux incident on the detector shown for two mean flux levels of 1 and 5 $\times$ 10^6 photons/second and for two PDEs of 0.3 and 1. A PPM order of 128 is assumed, and data rates have a 3-dB link margin and a 2-dB implementation loss.

art spacecraft laser transmitter will result in performance similar to that depicted in Fig. 6-54. Data rates below 100 kbps are not plotted. A detailed description of how these data rates can be derived is given in references [75] and [76]. The intent here was to emphasize the influence of the detector on the link performance.

6.2.1.2 Detector System Dependencies. Critical dependencies for a detector embedded in a deep-space optical communications system are briefly pointed out in this section.

Foremost among these detector dependencies is the ability to collect sufficient signal photons to satisfy the required data rates, in the presence of additive background. Fig. 6-54 is an example of how viable operating points for a set of coding and modulation constraints such as the code rate, PPM-order and slot width can be computed. Large effective collection areas are invariably required to achieve viable operating points. Large effective collection areas can be achieved by using large-diameter single monolithic photon collectors or alternatively, through the use of an array of smaller apertures. Studies and research on both these approaches show considerable promise [77,78,79]. The specific approach chosen for implementing large effective apertures will influence the selection of detectors. Firstly, the coupling of light from relatively smaller diameter versus larger diameter apertures on to the detector surface must be accounted for. Secondly, arrays involve combining the output from a large number of detectors. Therefore, in the latter case, low-noise or noiseless photon counting is desired in order to prevent an additive dark noise contribution when combining the outputs from the elements of the array. Finally, signal combination in arrays is complicated by the requirement to maintain precise temporal alignment between the incoming pulse streams. Either the ability to precisely distribute a clock to an extended array of receivers must be met, or local clock synchronization at each array element must be implemented. Note that for the latter to occur, the signal-to-noise ratio must be 1/N that for an N-element array compared to an equivalent single aperture.

Because detectors are expected to operate in harsh background light environments, every effort is made to reject or suppress background photons. Polarization discrimination and optical filtering are both utilized. Thus, transmitting circularly polarized laser light can in principal afford a 3-dB reduction in unpolarized background that reaches the detector. Narrow-band optical filtering is a highly desirable means of suppressing background as well.

The use of powerful near capacity achieving codes has been pointed out before. In order to implement these codes, the detected signal must be temporally synchronized, and suitable high speed hardware and software are needed to accomplish this.

Early in this discussion the laser transmitter wavelength was presumed to be in the near infrared, specifically near 1060 or 1550 nm. This deserves a brief

mention since the argument can be made that laser wavelengths be chosen for compatibility with available high efficiency detectors. From a detector standpoint, this would of course be attractive; however, optical communications systems design issues do not favor this rather complex trade. Selecting shorter wavelengths entails larger background contributions since the solar flux increases towards the visible; moreover, the optical tolerances at shorter wavelengths become tighter. Atmospheric turbulence also is more severe at shorter wavelengths. On the other hand the option of going to longer wavelengths that would favor background, turbulence and optical tolerances, is limited by the lack of lasers and efficient high-bandwidth detectors.

6.2.1.3 Detectors for Deep-Space Communications. In this section a brief description of potential photon counting devices is presented. It should be stated at the outset that at the time of writing this text, photon counting devices with desired characteristics for supporting deep space optical links were not commercially available. A number of efforts are underway to develop perhaps the first generation of deep-space optical detectors for ground-based receivers.

As shown in Fig. 6-54, the data rate required for an optical link will largely dictate the PDE of the detector. Other detector parameters (such as dark and excess noise) perturb the performance and are important, but PDE will prove to be the biggest driver in developing detectors. For example, increasing dark noise will shift the operating point in Fig. 6-54 towards the right on the abscissa, while an increase in excess noise may render the 2-dB implementation loss assumed in Fig. 6-54 inadequate, or it may require novel signal processing to achieve timing synchronization and decoding.

Photon-counting detectors can be subdivided into two categories, proportional or linear photon counters and Geiger mode photon counters. Linear photon counters output voltage pulses in response to single-photon incidence. Moreover, the voltage pulse height is proportional to the mean number of incident photons. Typically in these devices, a fraction of the incident photons is absorbed depending on the quantum efficiency. The absorbed photons result in primary photoelectrons that undergo internal gain or multiplication resulting in a secondary photoelectron current flow across a resistor that is sensed as a voltage pulse. The efficiency with which primary photoelectrons are converted to a sensible output signal contributes to a multiplicative term that determines the PDE. For an ideal proportional photon counter, the output pulse height would uniquely identify the number of absorbed photons. Practical device constraints (such as thermal noise and gain variance or excess noise) smear out the measured voltage response so that a distribution of voltages results from the incidence of a given number of photons. If the output voltage variance is narrow, the number of absorbed photons can still be uniquely determined. This usually occurs when the gain variance or excess noise is sufficiently low. Proportional photon counters

operating in this mode are sometimes referred to as photon-number detectors. On the other hand, if the excess noise is large, a unique relation between the output voltage and the number of absorbed photons ceases to exist. Fig. 6-55 shows a conceptual schematic of the different modes of operation of proportional photon counters relative to an ideal photon counter.

The temporal response of proportional photon-counters is limited by the ability to resolve photon time of arrivals, also referred to as detector bandwidth. Detector bandwidth limitations can reside in the detector or be introduced by post-detection analog conditioning electronics. Photon arrival time separations exceeding the detection bandwidth causes variance in the pulse shape output by the detector. For pulse position modulation, this pulse-shape variance can give rise to inter-symbol interference effects. Examples of proportional photon counters are photomultiplier tubes and avalanche photodiodes.

Geiger mode avalanche photodiodes (GMAPDs) are operated biased close to breakdown, so that incidence of a single photon triggers an avalanche or breakdown resulting in a gigantic output pulse. In this mode, detectors do not instantaneously recover from breakdown. Active and passive quenching circuits are implemented on the backend of GMAPDs that determine the dead time and after-pulsing characteristics. After-pulsing results from trapped charge that is emitted at a delay following breakdown. Thus, GMAPDs provide infinite gain (or virtually noiseless performance) but cannot provide a photon-number response since the dead times vary from tens to thousands of nanoseconds.

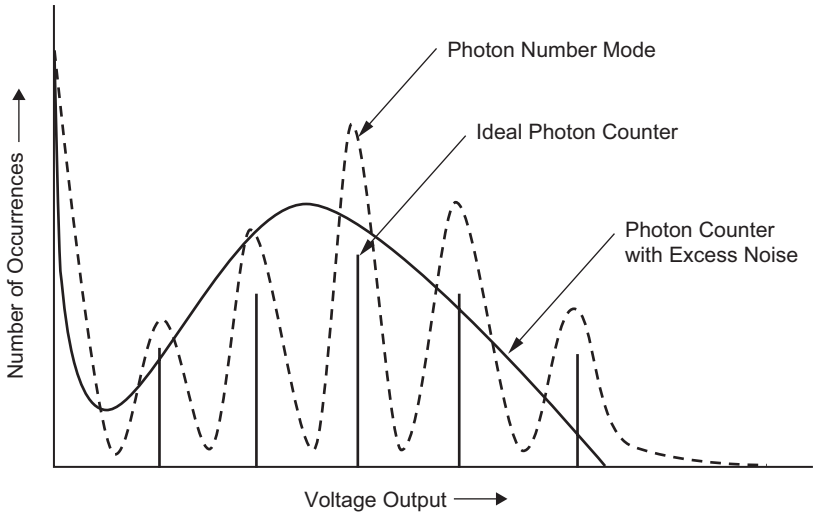


Fig. 6-55. A conceptual view comparing the histograms expected for the output of a linear mode proportional photon-counting detector for an ideal photon counter, a photon-number mode detector, and a photon counter with excess noise.

For pulse position modulation, the detector should count photon arrivals with temporal precision that is a small fraction of the slot width and generate output voltages proportional to the number of photons counted. Uncertainty in the number of counted photons and time of arrival will result in link performance degradation. A photon-number mode detector can meet these requirements if it (i) does not saturate due to the incidence of high photon fluxes, (ii) has low excess noise to count the number of photons over a slot interval accurately, (iii) has adequate bandwidth, and (iv) has a low dark noise. Provided the detector can meet these requirements, backend processing electronics in the receiver can perform the synchronization and likelihood estimation that are subsequently fed to the decoder for extracting bits of information. Given most available proportional photon counters, the fluxes typically encountered in day-time near-Sun links will tend to drive a single detector to saturation, and the use of a small array of detectors is required to handle the highest photon fluxes expected when operating close to the Sun. The array element outputs are processed in parallel and combined prior to processing. Provided two to four detector elements are used, the additive dark noise penalty is not excessive. For example, if a detector has 300 thousand counts per second (kcps) of dark counts then four of these will contribute to 1 million counts per second (Mcps), but this is still a small fraction of the 10^8–10^9 cps of background photons.

As pointed out in [74], GMAPD photon counters can be used with an extended array of receiving apertures. The photons arriving at the focal plane of each collection aperture are proportional to the diameter. A suitably sized array of GMAPDs at the focal plane of each aperture will ensure that the mean number of photon arrivals per detector dead time does not exceed 1. If this condition can be met, there will be no blockage loss. In reality, depending upon the link design, some blockage loss can be allocated.

Given the brief discussion on some very general approaches toward implementing photon counting schemes for deep-space optical links a few specific detectors will be briefly described.

6.2.1.3.1 Photomultiplier Tubes (PMT). Photomultiplier tubes (PMTs) for free-space optical communications were laboratory tested in the early 1980s at JPL [80,81]. It was shown in this early work that the photon counting channel capacity could indeed be validated through laboratory measurements. A laboratory link at 2.5 bits/photon was demonstrated using a photon-counting PMT with a quantum efficiency of 15 percent at 850 nm. Pulse widths of 100 ns were used, and both uncoded and coded (Reed-Solomon coding) bit-error rate performance matched that predicted by channel-capacity computations. Unfortunately, a big bottleneck to adapting PMTs has been the detection efficiency limitations at the wavelengths of interest.

More recently, commercial liquid-nitrogen-cooled PMTs with quantum efficiencies as high as 8 percent at 1064 nm have become available. Laboratory testing of such a device manufactured by Hammatsu Corporation, Hammatsu, Japan (Model Number R5509-42) was conducted [82] in the laboratory. The PMT was shown to behave like a proportional photon counter with the measured pulse height being proportional to incident photons per pulse, as suggested by the plot in Fig. 6-56. Figure 6-57 shows a 1064-nm laser pulse train measured in the laboratory using this PMT. The upper trace was generated by splitting a fraction of the incident laser pulse train onto a high speed photodiode in order to provide a reliable timing reference. The delay between the pulses in the upper trace and the PMT output represents the difference in light-travel time for the laser beam to reach the photodiode and PMT. A comparison of the traces in Fig. 6-57 illustrates the bandwidth limitations of the PMT. The temporal width of the high-speed photodiode output is far more representative of the laser pulse duration than the PMT bandwidth-limited pulse shapes in the lower traces. Furthermore, relatively larger fluctuations in pulse-height response of the PMT, associated with its gain variance are apparent from a comparison of the two traces. The gain variance causes the probability distribution function of the output to deviate from an ideal Poisson distribution. In this case a Polya-distribution function shows a reasonable fit to the measured data as shown in Fig. 6-58. In this plot, the parameter b is a free-fitting parameter that represents the uniformity of the photocathode and the subsequent dynode stages. Thus $b = 0.1,\ 0.25,\ 1,\ 0.01,\ 0.01,\ 0.01$ denotes a figure of merit with respect to uniformity for the photocathode and the

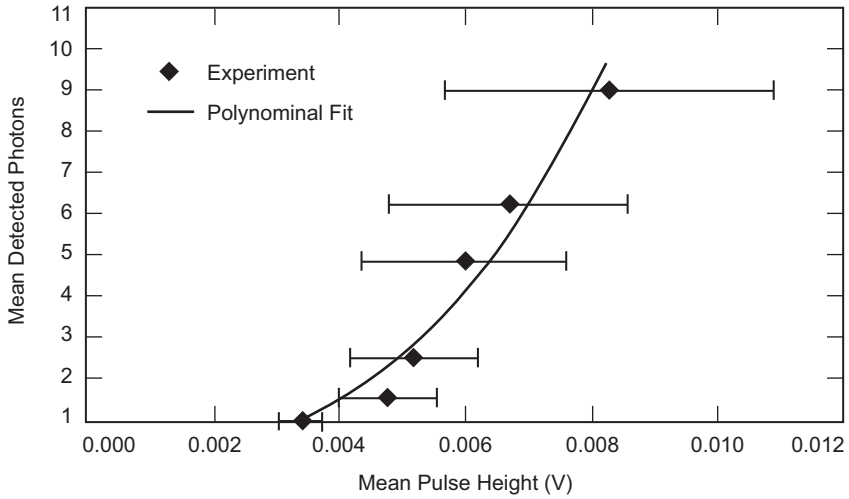


Fig. 6-56. A laboratory measured plot of output voltage versus mean number of incident photons for a photon-counting photomultiplier tube.

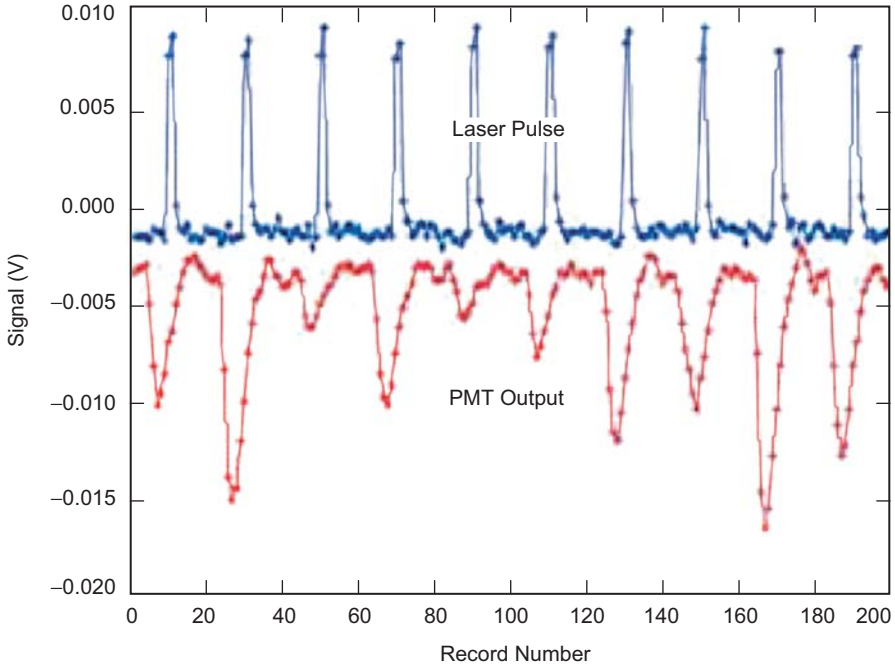


Fig. 6-57. An example of the output from a photon-counting photomultiplier tube shown in the lower trace. The upper trace is a time reference generated by splitting a fraction of the light incident on the PMT onto a high-speed photodiode. The delay between the pulses in the upper trace and lower trace is due to optical path differences.

subsequent dynode stages of the PMT. The higher numbers of 0.1 to 1 indicate a large degree of non-uniformity resulting in photoelectron emission deviating significantly from Poisson, while the smaller values for the subsequent stages is much more Poisson like.

While the PMT provides a reasonable example of a photon counter, its deficiencies in supporting a deep-space link are many. The bandwidth of the PMT would limit the minimum PPM slot width to ~ 9 ns while less than 1–2 ns are required to support the space-qualifiable lasers that are being considered. The anode current limit on these detectors is 2 microamperes (μA), equivalent to an incident photon flux of approximately 3×10^8 photons/second. In order to use such a device under severe background noise would require a small array (for example, 2×2) of devices. The largest drawback is the PDE, which is of the order of 8 percent.

Hybrid PMTs [83,84] are newly emerging photon counting detectors that hold considerable promise for providing efficient photon counting at 1060 and 1550 nm. Here a transmission-mode photocathode is used with a gallium arsenide (GaAs) Schottky avalanche photodiode that serves as the anode. These devices have bandwidths in excess of 1 GHz and excess noise as low as 1.023.

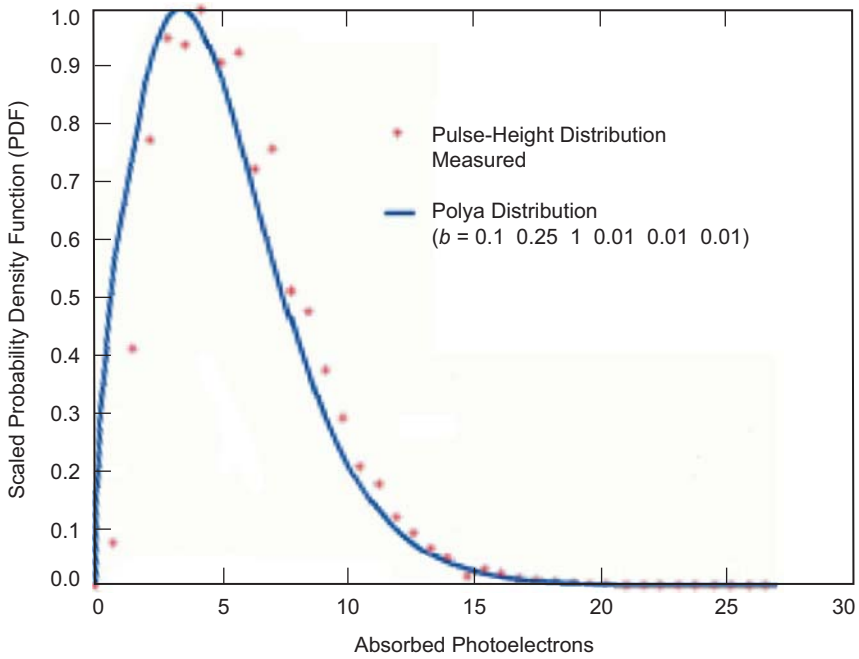


Fig. 6-58. A Polya-distribution function fitted to a measured PMT histogram. The parameter *b* is a free-fitting parameter representing a figure of merit of the uniformity of emission from the photocathode and subsequent dynode surfaces. A *b* of 0 would represent nearly perfect Poisson emission characteristics; deviation from *b* = 0 represents non-uniformity.

The internal gain occurs in two steps. The primary photoelectrons are accelerated by a high voltage (8 kV) and produce secondary photoelectrons by impact ionization with a gain of approximately 1000–1500. Subsequently, the anode provides an avalanche photodiode (APD) gain stage of about 10. These internal gains are sufficient to provide photon-counting sensitivities. The dark current can be reduced by cooling of the device. These devices are manufactured by Intevac Inc., Santa Clara California. They hold considerable promise of providing PDEs in the range of 30–40 percent at 1064 nm with the use of an InGaAsP photocathode.

6.2.1.3.2 Avalanche Photodiodes (APDs). Avalanche photodiodes are mature devices for many applications and typically operate under a bias voltage to provide high internal gain. The underlying device theory for APDs is treated in several text books [85,86] and is not treated here. Experimental work relevant to the use of APDs for deep space communications links has been reported [87]. These devices were limited by the Webb + Gaussian model and fall short of providing photon counting sensitivities.

More recently, visible light photon counters (VLPCs, manufactured by DRS Inc. of Anaheim, California) [87] have emerged as very high PDE photon-number devices that would be ideally suited for supporting deep space optical links. The only drawback of these devices is that their spectral response is very poor at 1064 nm. One drawback of these detectors is that they have to operate at cryogenic temperatures ranging from 6 to 8 K. Notwithstanding this drawback, if the spectral response could be extended into the near-infrared, they would prove attractive.

A GMAPD tailored for use with an array of smaller diameter (80 cm) apertures is underway at Lincoln Laboratories [88]. These detectors have PDEs of 45 percent and dark currents of the order of 50 kilocounts per second (kcps) at 290 K. The dead time on these detector elements is of the order of 1.6 μs. The planned implementation of these GMAPDs will utilize an array optically coupled to the collecting aperture with a lenslet array [74]. Refresh times of the order of 1 μs are considered possible for these detectors.

6.2.1.3.3 Detector Summary. In summary, it is appropriate to report that the planned upcoming deep space optical communications demonstration [72] has spurred the development of photon counting detectors [89,90] for deep space optical communications. In the next few years reports of progress on these developments should become available. Furthermore, the developments are expected to be ongoing since detectors that meet all the requirements discussed above will go a long way toward rendering optical communications links from deep space reliable and robust under the severe background contamination conditions that are expected to prevail for significant durations. These detectors will also be able to provide very high performance during the remaining duration when significantly reduced backgrounds are encountered.

6.2.2 Focal-Plane Detector Arrays for Communication Through Turbulence

Victor A. Vilnrotter and Meera Srinivasan

6.2.2.1 Introduction. Ground-based reception of optical signals from space suffers from degradation of the optical phase-front caused by atmospheric turbulence. This leads to a reduction in the effective diameter of the receiving telescope, and to random fluctuations of the receiver's "point-spread function" (PSF) in the focal plane. For example, the diffraction-limited FOV of a receiving telescope can be taken to be approximately $\theta_{dl} \cong \lambda / D_R$ which, for a 3-m aperture and 1-μm wavelength translates to 0.33 μrad. If the effective focal length of the telescope is 6 m, then a diffraction-limited PSF of 2 μm diameter, or 0.002 mm, will be produced in the focal plane. Thus, under ideal conditions, a very small detector could be used to collect virtually all of the signal energy, while at the same time spatially filtering out most of the background radiation.

However, atmospheric conditions rarely permit diffraction-limited operation of large telescopes. Even under "good" seeing conditions, the phase of the received signal field tends to become uncorrelated over distances greater than 20 cm, deteriorating to as little as 2–4 cm during the day [91]. Under these conditions, the dimensions of the PSF in the focal-plane tend to increase inversely with coherence length, as if the diameter of the collecting aperture were correspondingly reduced: the telescope still collects all of the signal energy propagating through its physical aperture, but the collected signal energy is now re-distributed onto a much larger spot in the focal-plane. An example of the increase in the effective dimensions of the receiver's PSF over its diffraction-limited value is shown in the contour plots of Fig. 6-59, obtained from a simulation using Kolmogorov phase-screens corresponding to 1-m telescope diameter and 4-cm atmospheric coherence length. For this example, the dimensions of the diffraction-limited PSF correspond roughly to a single pixel of the 16×16 detector array shown superimposed on the signal distribution.

In order to collect all of the signal energy, the dimensions of a single optical detector must be made large enough to encompass the degraded PSF as well as its random excursions in the focal-plane, which tend to change on time-

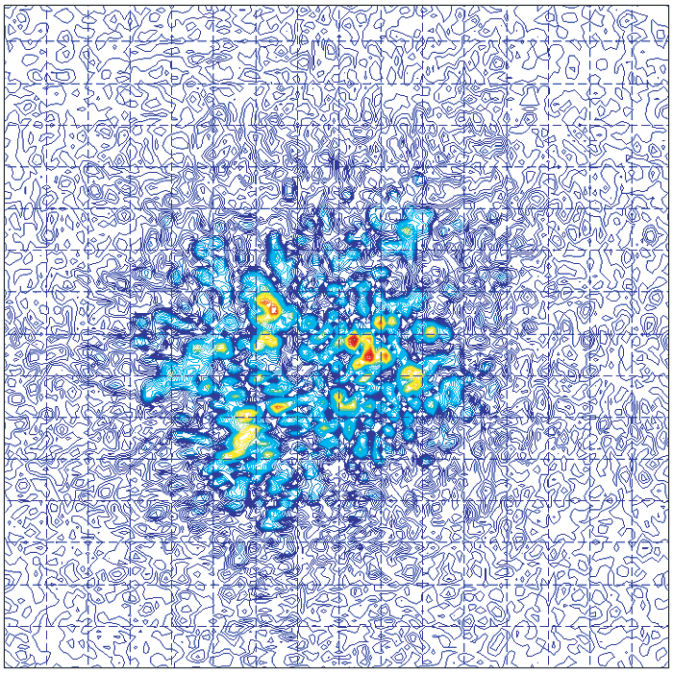


Fig. 6-59. Contour plot of instantaneous focal-plane signal distribution (covering 16×16 detector array) due to atmospheric turbulence, 4-cm coherence length, 45×45 μrad FOV.

scales of 10–100 ms. However, a large detector implies a large receiver FOV, which in turn implies a corresponding increase in the amount of background radiation admitted into the receiver. That, in turn, degrades communications performance. These problems are effectively mitigated by the use of a photon-counting detector array together with high-speed digital electronics capable of performing the signal-processing functions required for optimum or near-optimum detection.

6.2.2.2 Optical Direct Detection with Focal-Plane Arrays. In the following analyses we shall assume that a "multimode" assumption can be applied to both the signal and background fields. It has been shown that multimode Gaussian fields with suitably small average modal noise count generate approximately Poisson-distributed random point processes at the output of an ideal "photon-counting" detector [64]. This model is reasonable for communications systems operating even at mega-bit per second (Mbps) rates, and it justifies the use of the relatively simple Poisson model which, in turn, often leads to mathematically tractable solutions.

6.2.2.2.1 Single-Detector Model. Suppose that a single-detector element "measures" the number of photons contained in the received field by producing a stream of free electrons at its output terminal in response to the absorbed photons. If the occurrence-time of each pulse can be measured, and if the amplitude of each pulse is normalized to unity, then we can define a count accumulator function $N(t)$, consisting of positive integer-valued jumps occurring each time a photon is detected. As in [92], we can associate a conditional sample-function density with the detected process, conditioned on the intensity function $\lambda(t)$, defined as

$$p[N(t)\,|\,\lambda(t);0 \le t < T] = \begin{cases} \exp\left(-\int_0^T \lambda(t)\,dt\right); & N(T) = 0 \\ \left[\prod_{i=0}^{N} \lambda(w_i)\right] \exp\left(-\int_0^T \lambda(t)\,dt\right); & N(T) \ge 1 \end{cases} \tag{6.2-1}$$

where the set $\{w_i\}$ is the occurrence times of the detected photons, and $N(t)$ is the count accumulator function of the process over the time-interval [0,T). If instead of just a single detector, an array of detectors were used to detect the optical fields, then the notation must be suitably generalized to enable unambiguous description of the output of each detector element.

6.2.2.2.2 Array Detector Model. Consider a rectangular array of detectors consisting of $K \times L$ detector elements. For some applications, such as finding the "center" of the signal intensity distribution, it is important to know the

location of each detector element within the array; therefore, we include the subscripts *mn*, $1 \le m \le K$, $1 \le n \le L$, to denote the position of the detector element within the array. Thus, the sample function density defined in Eq. (6.2-1) can be written as $p[N_{mn}(t) | \lambda_{mn}(t); 0 \le t < T]$, which now represents the output of a particular element of the array. Assuming that each array element observes the sum of a signal-field plus multimode Gaussian noise-field with average noise-count per mode much less than one, the array outputs can be modeled as conditionally independent Poisson processes, conditioned on the average signal intensity over each detector element [64,93]. Hence, we denote the joint conditional sample function density of the array as

$$p\left[\mathbf{N}(t) \mid \lambda(t); 0 \le t \le T\right] = \prod_{m=1}^{K} \prod_{n=1}^{L} p\left[N_{mn}(t) \mid \lambda_{mn}(t); 0 \le t < T\right] \quad (6.2\text{-}2)$$

where $\mathbf{N}(t) \equiv (N_{11}(t), N_{12}(t), \ldots, N_{KL}(t))$, and each component on the right-hand-side is of the form defined in Eq. (6.2-1).

6.2.2.2.3 Hypothesis Testing for Poisson Processes. Consider *M*-ary pulse-position modulation (PPM), in which a signal pulse of duration τ seconds is transmitted in one of *M* time-slots, resulting in a PPM symbol of duration $T = \tau M$ seconds, after which the receiver attempts to determine the correct symbol based on observations of the array of count accumulator functions over each of the *M* time slots. It is assumed that the symbol boundaries are known and that the arrival time of each detected photon and total number of detected photons can be stored for a limited duration of time necessary for processing. Under the *i*-th hypothesis, the integrated intensity over the *mn*-th detector element is given by

$$\lambda_{mn}^{(i)}(t) = \begin{cases} \lambda_{s,mn}(t) + \lambda_b & (i-1)\tau \le t < i\tau \\ \lambda_b & \text{else} \end{cases} \quad (6.2\text{-}3)$$

where $\lambda_{s,mn}(t)$ is the received signal intensity function for the *mn*-th detector element, often assumed to be constant over the *i*-th slot duration, and λ_b is the background intensity per detector element. Suppose that each of the *M* messages is equally likely to be transmitted with probability M^{-1}, and that each message generates a unique vector of detector array intensities at the receiver, denoted by $\boldsymbol{\lambda}^{(i)}(t) = \left(\lambda_{11}^{(i)}(t), \lambda_{12}^{(i)}(t), \ldots, \lambda_{KL}^{(i)}(t)\right)$. At the end of *T* seconds, the post-detection processor selects that message corresponding to the greatest probability of having been received. Equivalently, the decoder selects the

message corresponding to the greatest "log-likelihood" function, $\Lambda_i(T)$, conditioned upon the signal occurring in the *i*-th time-slot:

$$\begin{aligned}\Lambda_i(T) &= \ln\left\{p\left[\mathbf{N}(t)\,|\,\boldsymbol{\lambda}^{(i)}(t);0\le t<T\right]\right\}\\ &= \sum_{m=1}^{K}\sum_{n=1}^{L}\ln\left\{p\left[N_{mn}(t)\,|\,\lambda_{mn}^{(i)}(t);0\le t<T\right]\right\}\\ &= \sum_{m=1}^{K}\sum_{n=1}^{L}\left(-\int_{(i-1)\tau}^{i\tau}\lambda_{mn}^{(i)}(t)\,dt+\sum_{w_{j,mn}\in((i-1)\tau,i\tau]}\ln\lambda_{mn}^{(i)}\left(w_{j,mn}\right)\right)\\ &\quad+\left(\text{terms that depend only on }\lambda_b\right)\end{aligned} \tag{6.2-4}$$

where $w_{j,mn}$ is the occurrence time of the *j*-th photon over the *mn*-th detector element within the same time-slot. By expressing the total intensity over the *i*-th time-slot as

$$\lambda_{mn}^{(i)}(t)=\lambda_b\left(1+\frac{\lambda_{s,mn}(t)}{\lambda_b}\right)=\ln(\lambda_b)+\ln\left(1+\frac{\lambda_{s,mn}(t)}{\lambda_b}\right)\quad(i-1)\tau\le t<i\tau \tag{6.2-5}$$

assuming constant signal intensity over the *i*-th time-slot, independent of the value of *i*, and ignoring terms that do not convey any information about the transmitted symbol, the log-likelihood function reduces to

$$\begin{aligned}\Lambda_i(T) &= \sum_{m=1}^{K}\sum_{n=1}^{L}\left\{\sum_{w_{j,mn}\in((i-1)\tau,i\tau]}\ln\left(1+\frac{\lambda_{s,mn}\left(w_{j,mn}\right)}{\lambda_b}\right)\right\}\\ &= \sum_{m=1}^{K}\sum_{n=1}^{L}\ln\left(1+\frac{\lambda_{s,mn}}{\lambda_b}\right)N_{mn}^{(i)}\end{aligned} \tag{6.2-6}$$

where $N_{mn}^{(i)}$ is defined as the total number of photons occurring over the *mn*-th detector element during the *i*-th time-slot. Note that with constant signal intensities the actual arrival-times of photons within each slot do not contribute to the decision; hence, only the total number of detected photons, $N_{mn}^{(i)}$, matters. The optimum detection strategy is to select the symbol corresponding to the greatest log-likelihood function.

6.2.2.2.4 Performance of the Optimum Detector-Array Receiver. The probability of a correct decision is just the probability that the log-likelihood function associated with the transmitted symbol exceeds all other log-likelihood functions, i.e., when the q-th symbol is sent, a correct decision is made if $\Lambda_q(T) > \Lambda_i(T)$ *for all* $i \neq q$. Denoting the logarithmic functions, or "weights," in Eq. (6.2-5) by u_{mn}, the log-likelihood function can be rewritten as

$$\Lambda_i(T) = \sum_{m=1}^{K} \sum_{n=1}^{L} u_{mn}\, N_{mn}^{(i)} \tag{6.2-7}$$

In this form, we can see that the log likelihood function is composed of sums of a random number of weights from each detector element. The probability density of the log likelihood function is the convolution of the probability densities from each detector element, and therefore all possible combinations of sums are represented. This means that the probability masses in the convolved density are defined over all possible sums of weights from every detector element, and the probability over each point is the product of the individual probabilities contributing to that point.

Define the set of ordered numbers over which the probability density of the i-th log likelihood function takes on values as $\{\alpha_0 = 0, \alpha_1, \alpha_2, \ldots\}$ and let $\Pr[\Lambda_i(T) = \alpha_k \mid H_q] \equiv p_i(\alpha_k \mid H_q)$. The received symbol is decoded correctly if the sum of weights from all detector elements over the signal-slot exceed the sum of weights from every other (non-signal) slot. Taking all cases into account as in [64], and assuming equiprobable signals, the probability of correctly decoding the received symbol is given by

$$\begin{aligned} P_M(C) = P_M(C \mid H_q) = \Bigg\{ & \sum_{r=0}^{M-1} \left(\frac{1}{r+1}\right) \binom{M-1}{r} \sum_{k=1}^{\infty} p_q\left(\alpha_k \mid H_q\right) \Big[\underset{i \neq q}{p_i}\left(\alpha_k \mid H_q\right) \Big]^r \\ & \times \Bigg[\sum_{j=0}^{k-1} \underset{i \neq q}{p_i}\left(\alpha_j \mid H_q\right) \Bigg]^{M-1-r} \Bigg\} \\ & + M^{-1} \Big\{ p_q\left(\alpha_0 \mid H_q\right) \Big[p_i\left(\alpha_0 \mid H_q\right) \Big]^{M-1} \Big\} \end{aligned} \tag{6.2-8}$$

where $p_q(-)$ and $p_i(-), i \neq q$, refer to the probability densities corresponding to the signal and null hypotheses, respectively. The probability of a symbol error

is simply $P_M(E) = 1 - P_M(C)$. Note that by counting all ties as errors in Eq. (6.2-8) a lower bound on the probability of correct detection, $P_M^l(C)$, is obtained that is much easier to compute, namely

$$P_M(C) \geq P_M^l(C) \equiv \sum_{k=1}^{\infty} p_q(\alpha_k \mid H_q) \left[\sum_{\substack{j=0 \\ i \neq q}}^{k-1} p_i(\alpha_j \mid H_q) \right]^{M-1} + M^{-1} \left\{ p_q(\alpha_0 \mid H_q) \left[p_i(\alpha_0 \mid H_q) \right]^{M-1} \right\} \tag{6.2-9}$$

leading to the following upper bound on the error probability:

$$P_M^u(E) \equiv 1 - P_M^l(C) \geq P_M(E)$$

6.2.2.2.5 The "Adaptive Synthesized Detector" Receiver. We observe from the preceding analysis that detectors containing much more background than signal intensity do not contribute significantly to the error probability, since the outputs of these detector elements are multiplied by weights that are close to zero. This observation suggests the following suboptimum decoder concept with greatly simplified structure:

1) List the detector elements in decreasing order of signal intensity.
2) Compute the probability of error for the first detector element plus background.
3) Compute the probability of error for the sum of signal energies from the first two detector elements (plus background for two detector elements).
4) Continue this process until the minimum error probability is reached.

Each set of detectors may be effectively considered to be a single detector, so that no weighting is applied to account for variations in the signal distribution over the detector elements included in that set. The set of detector elements that achieves the minimum probability of error is the best "synthesized single detector" matched to the signal intensity distribution. Note that this straightforward process of performing the optimization by actually calculating the error probabilities for each partial sum of detectors is not practical. Later in this paper we shall describe some practical methods for approximating this procedure.

In effect, we have partitioned the logarithmic weights into two classes: "large" weights were assigned the value one, while "small" weights were assigned the value "zero." We shall show that this simple partitioning achieves near-optimum performance in low to moderate background environments, but with greatly reduced decoder complexity.

For the "adaptive synthesized single detector," the probability of correct decision can be obtained directly from Eq. (6.2-8), by setting $\alpha_k = k$ in the probability densities, and assuming constant signal and background intensities over each time-slot, yielding

$$p_q\left(k \mid H_q\right) = \frac{\left(\lambda_s\tau + \lambda_b\tau\right)^k}{k!} e^{-\left(\lambda_s\tau + \lambda_b\tau\right)} \text{ and } p_i\left(k \mid H_q\right) = \frac{\left(\lambda_b\tau\right)^k}{k!} e^{-\lambda_b\tau} \quad (6.2\text{-}10)$$

Direct substitution of these Poisson densities into Eq. (6.2-8) yields

$$\begin{aligned} P_M(C) = \Bigg\{ & \sum_{r=0}^{M-1} \left(\frac{1}{r+1}\right) \binom{M-1}{r} \sum_{k=1}^{\infty} \frac{\left(\lambda_s\tau + \lambda_b\tau\right)^k}{k!} e^{-\left(\lambda_s\tau + \lambda_b\tau\right)} \left[\frac{\left(\lambda_b\tau\right)^k}{k!} e^{-\lambda_b\tau}\right]^r \\ & \times \left[\sum_{j=0}^{k-1} \frac{\left(\lambda_b\tau\right)^j}{j!} e^{-\lambda_b\tau}\right]^{M-1-r} \Bigg\} \\ & + M^{-1} e^{-\left(\lambda_s + M\lambda_b\right)\tau} \end{aligned} \quad (6.2\text{-}11)$$

where, again, $P_M(E) = 1 - P_M(C)$.

6.2.2.2.6 The Gaussian Approximation. When the array contains a large number of detector elements, the computation of the probability density of the weighted sum of Poisson random variables becomes prohibitively difficult. It is shown in reference [94] that a useful Gaussian approximation to the discrete density of the weighted sum of Poisson random variables may be derived from the characteristic function of the discrete density, leading to the following approximation:

$$\begin{aligned} P_M(E) &= 1 - P_M(C) \\ &\cong 1 - \int_{-\infty}^{\infty} dy \, \frac{e^{-y^2/2}}{\sqrt{2\pi}} \left\{1 - \frac{1}{2} Erfc\left(\frac{\sigma_{sb}}{\sigma_b} y + \frac{\eta_{sb} - \eta_b}{\sigma_b}\right)\right\}^{(M-1)} \end{aligned} \quad (6.2\text{-}12)$$

where the mean and variance of the approximating continuous random variables for signal and non-signal slots is η_{sb}, σ_{sb}^2, and η_b, σ_b^2, respectively, and where $Erfc(x) = 1/\sqrt{2\pi} \int_x^{\infty} e^{-y^2/2} dy$. In terms of the weighted signal intensities, the mean and variance are defined as

$$\eta = \sum_{m=1}^{K}\sum_{n=1}^{L} u_{mn}\,\lambda_{mn}\,\tau \text{ and } \sigma^2 = \sum_{m=1}^{K}\sum_{n=1}^{L} u_{mn}^2\,\lambda_{mn}\,\tau \qquad (6.2\text{-}13)$$

We expect this approximation to be accurate when either of two conditions are satisfied: when the average photon energies over each detector element are so high that the Poisson distribution can be well approximated by a Gaussian density (this would occur with intense background radiation), or when the conditions for the central limit theorem are satisfied. The central limit theorem applies if the average photon energies over the entire array are sufficiently similar, so that a large number of detector outputs can be considered identically distributed random variables, or when the array outputs can be partitioned into several groups of random variables, each containing a large enough number of random variables to justify the Gaussian model.

6.2.2.2.7 Performance Bounds for Poisson Detection. For the special case of binary PPM signaling, $M = 2$, the following useful form has been obtained by Hubbard [94]:

$$P_2(E) = \exp\left(-\Delta^2\right)\left\{\sum_{k=0}^{\infty}\left(\frac{\sqrt{\lambda_b\,\tau}}{\Delta+\sqrt{\lambda_b\,\tau}}\right)^k F_k(\psi) - \tfrac{1}{2}F_0(\psi)\right\} \qquad (6.2\text{-}14)$$

where $\Delta = \sqrt{(\lambda_s+\lambda_b)\,\tau} - \sqrt{\lambda_b\,\tau}$, $\psi = 2\tau\sqrt{(\lambda_s+\lambda_b)\,\lambda_b}$, $F_k(\psi) = e^{-\psi}\,I_k(\psi)$, and $I_k(\psi)$ is the modified Bessel function of order *k*. Since for any $\psi \geq 0$ and $k \geq 0$, $I_k(\psi) \leq I_0(\psi)$, a useful upper bound to Eq. (6.2-14) can be easily constructed as

$$\begin{aligned} P_2(E) &\leq \exp\left(-\Delta^2\right)\left\{\sum_{k=0}^{\infty}\left(\frac{\sqrt{\lambda_b\,\tau}}{\Delta+\sqrt{\lambda_b\,\tau}}\right)^k - \tfrac{1}{2}\right\}F_0(\psi) \\ &= \exp\left(-\Delta^2\right)\left(\frac{\sqrt{\lambda_b\,\tau}}{\Delta} + \tfrac{1}{2}\right)F_0(\psi) \equiv P_2(UB) \end{aligned} \qquad (6.2\text{-}15)$$

This upper bound on the binary PPM error probability can be further approximated in the limit of very small and very large background energies as

$$P_2(UB) \cong \begin{cases} \frac{1}{2}\exp\left(-\Delta^2\right) & \lambda_b \tau << 1 \\ \left(\frac{\sqrt{\lambda_b \tau}}{\Delta} + \frac{1}{2}\right)\frac{\exp\left(-\Delta^2\right)}{\sqrt{2\pi\psi}} & \lambda_b \tau >> 1 \end{cases} \tag{6.2-16}$$

The binary PPM error probability and its upper bound can also be used to further bound *M*-ary PPM performance for the single detector case, in a manner similar to the "union bound" familiar from the "additive Gaussian noise" problem. The result is an expression of the form $P_M(E) \le (M-1)P_2(E)$. The proof, presented in [95, Appendix B], is somewhat complicated by the fact that for the optical problem the probability of committing an error when attempting to resolve ties among maximal counts must also be taken into account.

6.2.2.2.8 Comparison of Exact and Approximate Performance Calculations. A performance comparison of the optimally weighted array receiver and the "adaptive synthesized single detector" receiver (also referred to as the "0-1" subarray in the figures) has been carried out for average background energies of $K_b \equiv \lambda_b \tau = 0.1$ and 1.0. Two different signal models were used: a simple "test" model where only 5 of the $16 \times 16 = 256$ total detector elements were assumed to contain signal energy while the rest are assumed to contain no signal, and a more realistic 16 × 16 detector array model where the signal distribution over the array was simulated using a Kolmogorov turbulence model as described in [96], and all 256 detector elements may contain some signal.

For the test model the proportions of the total average absorbed signal energy $K_s \equiv \lambda_s \tau$ over the five detector elements were assumed to be (1.0, 0.3, 0.2, 0.05, and 0.02). Equation (6.2-8) was evaluated using this model, and compared with results obtained via Monte-Carlo simulations. The results are shown in Fig. 6-60 as a function of the total average absorbed signal energy K_s. It is evident that optimal weighting (represented by large squares) yields somewhat better performance than the suboptimum "0-1 subarray" computed according to Eq. (6.2-11), and that greater improvements occur at greater background intensities: however, the improvements due to the significantly more complicated optimally weighted array are only about 0.3 dB at an error probability of 0.001 for the high background case.

The reason for using only five detectors in the examples of Fig. 6-60 is that convolving more than five weighted Poisson densities rapidly becomes prohibitively difficult due to excessive demands on computer memory. Therefore, only this "five-detector" example could be evaluated computationally through the use of the bound on the probability of correct

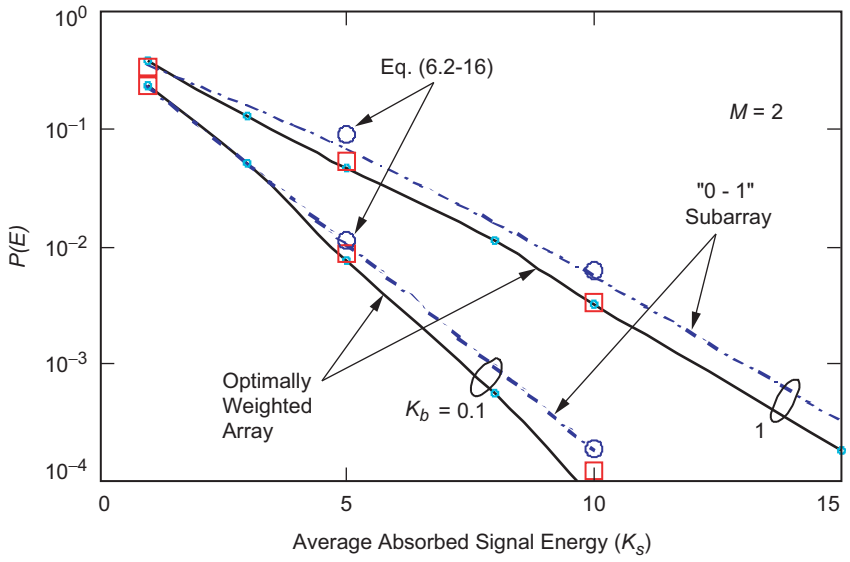


Fig. 6-60. Performance of optimally weighted array and adaptive "zero-one" subarray, binary PPM.

detection defined in Eq. (6.2-9), and only for background intensities not exceeding an average of 2 photons per detector per slot. Note the excellent agreement between the calculated error probabilities for the "0-1" subarray case and the approximate upper-bound of Eq. (6.2-16), represented by the large circles in Fig. 6-60 at average signal energies of 5 and 10 photons.

In Fig. 6-61, a realistic spatial distribution of the signal intensity over the focal-plane was generated using Kolmogorov phase-screens. Monte-Carlo simulations were performed to evaluate the error probability for the optimally weighted array. The "Gaussian" approximation to the error probability defined in Eq. (6.2-12) has also been evaluated for the binary PPM case, $M = 2$, with average background energies (per detector element) of $K_b \equiv \lambda_b \tau$ equal to 0.1, 1, and 5 photons/time-slot, as a function of the total average signal energy. It can be seen that the Gaussian approximation is close to the exact values obtained from Monte-Carlo simulation, and that good agreement is obtained even for small background energies per detector element, as direct comparison with the simulation results (large diamonds) indicates. In fact, it appears that this Gaussian approximation provides useful results over the entire range of background and signal energies represented in Fig. 6-61. The performance of the "0-1" subarray is also shown as the dashed curves: as expected, performance is somewhat worse than that of the optimally weighted array, but not significantly so.

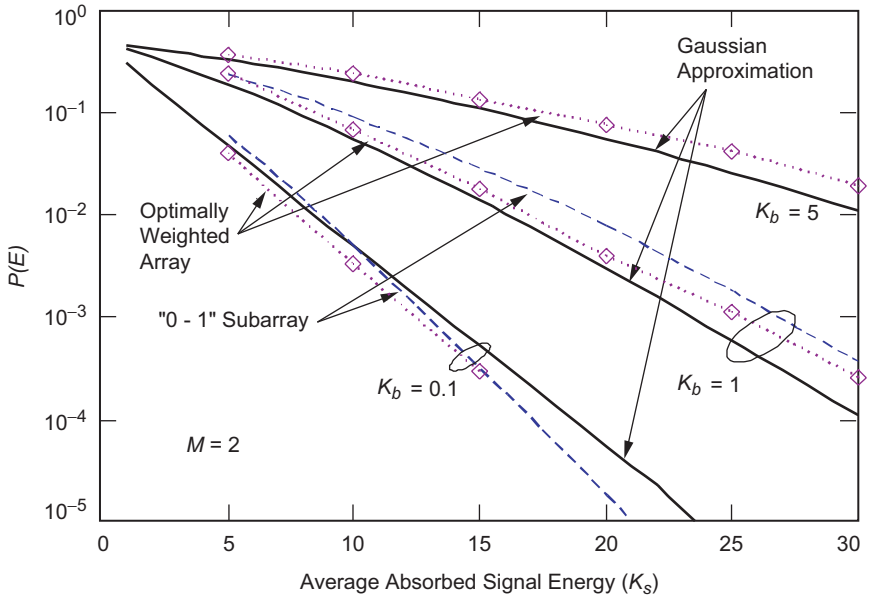


Fig. 6-61. Exact error probabilities, Gaussian approximations to the weighted sum, and simulation results for the "realistic" signal distribution described in 6.2.2.3 Numerical Results.

6.2.2.3 Numerical Results

6.2.2.3.1 Performance of Optimally Weighted Array. Both analytical calculations and Monte-Carlo simulations were performed in order to obtain PPM error probabilities for the "adaptive synthesized single detector" subarray. Performance of the optimally weighted array receiver was obtained from simulations: for each PPM symbol, M Poisson random variables with the proper statistics were generated, the optimum weights were applied, and the symbol corresponding to the largest observable selected. Simulated turbulence-degraded signal distributions were generated over the 16 × 16 detector array for all subsequent results. With no loss in generality, the transmitted symbol was always assumed to be the one corresponding to a signal pulse in the first slot. The detection process was repeated a large number of times (until 100 errors were accumulated) and repeated for increasing average signal energy with various background levels. Figure 6-62 shows the results of this simulation for $M = 2$, 16, and 256. The probability of bit error is shown as a function of the receiver's "photon efficiency," ρ, which is a measure of the average number of bits of information carried by each absorbed photon. It can be seen that with background levels of 0.1–1.0 photons per slot, $\rho = 0.3 - 0.5$ bits/photon can be

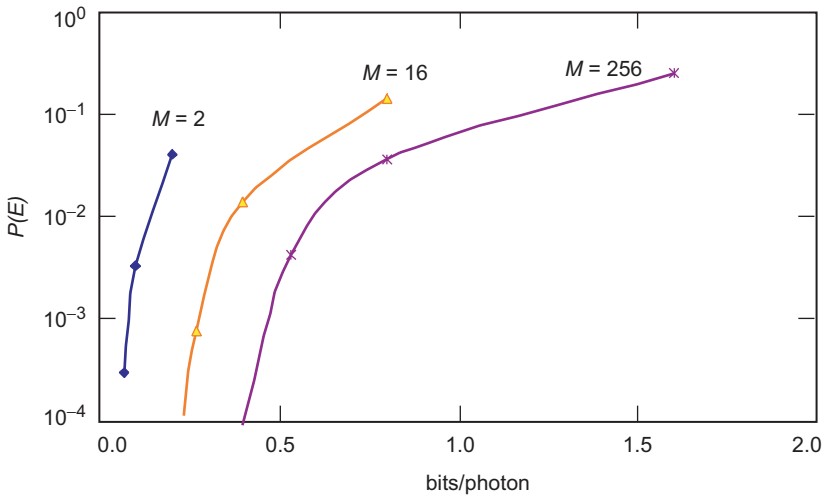


Fig. 6-62. Simulated PPM bit-error probability of optimum array receiver as a function of "photon efficiency" ρ (bits/photon), K_b = 0.1.

achieved with 256 PPM signaling at uncoded symbol error probabilities around 0.001–0.01.

6.2.2.3.2 Performance of "Adaptive Synthesized Single Detector" Receiver. In order to generate a spatial distribution of the signal incident upon the detector plane, a sample field was generated using a "Kolmogorov phase-screen" program [96], resulting in a matrix of complex signal amplitudes. For the simulation, an atmospheric correlation length of $r_0 = 4$ cm was assumed, which implies that the results should apply to any receiving aperture that is much greater than this correlation length [91]. The field intensity generated in the detector plane by the simulation is then integrated over the elements of a 16 × 16 detector array, which is assumed to encompass the extent of the signal distribution in the detector plane. Detector signal intensities are normalized so that for the *mn*-th detector we obtain an average number of absorbed signal photons of $\lambda_{s,mn}\tau$. A constant average background photon energy of $\lambda_b\tau$ is assumed over each detector element.

For a given sample function of the intensity distribution, the 16 × 16 = 256 detector elements were sorted in decreasing order of average signal energy, and *M*-ary PPM symbol error probabilities were calculated for increasing numbers of detectors, starting with the first detector, using Eq. (6.2-11). The lowest curve in Fig. 6-63 shows the symbol error probability for binary PPM ($M = 2$) as a function of the number of detector elements used, for the case $K_s = 10$ and $K_b = 0.1$ (that is, total average signal photons absorbed by the entire array is 10, and the average number of background photons *per detector element* is 0.1).

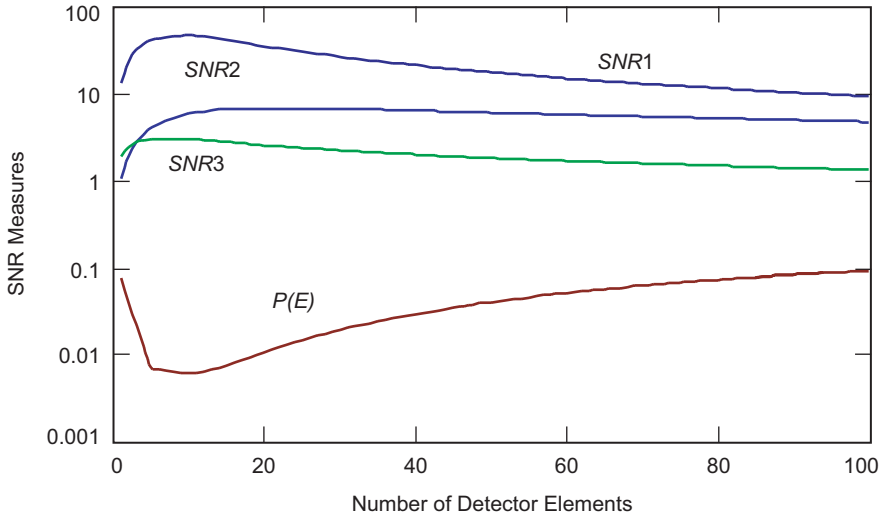


Fig. 6-63. Comparison of the optimum number of detector elements predicted by the three SNR measures.

It can be seen that for this case the smallest error probability of 0.0049 is achieved by assigning unity weight to the first 15 detector elements containing the greatest signal intensities, and zero to all the rest.

In Figs. 6-64 and 6-65, binary PPM symbol error probabilities are shown as a function of total average number of absorbed signal photons for four cases:

1) When the optimum number of "0–1" weighted detector elements are used,
2) When simulating the optimally weighted array,
3) When all 256 detector elements are given unity weight (synthesizing a large, nonadaptive single detector element), and
4) When an ideal "adaptive optics" system succeeds in concentrating all of the available signal energy into a single detector element, which then is the only detector element that is observed.

Using the same focal-plane signal distribution as before, error probabilities were computed for average background photon counts of 0.1 and 1.0, shown in Figs. 6-64 and 6-65, indicating performance gains by the "adaptive synthesized single detector" over a single "large" non-adaptive detector of 2 and 2.8, respectively, at an error probability of 0.001, corresponding to 3 dB and 4.5 dB of performance improvement. Note the excellent agreement of the approximation of Eq. (6.2-16) (large circles) with the computed values. When compared to the ideal "adaptive optics" receiver that concentrates all of the collected signal energy in a single element of the array, the gains are 3.8 and 8.2, corresponding to 5.9 dB and 9.1 dB of improvement. Note that the optimal weighted array yields only about 0.3 dB improvement over the optimized "0–1"

subarray at a symbol error probability of 0.001, even with relatively high background energy of $K_b = 1$.

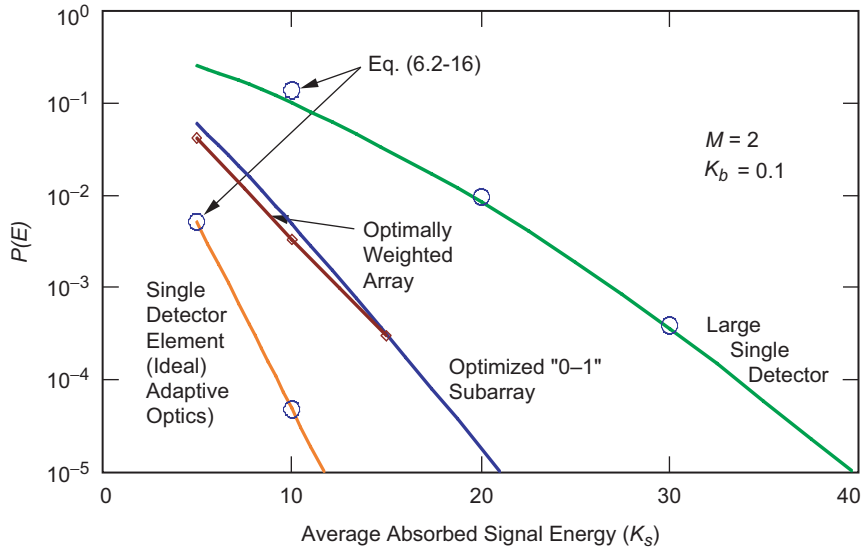


Fig. 6-64. Binary error probability of "large single detector," optimally weighted array, optimized "0–1" subarray, and single detector element with ideal adaptive optics: $M = 2$, $K_b = 0.1$.

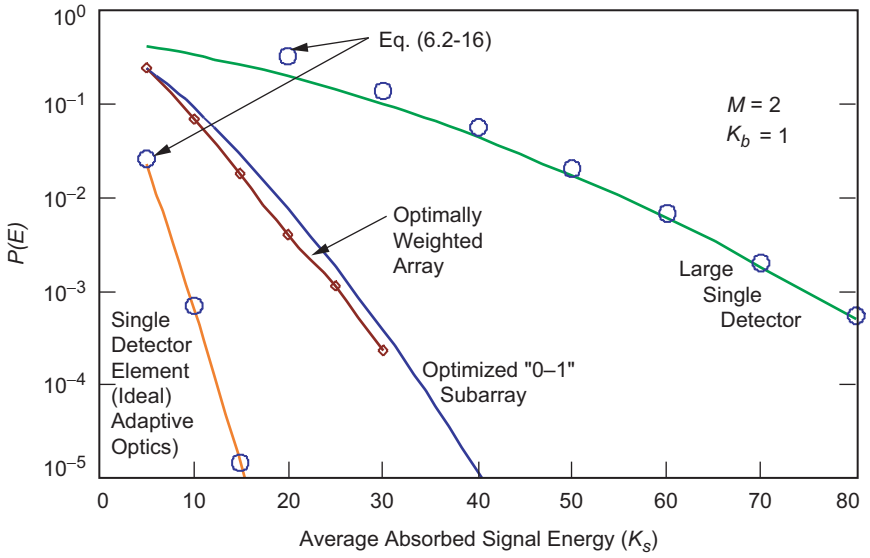


Fig. 6-65. Binary error probability of "large single detector," optimally weighted array, optimized "0–1" subarray, and single detector element with ideal adaptive optics: $M = 2$, $K_b = 1$.

Similar gains are evident in Fig. 6-66, which represent the symbol error probability, $P(SE)$, of the optimized subarray observing 16-dimensional ($M = 16$) PPM. The accuracy of the "union bound" evaluated for the case $\lambda_b \tau >> 1$ is evident (large circles), especially at the lower error probabilities.

Performance improvements were also obtained for several different focal-plane distributions at an average background energy of one photon per detector per slot, in order to verify that the above results were typical. Numerical results have shown that three out of four simulations yielded performance comparable to that of Fig. 6-65, requiring approximately 26 signal photons to achieve an error probability of 0.001 while utilizing 9 to 13 elements of the array in the region of error probabilities examined. One of the phase-screen simulations yielded an unusually favorable signal distribution that achieved optimum performance with only 4 detector elements, but still needed roughly 22 signal photons for 0.001 error probability. We can conclude, therefore, that for different distributions generated using the same atmospheric and receiver parameters, similar receiver performance is obtained; hence it is reasonable to use a "representative" signal distribution for the subsequent numerical analysis. Note that as more detector elements are used the performance tends to deteriorate, since a greater amount of background energy is collected by the receiver.

6.2.2.4 Summary and Conclusions. The problem of improving the performance of ground-based optical receivers through the use of photon-

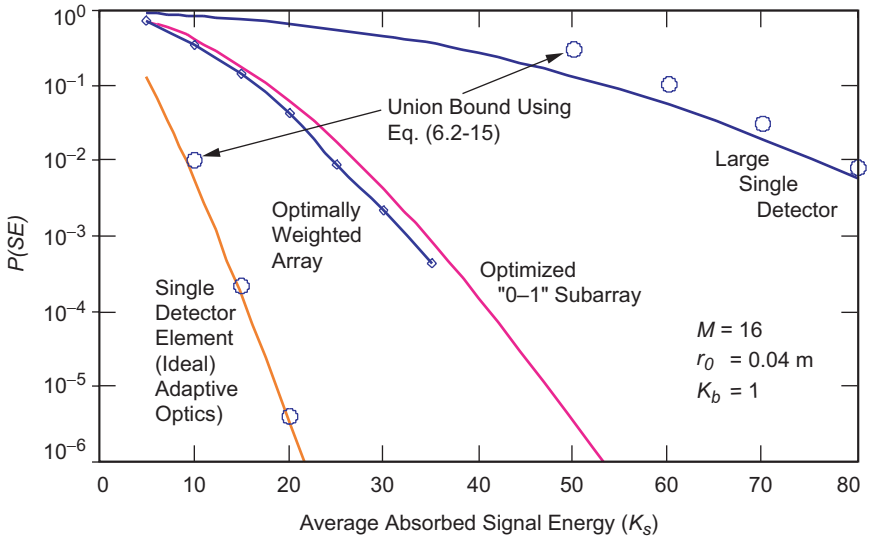


Fig. 6-66. Error probability of "large single detector," optimally weighted array, optimized "0–1" subarray, and single detector element with ideal adaptive optics: $M = 16$, $K_b = 1$.

counting detector arrays together with optimum signal processing algorithms has been addressed. The optimum array detection algorithm was derived, and a simpler suboptimum structure based on the optimum algorithm was also defined. Exact and approximate expressions for the error-probability performance of these structures were derived, and simulations were carried out to verify the analytic performance calculations. Realistic sample-functions of turbulence-degraded focal-plane signal distributions were generated using the Kolmogorov phase-screen algorithms described in [96], corresponding to moderate daytime turbulence (coherence-length of 4 cm), and used to evaluate the performance of optimum and suboptimum array detection algorithms designed for PPM signals. Performance improvements of up to 5 dB were demonstrated over a single "large" detector designed to collect most of the turbulent signal, when operating in the presence of moderate to strong background radiation. The use of "ideal adaptive optics" in front of the detector, which concentrates most of the signal energy into a single detector element, was evaluated and found to provide approximately 3 dB of additional improvement at an error probability of 0.001. It was shown that in cases of interest, the simpler suboptimum detector array algorithm performs nearly as well as the optimal array, with considerable savings in computational complexity.

6.3 Receiver Electronics

Andrew A. Gray, Victor A. Vilnrotter, and Meera Srinivasan

6.3.1 Introduction

The objective of this chapter is to describe discrete-time demodulator architectures for broadband optical pulse-position modulation (PPM) that are capable of processing Nyquist or near-Nyquist data rates. These architectures can provide numerous advantages for realizing communications demodulators in digital very large scale integrated (VLSI) circuits. The architectures are developed within a framework that encompasses a large body of work in optical communications, synchronization, and multirate discrete-time signal processing and are constrained by the limitations of the state-of-the art in digital hardware. This chapter attempts to create a bridge between theoretical communication algorithms and analysis for deep-space optical PPM and modern digital VLSI. The primary focus of this work is on the synthesis of discrete-time processing architectures for accomplishing the most fundamental functions required in PPM demodulators, post-detection filtering, synchronization, and decision processing. The architectures derived are capable of closely approximating theoretical performance of the continuous-time algorithms from which they are derived. The work concludes with an outline of the development path that leads to hardware.

This work builds on a large body of previous work in optical communications in particular *Optical Communications* by Gagliardi and Karp [97], "Design and Analysis of a First-Generation Optical Pulse-Position Modulation Receiver" by Vilnrotter and Srinivasan [98], and many others that are referenced later in this work. The focus in this work is not on the analysis of performance of optical communications systems but rather synthesis of discrete-time architectures suitable for realizing the functions of certain continuous-time processing using modern digital VLSI. The primary discrete-time signal processing building blocks are presented in this work along with preliminary performance results; the tasks required for synthesizing more complete demodulator architectures and analyzing their performance are outlined in the conclusion.

Figure 6-67 illustrates a block diagram of the optical transmitter and ground receiver system. The development path of the digital demodulator is also indicated from an oversimplified high level; a more complete description of this process is presented in the conclusion. The work presented here is one step in the development process starting from theoretical formulation and ending with implementation of the digital demodulator. Many elements of the discrete-time architectures developed here are fundamental and remain valid independent of whether or not the communications channel is radio frequency or optical, and some of these elements are derived with the intent to overcome limitations of state-of-the-art digital hardware.

The architectures are parameterized to utilize the tremendous flexibility achievable with modern digital hardware and to satisfy a large range of system requirements. System requirements play an increasingly critical role in the development as the steps in the process of Fig. 6-67 get closer to implemented hardware. Many system requirements may be met with numerous variations on discrete-time architectures developed herein. The final determination of which architecture is to be fully developed to implementation and the determination of numerous parameters are determined from specific requirements, and constrained by implementation considerations such as commercially available high-speed chips and board technology.

Next, a brief overview highlighting the objective of each section is given. Section 6.3.2 is an introduction to discrete-time architectures. The motivation for their development is discussed and includes the many advantages of implementation of signal processing using digital VLSI circuits.

Section 6.3.3 introduces discrete-time synchronization and slot filtering, and it introduces the impact of signal dynamics with fixed sample rate systems employing this processing. The motivation for fixed-phase analog-to-digital (A/D) sampling and the time-varying slot or interpolation filter is presented.

Section 6.3.4 extends the results of Section 6.3.3 to more complete demodulator models. Simplified models of demodulators are presented here, and these are extended throughout the remaining sections.

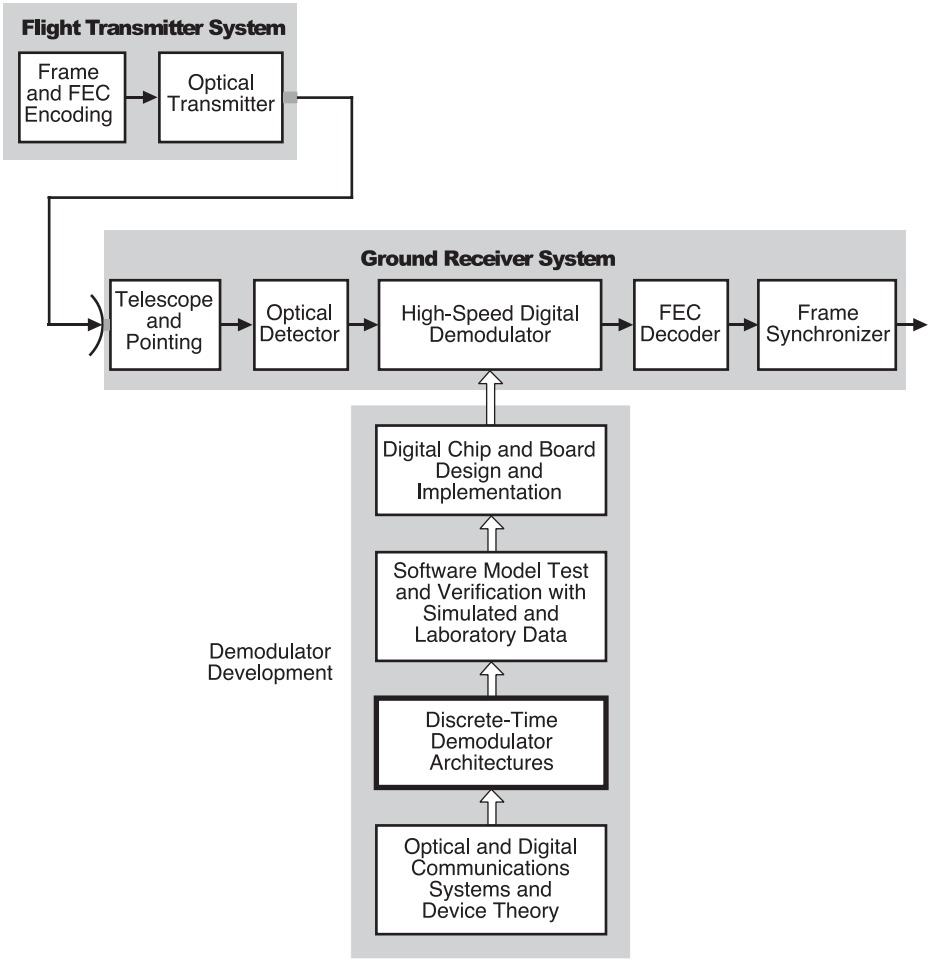


Fig. 6-67. Block diagram of the system and simplified development path for the digital demodulator.

In Section 6.3.5 problems identified in Sections 6.3.3 and 6.3.4 are addressed by various signal processing methods including non-linear processing. The methods for combining post-detection filtering and correction of synchronization errors are addressed with the time-varying post-detection or interpolation filters.

Section 6.3.6 presents parallel discrete-time demodulator architectures. The serial processing results are extended to parallel or vector processing which is required to achieve the processing necessary for broadband pulses requiring very high sample rates. This parallelization is performed on the core processing of slot synchronization and post-detection filtering. Symbol synchronization

and other demodulator algorithms are not explicitly addressed here but are considered much more straightforward to design in discrete-time.

In Section 6.3.7 asynchronous discrete-time processing is addressed. The fixed-rate processing of a signal that contains modulated data with a rate that is asynchronous to the sample and system clocks creates significant challenges, specifically asynchronous digital design and implementation. The asynchronous processing is shown to create particular challenges for the parallel architectures.

Section 6.3.8 introduces representative discrete-time architectures derived using the methods, framework, and signal processing designs developed herein. In this section we use generic parallel discrete-time algorithms and methods developed earlier and existing in the literature to synthesize specific designs that incorporate capability for trading processing rate with complexity. The performance of a software model of a receiver architecture is given that includes a simplified model of the optical channel. While the demodulator architectures presented are not complete, they encompass many required functions, and the evolution to more complete architectures is presented in brief.

Section 6.3.9 presents system models along with certain design/ implementation equations. Frequency and digital transmission line models are presented. The parallel discrete-time demodulator is described in terms of bandwidth and clock rates. This model is useful in establishing specific design parameters of the parallel digital demodulator and implementation platform from requirements such as data rates, PPM pulse bandwidth, and other requirements. The transmission line model is useful for high-speed digital platform design by establishing the primary input/output (I/O) requirements on the implementation platform.

In Section 6.3.10 the conclusion is presented that provides an overview of the work presented and indicates the next steps in the development of a broadband hardware implementation. These steps encompass discrete-time design and analysis, extensive software modeling and simulation, and hardware design tools and methods and realization. The conclusion includes an overview of the state-of-the art processes that lead to implementation in modern VLSI platforms and devices. Parts of this process are derived from targeting VLSI implementation in field programmable gate arrays (FPGAs). These devices have in many instances enabled improved development strategies of complex VLSI systems, reducing development risk while facilitating more aggressive schedules and development processes. The development process proposed incorporates extensive use of a variety of software models and tools, some of which are generic and used for system and subsystem validation and others that are device specific. Modern communications systems, such as the one developed here, are generally beyond any type of comprehensive closed-form analytical analysis or performance evaluation and rely extensively on Monte-

Carlo simulation. However, as part of the process, illustrated analytical models and bounds are used extensively to validate various subsystem performances.

6.3.2 Introduction to Discrete-Time Demodulator Architectures

Receivers are arguably the most complicated processing element (hardware or software) in a communications system. Modern digital receivers must be flexible enough to process parameterized modulation schemes, pulse shapes, and data rates. Receiver complexity is directly related, although not necessarily linearly, to the complexity of the modulation type used in the system. Many modulations developed for optical communication are complex, with some so complicated that it may be impractical to implement them in hardware. The PPM modulation was selected for this monograph because it appears to be the modulation of choice for most future deep-space optical satellite communications systems. In addition, the demodulator architectures developed are based on the assumption of single-symbol decision processing, as opposed to symbol decisions that are based on maximum-likelihood sequence detection [99,100]. However, the demodulator may process input signals to generate soft symbols at the output of the post-detection filter that may be further processed by a forward error correction decoder (maximum-likelihood sequence decision processing).

With the availability of VLSI technology, coupled with the flexibility of discrete-time signal processing algorithms, it is very desirable to implement receivers with as much digital processing capability as possible [101]. Dynamic range in filter bandwidths of greater than 8 orders of magnitude is feasible, and similar flexibility exists with other processing functions when implemented in modern digital circuits. In addition, an all-digital demodulator implemented using complementary metal oxide semiconductor (CMOS) technology has great advantages in size and reliability and greatly reduced reproduction costs over analog demodulators as well as other digital technologies [102,103]. The number of options for realizing digital VLSI circuits is extensive, and ranges from many FPGAs, commercially available digital signal processors (DSPs), and applications specific circuits (ASICs). FPGAs in particular have increased the flexibility of high speed digital processing, with large scale reconfiguration of VLSI possible with relatively rapid design cycles.

Often the only true limitation to the data rates an all-digital demodulator can process is the A/D converter. However, conventional CMOS digital demodulators currently have substantially lower clock rates than the fastest commercially available A/D converters. In such a high rate system, the minimum number of samples required for discrete-time processing, the Nyquist rate, is required to process the maximum data rate possible, the Nyquist data rate. Because commercially available CMOS hardware generally has lower clock rates than the fastest A/D converters, CMOS digital demodulators using

traditional serial algorithms for digital communications process data rates many times lower than the maximum Nyquist data rate. The rate difference motivates the development of parallel processing architectures.

It should be noted that, strictly speaking, there is no such thing as an all-digital receiver. There is always some analog processing for the optical-to-digital conversion, typically accomplished by optical-to-electrical (analog) and then conversion to a digital signal. For the purposes of this work, the definition of an *all-digital demodulator* is such that the modulated optical waveform is detected and converted to an analog current and then voltage signal. This voltage is sampled, and then the demodulator functions of post-detection filtering, symbol-decision processing, synchronization (including slot-timing and symbol recovery for PPM), and symbol-to-bit conversion are performed using exclusively digital processing. The continuous-time A/D clock phase and frequency are not adjusted using feedback from the digital demodulator. There are many other functions of a demodulator that, although necessary, will prove more trivial to implement and are not dealt with in this work. These functions include slot and symbol-synchronizer lock detection, power estimation, error-control decoder preprocessing, and others. Here we derive discrete-time architectures for realizing the primary demodulator functions using known continuous-time processing, acknowledging that the processing utilized in a demodulator and the performance obtained are highly dependant on the optical channel used in the system. In the architecture development, we place particular emphasis on the two greatest and related challenges of a Nyquist data rate digital implementation, Nyquist rate sampling of communications signals and very high rate (bandwidth) processing.

6.3.3 Discrete-Time Synchronization and Post-Detection Filtering Overview

The PPM slot and symbol synchronization, post-detection filtering of the input signal, and symbol decision processing form the core of the optical PPM satellite-communications receiver. Synchronization and post-detection filtering are often thought of as separate operations; however, we shall demonstrate here that they may be combined in an intuitive way to create time-varying discrete-time filter architectures. Furthermore, these architectures are well suited for and motivated by all-digital implementation.

The post-detection filter is a specified operation that is used to process the received signal in a precise time-synchronized fashion. The incoming waveform or received signal is often convolved with the post-detection filter, although there are other types of processing employed depending on the system [97]. In the development of the discrete-time architecture here we assume that the incoming sampled waveform is convolved with a post-detection filter. The waveform analysis is greatly simplified and ignores the statistical nature of the

signal at the output of the optical detector. This simplified analysis should be thought of as a conceptual tool to reveal many fundamental elements of the parallel processing discrete-time architecture design that are largely independent of the statistical nature of the input signal even if the channel is radio frequency or optical. This approach results in a general architecture for accomplishing time-synchronized processing that encompasses or is readily extended to other processing that does incorporate the detailed nature of the input signal. Regardless of the post-detection filter coefficients (or even if it is not a linear filter/convolution, e.g., threshold processing), precise time synchronization of the required signal processing in the presence of signal dynamics is generally required. Understanding the implications and properties of this precise time synchronization with Nyquist or near-Nyquist sampling motivates much of the discussion and analysis here; this understanding is critical in developing discrete-time demodulator architectures for Nyquist data rates. The results of this section are used in Section 6.3.8 to develop architectures with generalized slot-synchronization and post-detection filtering.

6.3.3.1 Discrete-Time Post-Detection Filtering. In the generic digital communications transmitter, baseband bits are mapped to modulation symbols, A_k. The transmit pulse shape is give by $p(t)$. The output of the transmitter is then

$$Q(t) = \sum_{k=-\infty}^{\infty} r_k \, p\left(t - kT_{sym} - A_k T_{slot} + \phi_k\right) \tag{6.3-1}$$

Here T_{sym} is the symbol period, T_{slot} is the slot period, r_k represents the intensity variations, and ϕ_k is a timing jitter sequence [104]. The timing jitter and intensity variation terms will be excluded from the simplified analysis and discussion here. As example, consider the 8-PPM signal set in Fig. 6-68:

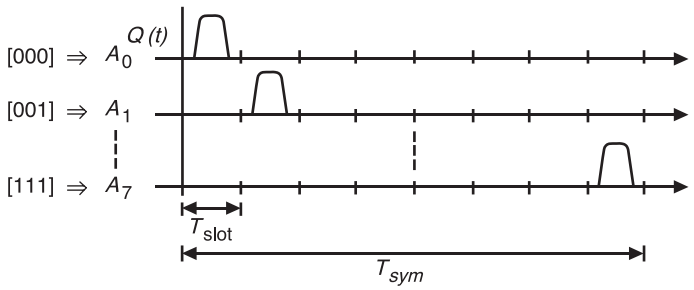


Fig. 6-68. Example PPM symbol set.

Transmitting the signal $Q(t)$ through a linear noiseless channel with impulse response, $b(t)$, will result in a received signal

$$R(t) = \int_{-\infty}^{\infty} b(\tau) \sum_{k=-\infty}^{\infty} p\left(t - kT_{sym} - A_k T_{slot} - \tau\right) d\tau \qquad (6.3\text{-}2)$$

This can be rewritten as

$$R(t) = \sum_{k=-\infty}^{\infty} h\left(t - kT_{sym} + A_k T_{slot}\right), \quad h(t) = \int_{-\infty}^{\infty} b(\tau)\, p(t-\tau)\, d\tau \qquad (6.3\text{-}3)$$

where $h(t)$ is the received pulse. The result of the continuous-time input filtered with some post-detection filter is then

$$y(t) = \int_{-\infty}^{\infty} R(t-\tau) f(t) d\tau \qquad (6.3\text{-}4)$$

Where the filter $f(t)$ may be designed from some optimality criteria or criterion derived for the optical channel. Here we derive the time-varying slot filter assuming a classical matched filter approach common in radio-frequency communications; this derivation will illustrate fundamental problems with discrete-time synchronization for which architectures are then designed to solve. The post-detection filter peak SNR or optimal output is obtained by sampling at the slot rate,

$$y_M(n) = y\left(T_{slot} n\right) \quad n = 1,2,3... \qquad (6.3\text{-}5)$$

It should be noted that in an operational wireless communications system with channel noise, including timing jitter, the desired signal or sample is generally not obtained, but some estimate or approximation is. For example, with jitter ϕ_k that varies from one pulse to the next, the optimal sample estimate $\hat{y}_M(n)$ will generally not equal the true optimal sample point, $\hat{y}_M(n) \neq y_M(n)$. If the system is designed properly, the estimate of the optimum sample (or samples) will have zero average error assuming there are no other unaccounted for dynamics or signal distortions and it is an unbiased estimator [105,106,107]. We progress assuming no jitter in the input signal so as to derive necessary fundamental concepts in a simplified fashion, but it must be understood that the slot synchronization algorithm design should average the effects of jitter and other channel noise sources in such a way that an unbiased estimate of timing offset is obtained and then corrected for. For a discrete-time demodulator system with sample period T_s,

$$h_d(n) = h\left(T_s n + \Delta t_R\right) \tag{6.3-6}$$

$$f_d(n) = h\left(-T_s n - \Delta t_M\right) \otimes K\left(T_s n\right) \tag{6.3-7}$$

$$R_d(n) = \sum_{k=-\infty}^{\infty} h_d\left(n - kT_{sym} + A_k T_{slot}\right) \tag{6.3-8}$$

$$y_d(n) = \sum_{k=-\infty}^{\infty} R_d(k) f_d(n-k) \tag{6.3-9}$$

Here Δt_R is the time offset of the received signal $h_d(n)$ and Δt_M is the time offset of the post-detection filter $f_d(n)$. In general, for $f_d(n)$ to have equivalent performance to the continuous-time post-detection filter the following condition must be met, $\Delta t_M = \Delta t_R$. (Other conditions must also be met, such as band limited signals.) If this condition is not met, the discrete-time system realizing $f_d(n)$ will generally result in different (probably reduced) performance compared with the continuous-time system with $f(t) = h(-t) \otimes K(t)$ [100]. Note that $K(t)$ is a designed impulse response derived from system specific criteria, the simplified case is an impulse and could perform the same role as a constant for scaling. So $f(t)$ is a convolution of the time-reversed received pulse and a design pulse $K(t)$. The output of the discrete-time filter may be downsampled at the slot rate,

$$y_D(n) = y_d(Dn) \tag{6.3-10}$$

Obtaining synchronized samples at the output of the post-detection filter is required to obtain the optimum filter output defined by some criterion. Obtaining this correct sample or samples (if no downsampling or another downsample rate is used) is the objective of slot synchronization. Any sampling offset [108] may result in performance degradation. This performance degradation may occur in varying degrees to the decision process and the slot-synchronization algorithm itself as some closed-loop slot-synchronization algorithms contain the post-detection filter in their feedback path [106]. Figure 6-69 illustrates a conceptual (post detected) received waveform and post-detection filter demonstrating ideal and non-ideal slot timing at some time t_0. Note that if $\Delta t_M \neq \Delta t_R$, the peak will not be represented by a discrete-time sample, but the error may be removed or minimized via a variety of discrete-time processing methods; however, these may entail significant challenges to realizable in real-time hardware with very high-rate sampled systems.

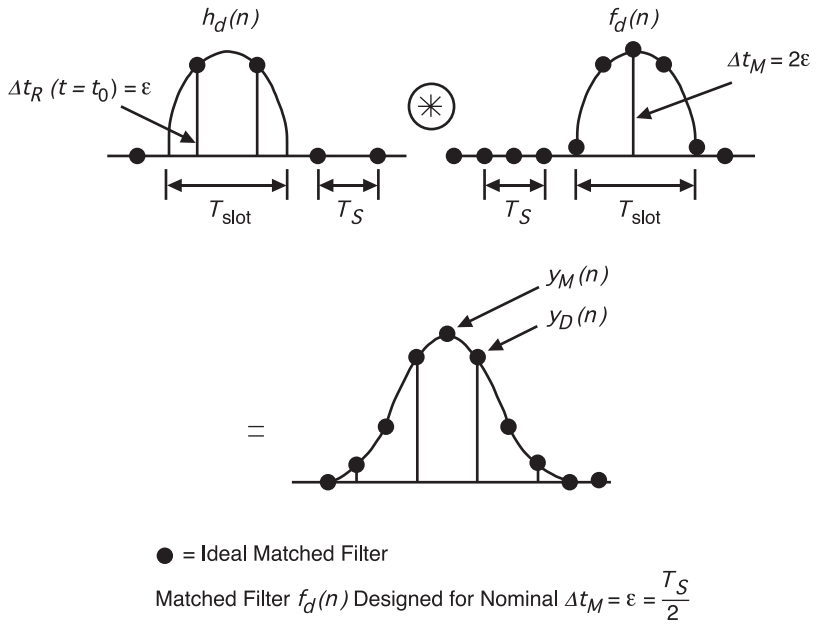


Fig. 6-69. Illustration of discrete-time post-detection filtering.

In this conceptual example, an arbitrary slot or receive filter was used. This could just as well be an integrate-and-dump filter (moving average filter with downsample); in which the signal may be synchronized before the integrate-and-dump by some time-varying interpolation filter to adjust the phase of the signal (an example of this is given in Section 6.3.7). Recall that in a discrete-time communications receiver with very high-data (sample) rates when flexibility is required or when the receiver is used for ranging, it is very desirable to fix the sample clock. In addition, adjusting the oscillator may be very difficult or impractical in near-maximum-Nyquist data rate receivers due to the very high clock rate, significantly greater than 1 GHz with modern A/Ds. However, when sampling the received signal at a constant sampling rate f_s, the sample offset Δt_R defined earlier changes with time, that is $\Delta t_R(t)$. This is due to the non-linear Doppler channel created by the relative velocity between ground stations and the transmitting satellite and the difference between the transmitter and receiver clocks. These differences are impossible to predict precisely. The difference between the discrete-time post-detection filter output and the ideal continuous-time post-detection filter output is the *fractional sampling offset*. The fractional sampling offset (or simply sampling offset) can be no larger than one half a sample period, or $T_s/2$. For the downsampled discrete-time post-detection filter output to be equal to the continuous-time matched filter sampled output, $y_D(n) = y_M(n)$, from the previous equations,

$$
\begin{aligned}
y_D(n) &= y_d(nD) \\
&= \sum_{k=-\infty}^{\infty}\left(\sum_{m=-\infty}^{\infty} h\left(T_s\left(k - mT_{sym} + A_m T_{slot}\right) + \Delta t_R\right) f\left(nT_S D - k\right)\right) \qquad (6.3\text{-}11) \\
&= \int_{-\infty}^{\infty}\left(\sum_{m=-\infty}^{\infty} h\left(\left(nT_{sym} - \tau\right) - mT_{sym} + A_k T_{slot}\right)\right) f(\tau)\, d\tau = y_M(n)
\end{aligned}
$$

This equality cannot be valid unless the following condition is met:

$$T_s nD + \Delta t_R = T_{slot} n \qquad (6.3\text{-}12)$$

Note that Δt_R is varying with time, and so is T_{slot}, but in a continuous-time processing system it is assumed that the oscillator frequency and phase are adjusted such that the matched filter output sampling time T_{slot} is optimal. Recall that T_s is fixed for the discrete-time system proposed. This fact (combined with a time-varying Δt_R with fixed Δt_M) implies that the discrete-time matched filter peak, or other optimal sample or samples if another optimality criterion is used, in general will not be represented by the discrete-time samples of the post-detection filter output. Furthermore, it should be noted that when Nyquist rate sampling is utilized, the sample period is the largest possible, making the potential sampling offset the largest possible. The A/D oscillator phase could theoretically be adjusted such that Δt_R is obtained to match Δt_M.

The slot synchronization algorithm estimates the slot time delay ($\Delta t_\varepsilon = \Delta t_M - \Delta t_R$) and frequency $1/T'_{slot}$ from the received signal $R(t)$ where the true slot frequency is $1/T_{slot}$. Slot synchronization for the optical channel may be accomplished by a number of algorithms, ranging from optimal estimators to highly sub-optimal estimators, and may be an open-loop estimator or a closed-loop estimator with feedback. It turns out that detecting or estimating the error in a discrete-time or digital implementation with Nyquist-rate sampling at very high bandwidths is often less challenging than correcting for it in real-time. While the type of estimator and its estimation performance are key considerations in any demodulator design and implementation, a nominal algorithm is chosen from [100], and the development presented here is focused primarily on how to correct for time-varying synchronization errors that the algorithms estimates. It will become apparent that most known estimation algorithms can be used in conjunction with the architecture developed with proper modification and/or discrete-time design. However, determining the performance of the resulting discrete-time system may be very challenging.

A property of the PPM demodulator is that once slot synchronization is estimated, and the slot clock determined, then the symbol clock may generally be readily derived, and hence the bit clock may also be derived. Symbol synchronization is required to determine symbol boundaries so that the slot position of the pulse in the symbol can be estimated; the estimate $\hat{A}_k$ can be made. This process is referred to as the decision process, refer to Fig. 6-70. Following this estimate, symbol-to-bit mapping is performed; this operation inverts the bit-to-symbol mapping at the transmitter. Following this processing forward error correction decoding may be performed if coding is used in the system. In modern systems the process of symbol-to-bit mapping and decoding may be performed in a coupled fashion. Figure 6-70 illustrates a simplified optical communications receiver. Note the two-step symbol process: first, slot synchronization (slot clock recovered) and pulse filtering are performed on all slots to determine $y_M(n)$; second, symbol synchronization and symbol decision processing are accomplished to form the symbol estimates $\hat{A}_k$. Other processing follows, such as forward error correction and frame synchronization.

Figure 6-71 illustrates a simplified discrete-time receiver with discrete-time post-detection filter. This system employs a constant clock rate and synchronous processing, and therefore it is a greatly simplified model. The slot

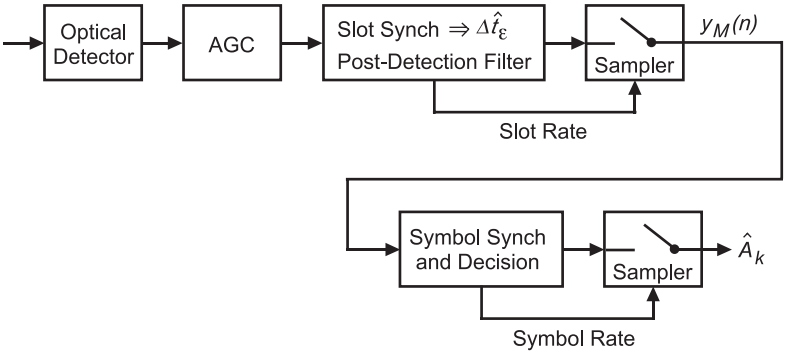


Fig. 6-70. Simplified primarily continuous-time receiver.

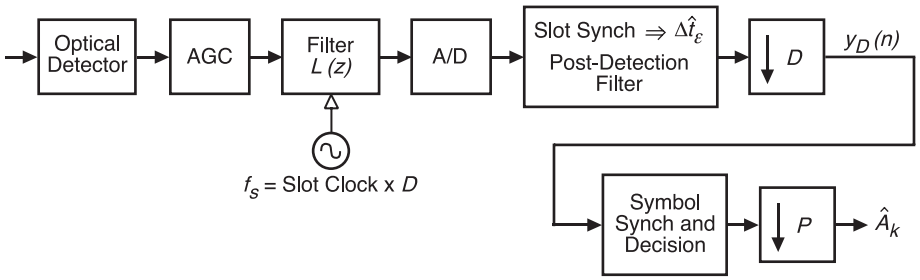


Fig. 6-71. Simplified discrete-time receiver.

synchronization algorithm recovers the true slot clock by estimating slot phase and frequency offset from the nominal slot clock, this nominal clock frequency is $1/T_S$. By obtaining the estimate $\Delta\hat{t}_\varepsilon$ of Δt_ε the phase of the slot clock is determined and the frequency may also be estimated using successive $\Delta\hat{t}_\varepsilon$ estimates. The number of samples per slot for this signaling is $D = T_{slot}/T_S$, the slot period divided by the sample period. The number of samples per PPM symbol is $C = DP$, where P is the number of slots per symbol, that is $P = 2^m$. Next, we will analyze the critical deficiencies in this model and use the results to develop appropriate discrete-time signal processing algorithms and architectures to overcome these deficiencies.

If the Nyquist criterion is met, the peak, or any other sample(s), may be obtained or estimated using discrete-time phase delay methods [109,110]. One method for realizing the latter is to change the discrete-time slot filter, $f_d(n)$, with time to match the time-varying incoming sampled received pulse shape such that $\Delta t_\varepsilon(t) = \Delta t_R(t) - \Delta t_M(t) = 0$ or with noise sources in the system the mean estimate $\overline{\Delta} t_\varepsilon(t) = \overline{\Delta} t_R(t) - \overline{\Delta} t_M(t) < \varepsilon_0$ and the variance of the estimate $\mathrm{var}\left\{\overline{\Delta} t_\varepsilon(t)\right\} < \varepsilon_1$ for some design parameters ε_0 and ε_1. Another method is to use a generic time-varying interpolation filter. Therefore, in a discrete-time system slot-synchronization can be thought of as having two components. The sampling offset must be removed ($\Delta t_M(t) = \Delta t_R(t)$) or minimized to ensure that one of the discrete-time sample (or samples), determined by some selection criterion, represent the slot energy in the decision process. The discrete-time architecture must minimize the fractional sampling offset in order to approximate the performance of the ideal continuous-time system.

Static timing or phase offset has been considered under the assumption the received signal phase has first and second derivatives that are zero. This is just the first step in developing effective demodulator architectures, but it should be realized that in an operational system with scenarios including Doppler, the first and second derivatives of phase are generally not zero (although higher derivatives can often be ignored [111]). There are frequency offsets and frequency rate changes. These correspond in not only a time-varying phase shift of the post-detection filter but also possibly a change in the post-detection filter impulse response itself as the pulse, slot, and symbol periods either expand or contract with Doppler. These nonlinear pulse distortion considerations are beyond the scope of this work. However, it will be become clear that once the basic framework for developing, analyzing, and implementing a discrete-time architecture with time-varying phase capability is completed, extending it to higher order dynamics is possible and is very similar to previous work in the field of synchronization.

In summary, we have demonstrated that in a discrete-time system with signal dynamics and a fixed sample-clock, slot synchronization/post-detection

filtering poses challenges to discrete-time design and implementation, particularly with Nyquist sampling. These concepts are considered further in Section 6.3.5 in the development of discrete-time architectures for accomplishing synchronization error correction for broadband pulses.

6.3.3.2 Slot and Symbol Synchronization and Decision Processing. Here a brief summary is presented of known signal processing for performing slot and symbol synchronization, post-detection filtering, and decision processing outlined in [97,99,100] for a simple optical channel model. References for previous work include [112–127]. Some of the structures presented here do not represent unique or optimal solutions but may represent good design choices based on analysis and systems presented in the references given. Figure 6-72 illustrates the conceptual continuous-time slot synch, a variation of which is analyzed in [128].

The discrete-time or sampled version of Fig. 6-72 is given in Figs. 6-73, 6-74, and 6-75. To date, performance of the discrete-time system has not been analyzed in closed form. The filter structure is the generic discrete-time version (using the impulse invariant transformation) of the continuous-time closed-loop second order loop filter used commonly in phase-locked loop, Costas loops, data-transition tracking loops, early-late gate type loops, and many other synchronization loops. The performance and loop-filter design must be determined through analysis analogous to that performed numerous times on similar synchronization loops that exist both in published research and analysis and in operational systems assuming the additive white Gaussian noise channel; examples of these derivations may be found in [129,130]. This filter structure is derived from the two-pole continuous-time loop and may be used to track both phase and frequency. Note that the loop filter $L(z,l)$ in Figs. 6-74 and 6-75 may be extended to include more poles (increasing the order of the loop) and is

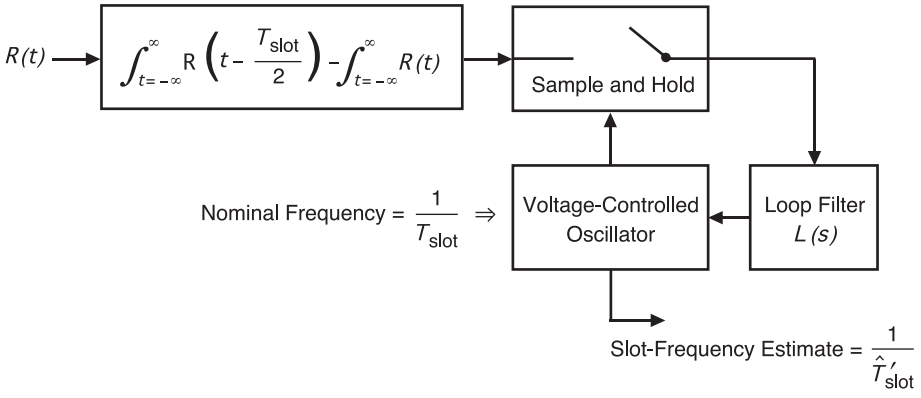


Fig. 6-72. Continuous-time model of the slot-synchronization loop.

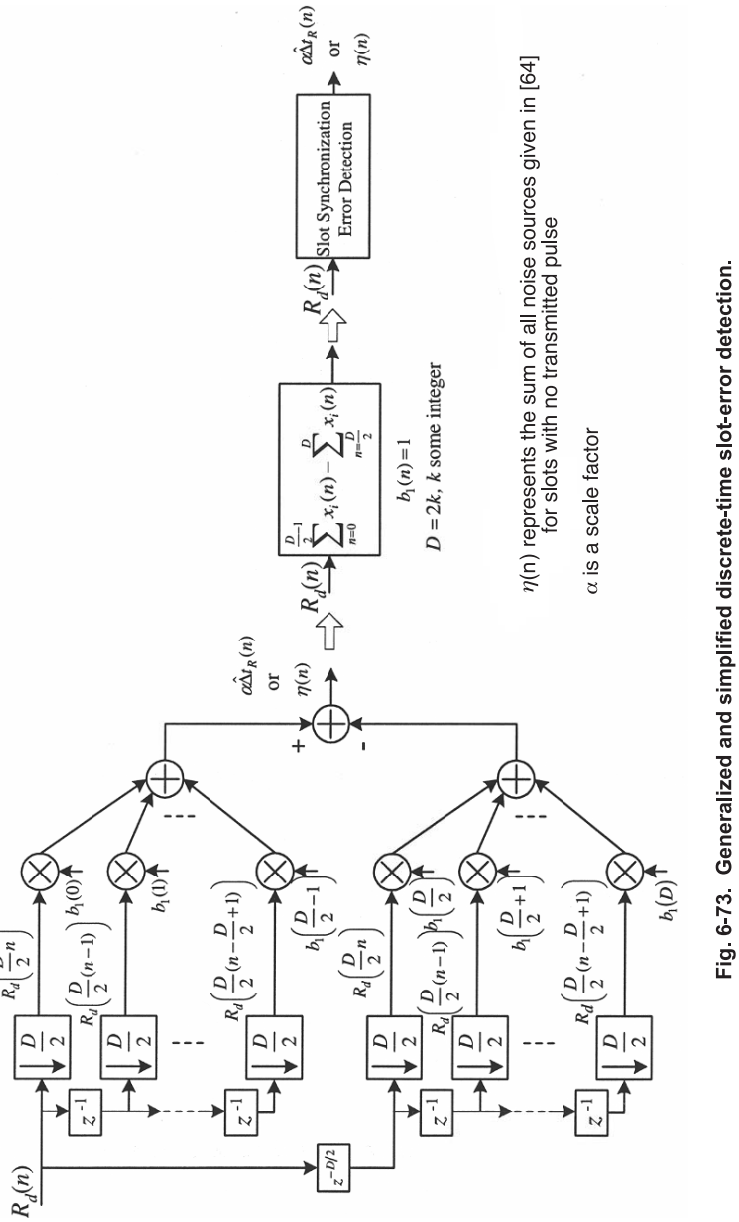


Fig. 6-73. Generalized and simplified discrete-time slot-error detection.

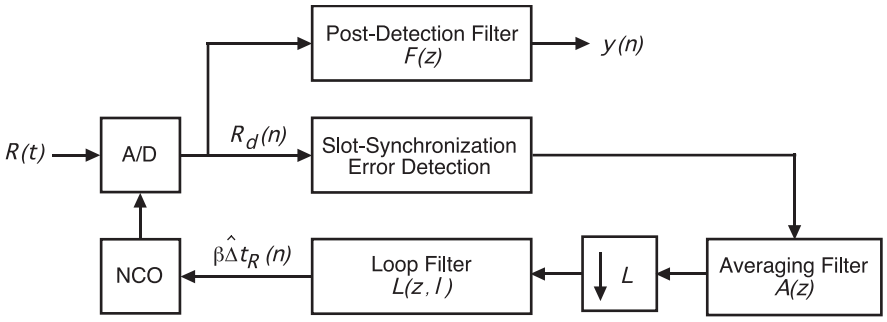


Fig. 6-74. Discrete-time slot-synchronization loop.

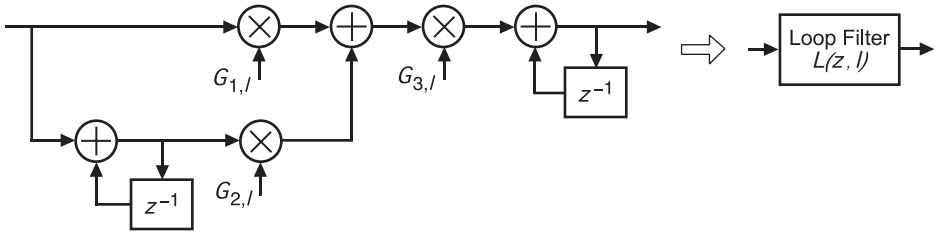


Fig. 6-75. Discrete-time and time-varying loop filter structure.

a function of *l*. The latter characteristic indicates the ability to switch on-the-fly the filter bandwidth, which is commonly done in communications systems. The filter coefficients may be changed on the fly to achieve different bandwidths or damping factors and increase performance during the various regions of operation of the receiver. For example, tracking bandwidths are often different than acquisition bandwidths, and it may be desirable to change bandwidths or even tracking loop order with signal-to-noise ratio and/or signal dynamics. Obviously an intelligent controller, such as a state-machine or human-in-the loop, must be used to control such changes.

Figure 6-76 illustrates the operations of symbol synchronization and decision processing. The various algorithms fall into two general categories, those requiring a pilot sequence to acquire and track and those that operate on random data sequences, the so-called blind acquisition and tracking algorithms. There is relatively little previous work on evaluating the performance of these algorithms with the specific optical channel as described in [97,99]. It is assumed in this development that an algorithm using a periodic pilot sequence will be utilized, making the acquisition and tracking of a large range of PPM orders relatively straightforward.

From [97] the post-detection and decision processing is highly dependant on the characteristics of the optical channel and the optical detector used. Given the channel and operating scenarios outlined in [99], the slot-synchronized

decision process of choosing the slot with the greatest energy from the slots within a symbol boundary to estimate which symbol was transmitted is nearly optimum for a large range of system parameters. This is fortunate as this decision processing is readily accomplished with digital processing and may be easily parallelized. Figure 6-77 illustrates a simple example of the detection and decision process. Here ideal slot and symbol synchronization are assumed. The filter $F(z)$ is a simple moving average filter with D coefficients $(f(n)=1)$, where $D=4$ is the number of samples per slot, and the output is downsampled by D. The output of the downsampler is the sum of all the samples in a slot. These samples are then processed to decide which slot has the greatest energy. The samples in the slots other than the slot containing the transmitted pulse are non-zero due to the noise on the input signal.

Although symbol synchronization and detection/decision processing are essential processing elements, and their performance must be evaluated carefully in any theoretical or practical receiver development, they generally are

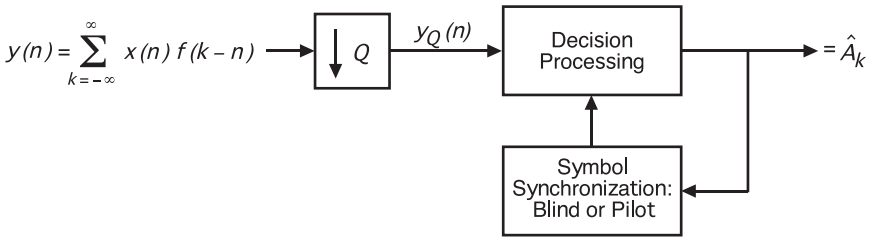


Fig. 6-76. Discrete-time symbol synchronization and decision processing.

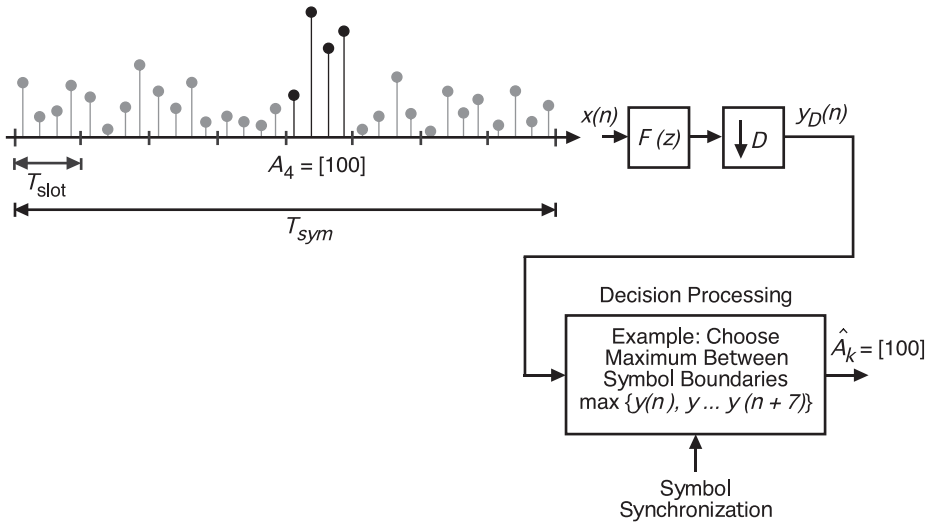


Fig. 6-77. Example symbol synchronization and detection, $Q = D$ and choose maximum detection.

not as challenging to realize in digital VLSI as slot synchronization and post-detection filtering. These former functions are generally processed at much lower rates, depending on PPM order, than slot filtering and timing error estimation and correction processing. Of course if other types of decision algorithms that require more complex processing are needed, this may not be the case.

Finally, as stated previously, many of these algorithms are not uniquely optimum but dependant on the channel characteristics, dynamics, etc. The algorithms presented very briefly in this section are meant to be representative of the signal processing structures used for accomplishing slot timing synchronization, post-detection filtering, symbol synchronization, and decision processing. One of many possible modifications is the integrated symbol and slot synchronization algorithm that may offer increased performance particularly with high-order PPM. The structures presented may be readily modified to accommodate the feedback from symbol synchronization to slot synchronization, and an example of this type structure is given in Section 6.3.8. The optimal slot and symbol synchronization and detection are highly dependant on the optical detector and other channel characteristics. This provides motivation for highly flexible, that is parameterized and reconfigurable, discrete-time processing architectures. We now introduce demodulator variations based on these "base-line" algorithms in a more global context and identify significant design choices along with specific challenges to their realization in discrete-time or digital VLSI.

6.3.4 Discrete-Time Demodulator Variations

Figure 6-78 illustrates a discrete-time demodulator for PPM type modulation. Based on the definition of the all-discrete-time demodulator developed in this work, Fig. 6-78 is not strictly an all-discrete-time

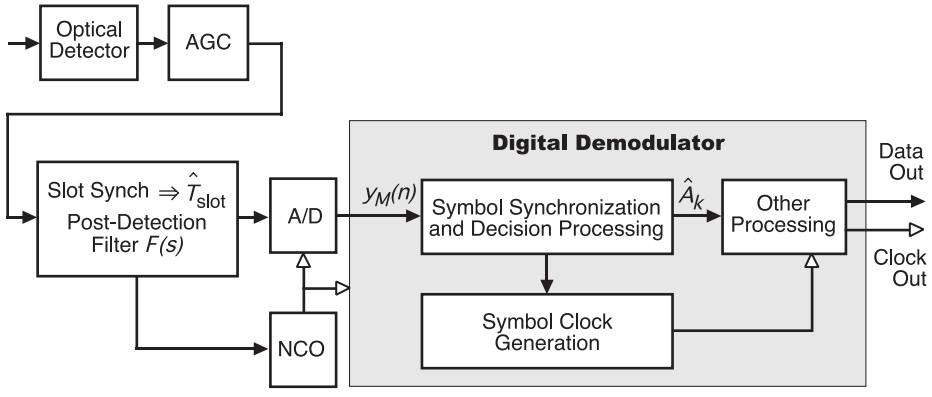


Fig. 6-78. Hybrid analog and discrete-time PPM receiver.

demodulator but a hybrid demodulator, as the slot-timing recovery is performed with continuous time processing. Figure 6-79 is a variation on this same theme but incorporates more discrete-time processing. Figure 6-80 illustrates an all-digital demodulator for a communications system with time-varying slot phase and period and may represent a reasonable design choice for some systems. For example, this demodulator may provide adequate performance in systems with many more samples per slot than required by the Nyquist rate. Nyquist rate sampling in such a system implies that some additional discrete-time signal processing is necessary to synthesize continuous-time phase adjustments of the A/D converter in order to approximate performance of the system in Fig. 6-79.

For very high sample rates, ranging applications where one desires an ultra-stable clock source, a large dynamic range in data rate, a large dynamic range of input bandwidth, or other flexibility (high parameterization or reconfigurability), the receiver in Fig. 6-80 is the desired implementation [106]. At very high sample rates (>>1 GHz) adjusting the A/D oscillator phase to correct for synchronization errors as in Figs. 6-78 and 6-79 may be very difficult. The analog circuits necessary to accomplish this function are also generally very limited in dynamic range of bandwidths, signal amplitude variations, etc. The architecture in Fig. 6-80 is the base line all-discrete-time demodulator architecture which we will further develop to mitigate losses in slot synchronization performance, post-detection filtering, and decision processing due to a fixed-sample rate near Nyquist rate sampling. In addition to operate at very high sampling rates (>>1 GHz) parallel processing is required given the current state-of-the art in commercial-off-the-shelf CMOS VLSI circuits.

Finally, the three receivers in Figs. 6-78 through 6-80 are greatly simplified versions of actual receivers and represent the variations of primary interest from a top level architecture development. There are many additional variations, functions (many diagnostic in nature) necessary in an operational receiver. However, Fig. 6-80 represents the core of the most challenging discrete-time processing that must be further developed to operate with Nyquist rate sampling and with processing rates 10–20 times slower than the required A/D converter rates for sampling broadband PPM (>500 MHz).

6.3.5 Discrete-Time Demodulator with Time-Varying Post-Detection Filter

Now we derive a serial system such as that illustrated in Fig. 6-80 with a nominal sample rate $D(1/T_{slot})$ such that with Nyquist or near-Nyquist sampling the system has the performance of a system employing a sample rate many times the Nyquist sample rate. That is to say, the system has a sample rate $D(1/T_{slot})$ that operates with identical or nearly identical performance to the system with sample rate $DN(1/T_{slot})$, where N is some large positive

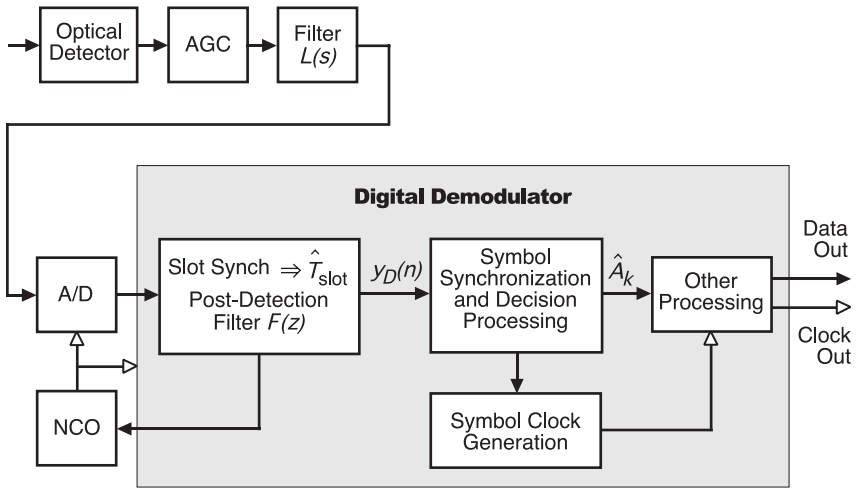


Fig. 6-79. Discrete-time demodulator with feedback to A/D.

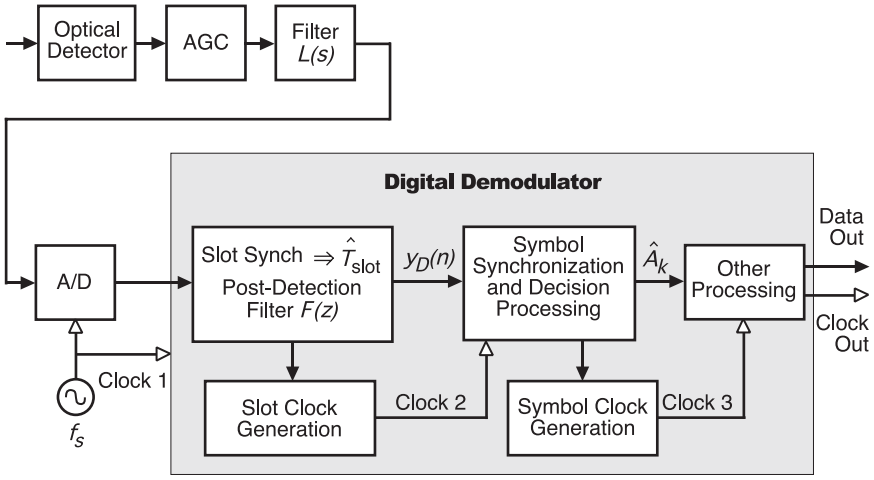


Fig. 6-80. All discrete-time demodulator.

integer (for example >100). Here the assumption is made that the received pulse is full-response and band-limited (which is impossible in theory but can be approximated in practice). For the scenario such that the signal is band-limited but not full-response [101] the system developed will not be identical but may be designed to be nearly equivalent. Several properties of multirate systems relevant to this development are illustrated in Figs. 6-81(a) through 6-81(e) [111].

Figure 6-81(a) illustrates some discrete-time sequence $x(n)$ convolved with another discrete-time sequence $h(n)$. Figure 6-81(b) illustrates the

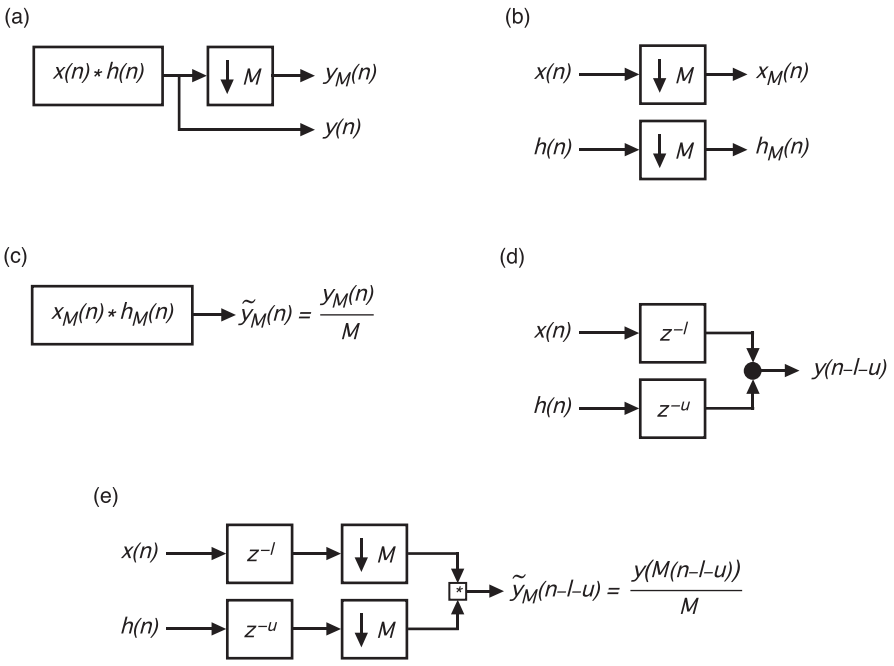


Fig. 6-81. Multirate signal processing properties: (a) downsampled convolution, (b) downsampled input, (c) convolved downsampled signals, (d) convolved time-shifted signals, and (e) convolved time-shifted and downsampled signals.

notation for the downsampled sequences. If $x(n)$ and $h(n)$ are perfectly bandlimited such that no aliasing results from this downsampling (implying infinite sequences), the result of this convolution is related to $yM(n)$ as demonstrated in Fig. 6-81(c) by a constant scale factor, *M*. If $x(n)$ and $h(n)$ are not infinite in time extent, convolving their downsampled versions as in Fig. 6-81(c) will result in a sequence that can be related to $yM(n)$ by some distortion function that incorporates the effects of aliasing caused by the downsampling. Under the assumption of this development, there is no aliasing. Figures 6-81(d) and 6-81(e) illustrate the effect that independent sample delays have on the convolved output and how such a delay relates to $y(n)$. As illustrated in Fig. 6-82, the sampled received pulse shape $hdN(n)$ can be approximated from $hd(n)$ as illustrated using the expander and interpolation filter. If we make the simplifying assumption that $h(t)$ is perfectly band-limited (with *D* samples per slot) and full-response (although theoretically not possible) and the perfect interpolation filter is used (infinite time extent) [111], then $\tilde{h}_{dN}(n) - h_{dN}(n) \approx 0$.

Now the time-variable *slot* or *post-detection filter bank* ($f_d(n,u)$) can be developed from the received pulse, $h_{dN}(n)$ (or $\tilde{h}_{dN}(n)$), as illustrated in

Fig. 6-83 [112]. The filter bank $f_d(n,u)$ may be developed from the time reversed $h_{dN}(n)$ or another filter $v(n)$. The filter $v(n)$ for example could be an interpolation filter. Each sampled filter *u* is shifted in time by $-T_S/N$ from *u*-1 (recall the sample period of T_S and *D* samples per slot).

It is clear from this development and the properties in Figs. 6-81(c) through 6-81(e) that the output of Fig. 6-84(b) is equivalent to the output of Fig. 6-81(a) divided by the scale factor *N* [112]. Assuming the delays *l* and *u* are correct, such that the output is the desired sample (possibly with residual fractional sampling offset, $\Delta t_e \le T_S/2N$), any sample that can be obtained with the $DN(1/T_{slot})$ sampled system can now be obtained from the $DN(1/T_{slot})$ sampled system due to the bank of time-shifted filters.

These properties are combined to generate the architecture in Fig. 6-85 that utilizes primarily synchronous processing and can operate with Nyquist rate sampling while approximating the performance of a system with a much higher sample rate (*N* times more samples per slot).

There are many other aspects of the time-varying post-detection filter and slot-synchronizer design that require analysis and further development on the

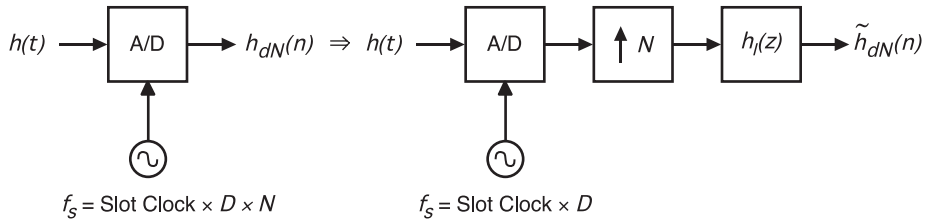


Fig. 6-82. Approximating high-sample-rate system with low-sample-rate system.

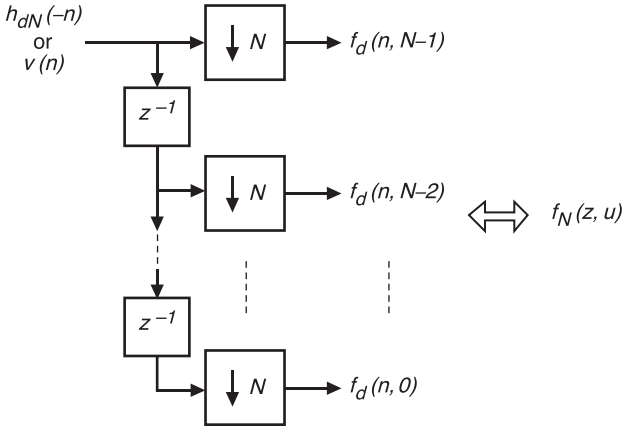


Fig. 6-83. Filter bank generation.

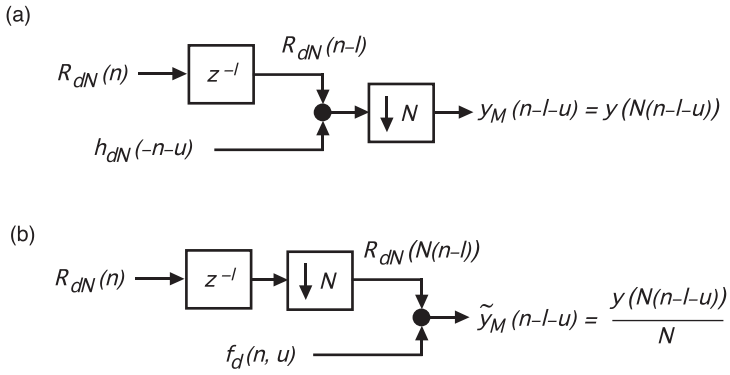


Fig. 6-84. Multirate signal processing properties: (a) delay, convolution, and downsample, and (b) delay, downsample, and convolution.

path to practical implementation and to determine performance with realistic channel conditions. There is one issue in particular that requires further careful consideration in both the architecture development and hardware design and implementation. Recall that the reason that post-detection filter $F(z)$ needs to vary with time is because the transmitter and receiver are fundamentally asynchronous, and the actual received data rate will generally differ and vary with time from the expected data rate chosen to set the A/D and system clocks. Consider the design of Fig. 6-85, as part of the time-varying detection filter and slot synchronization algorithm the true slot clock is recovered. The true symbol clock is also recovered and is synchronous to the slot clock; however, both of these clocks are asynchronous to the A/D or system clock. Thus the implementation of the architecture of Fig. 6-85 requires high speed digital design with asynchronous clocks. The asynchronous nature of the processing also creates additional design challenges when a parallel- or vector-processing architecture is developed. These challenges are addressed in detail in Section 6.3.6.

Finally, of course, it is not possible to achieve perfectly band-limited pulse shapes that are also full-response or to implement perfect (infinite in time extent) interpolation filters, all of which was assumed for simplicity in the derivation of Fig. 6-85. However, these can be approximated very closely in reality, and generally such a system based on these principles results in negligible implementation losses if designed and implemented properly.

6.3.6 Parallel Discrete-Time Demodulator Architectures

Due to the very high sampling rates required by broadband optical pulses processing rate reductions of 10–20 are typically required given the current state-of-the art if the targeted processing implementation is commercially

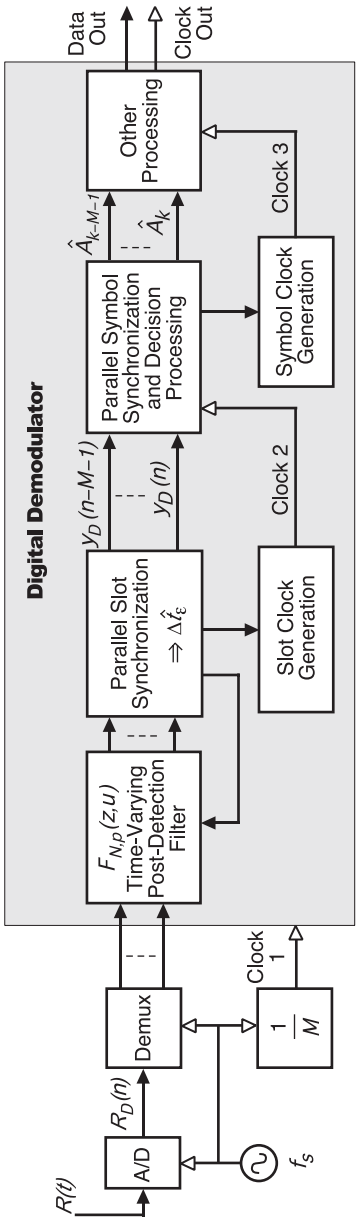


Fig. 6-85. Nyquist-rate discrete-time demodulator.

available CMOS (FPGAs for example). Some demodulator functions are trivial to parallelize, and others are not. The process of parallelization generally increases complexity; depending on the methods used, the increase in complexity may be linear or nonlinear [105,106,111,112]. There are numerous methods, many based on frequency domain or dual time-frequency domain processing, for parallelizing the convolution/correlation processing that reduce the complexity or transistor count of parallel implementations [106,111,112, 131–148]. Figure 6-86 illustrates the most straightforward time-domain method of parallelizing the FIR filter structure illustrated in Fig. 6-87.

There is a much more elegant way of representing parallel filters and filter banks that is based on multirate notation. The vector downsample operation is given in Fig. 6-88.

The parallel filter can then be represented by the structures given in Fig. 6-89, where the demultiplexer operation represents the delay and vector downsample chain (ordered serial-to-parallel conversion).

Using this notation, the parallel version of the demodulator given in Fig. 6-85 is given in Fig. 6-90. A core processing element of the demodulator is the time-varying parallel filter bank, $F_{N,p}(z,u)$. This is also the processing element that may dominate the complexity (transistor) count of the VLSI implementation. There are numerous potential simplifications to the filter structure in Fig. 6-89. For example, if the impulse response is symmetric, the multipliers and adders may be reduced by a factor of 2. In addition, as indicated earlier, there are other options for performing the parallel filtering, and many have greatly reduced multiplier requirements depending on the filter type and order.

As mentioned earlier, one of the additional challenges of realizing the architecture in Fig. 6-85 or Fig. 6-90 in digital VLSI is appropriate asynchronous clock management. Additionally, the parallel implementation is much more straightforward to realize if the number of parallel paths (vector length) remains constant throughout a given processing operation (i.e., convolution). Obviously this creates challenges as each slot clock and symbol clock represents a fixed-vector length of data samples and both of these are asynchronous with the A/D (system) clock. From a high level, the system of Fig. 6-90 is primarily composed of synchronous clocks and fixed-vector length

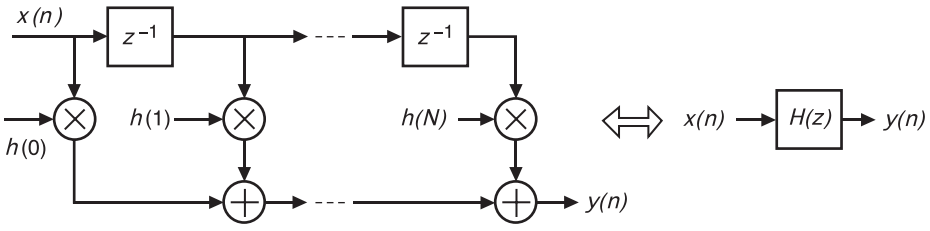


Fig. 6-86. FIR filter structure.

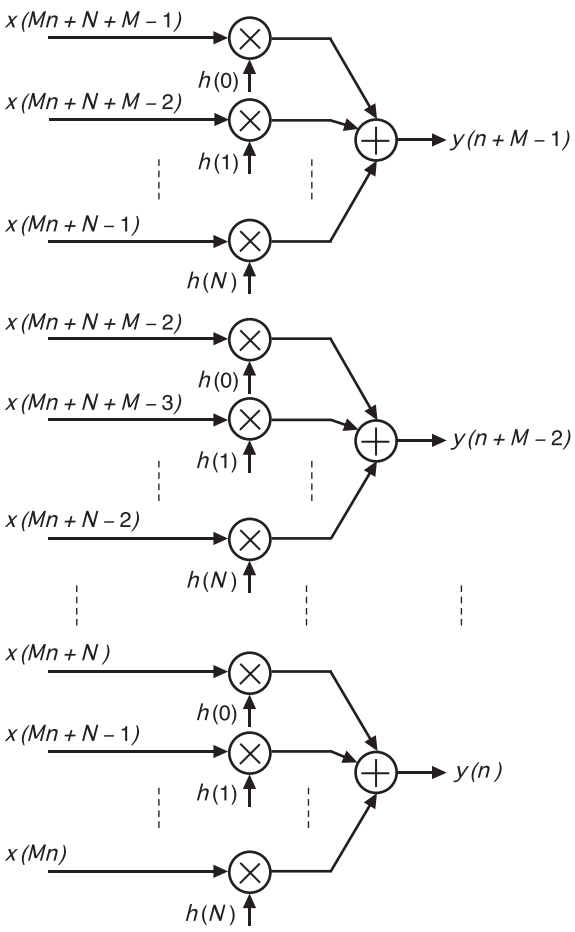


Fig. 6-87. Parallel FIR filter structure.

parallel processing. However the data symbols must be clocked out at the true symbol rate that is asynchronous to the A/D clock and the 1/*M*th rate system clock. One method of re-clocking the vector of symbols at the recovered data rate is with the asynchronous first-in first-out (FIFO) circuit, and there are many ways to realize this circuit. More details of the slot and symbol clock generation and the conceptual asynchronous FIFO type circuits are presented in the next section.

6.3.7 Asynchronous Discrete-Time Processing

The input signal has a dynamic nature due to the Doppler in the channel and the absence of synchronization between the oscillators in the transmitter and receiver. As a result of this dynamic signal the error $\Delta\hat{t}_{\varepsilon}$ varies with time,

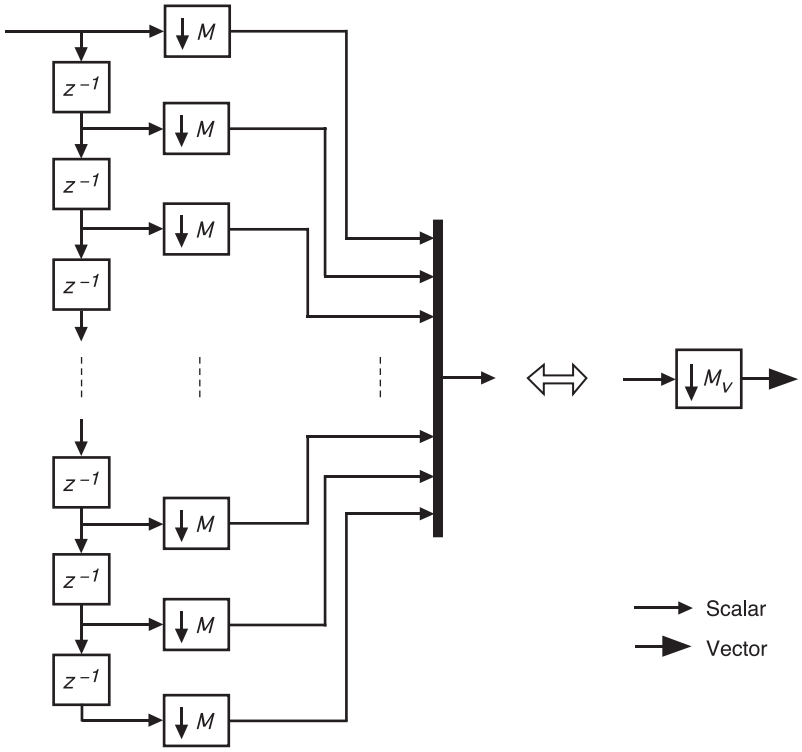


Fig. 6-88. Vector downsample operation.

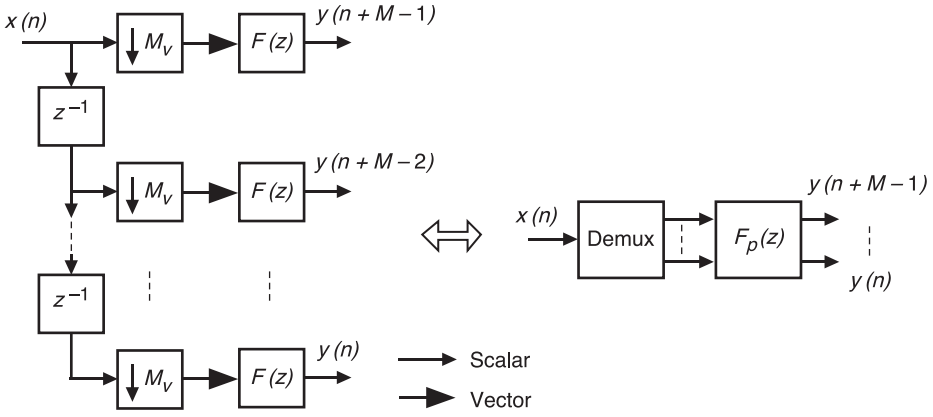


Fig. 6-89. Parallel filter bank representations.

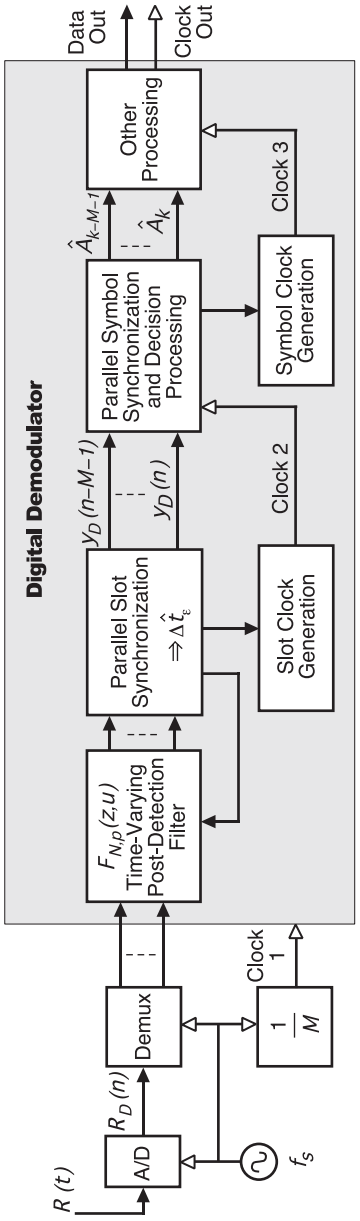


Fig. 6-90. Parallel discrete-time demodulator.

and given the fixed-sample clock A/D of Fig. 6-90, this results in a discrete-time system with asynchronous sample rates and a digital implementation with asynchronous clocks. The simplified error signal model may be represented as

$$\Delta t_\varepsilon(t) = T_{slot}(t) - T'_{slot}(t) + \tau \tag{6.3-13}$$

where T'_{slot} is slot period due to Doppler, and τ is a fixed time offset. The slot period with Doppler is given by,

$$T'_{slot} = T_{slot} + T_{slot}\left(\frac{v}{c}\right) \tag{6.3-14}$$

where *v* is the relative velocity of the transmitter relative to the receiver and *c* is the speed of light. Note that the relative velocity may be positive or negative. However, for reasons that will be demonstrated, significant simplifications result if the received sample clock frequency is set such that,

$$1/T'_{slot} \le \frac{f_{S\_\text{nominal}} + r_1}{D} = \frac{f_S}{D} \tag{6.3-15}$$

The sample frequency is set such that the nominal sample rate assuming no Doppler, f_{S_nominal} is increased by the real constant r_1 such that the true sample frequency f_S divided by the designed number of samples per symbol is greater than the maximum slot frequency the demodulator processes. The slot synchronization algorithm creates a new estimate at the nominal symbol rate,

$$\hat{\Delta} t_\varepsilon\left(nT_{sym}\right) \approx T_{slot}\left(nT_{sym}\right) - T'_{slot}\left(nT_{sym}\right) + \tau + \eta_0(n) \tag{6.3-16}$$

Here $\eta_0(n)$ is some noise on the estimate. Successive timing error estimates must be accumulated, and this accumulation is represented as

$$\sum_n \Delta \hat{t}_\varepsilon\left(nT_{sym}\right) \approx \sum_n T_{slot}\left(nT_{sym}\right) - \sum_n T'_{slot}\left(nT_{sym}\right) + \tau + \sum_n \eta_0(n) \tag{6.3-17}$$

An accumulator such as this exists in the filter given in Figure 6-75. As stated previously, the error estimate is an approximation to the difference between the nominal slot rate and the actual slot rate. This estimate is not generally equal to the actual error as many signal components and signal distortions are not included in this simple model. These include multiple jitter sources, channel and amplifier noise sources, imperfect filters, etc. Much analysis must be performed to determine actual performance of the estimator as a function of integration time or filtering and the statistical properties of the numerous random variables in the communications system. As shown in [99],

this filtering may be accomplished with a closed-loop feedback system yielding very desirable Doppler tracking capabilities. Here we continue to address the implication of asynchronous processing on the overall architecture and ignore random noise and distortions.

The accumulated error may become arbitrarily large as n becomes arbitrarily large. For practical purposes the error must be accumulated in a cyclic fashion, that is modulo some multiple of $-T_S$. Recall the system clock rate of the serial demodulator in Fig. 6-85 is $1/T_S$. It is now demonstrated that it is desirable to alter the slot clock at time increments given by the following,

$$\text{round}\left\{\sum_n \Delta\hat{t}_\varepsilon\left(nT_{sym}\right)\right\} = -\frac{L}{N}T_S \tag{6.3-18}$$

where $L = kN$ and k is some integer greater than or equal to 1. For the demodulator in Figure 6-85, there are clocks with periods T_S, T_{slot}, and T_{sym} as illustrated in Fig. 6-91.

Recall the slot period is D times the sample period, $T_{slot} = DT_S$. Here $L/N = 1$, $L/N = D$, and $L/N + DP$ in the accumulated error estimate represents the time of one sample clock, slot clock, or symbol clock cycle, respectively. The timing error in the signal is partially corrected with the filter bank $F_{N,p}(z,u)$; the error is quantized to one of the time-shifted versions of the detection or interpolation filter. This error correction must be made modulo some time LT_S/N. Otherwise arbitrary filter lengths and complexity are required if $1/T'_{slot} < 1/T_{slot}$ and the receiver runs for an arbitrarily long time. When the accumulated error time reaches its minimum value, it must *wrap* or *roll-over*; otherwise, it will become arbitrarily large. Figure 6-92 illustrates the accumulation of the error estimate $\Delta\hat{t}_\varepsilon$ modulo $L/N\,T_S$.

While the filter bank of Fig. 6-83 indicates a time-delay of one sample period $(N(T_S/N))$ from the first filter $(u = 0)$ to the last $(u = N-1)$ it may be

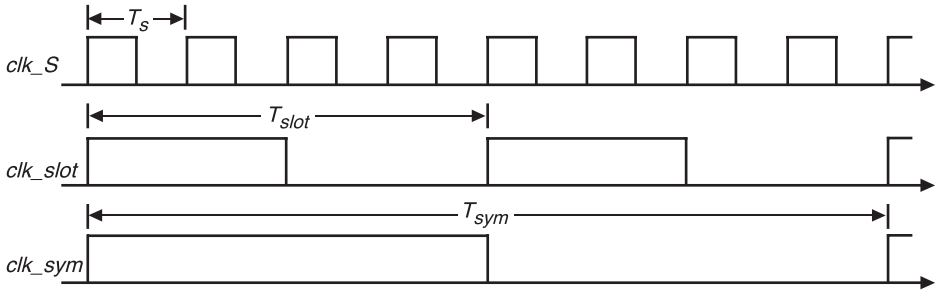


Fig. 6-91. Sample, slot, and symbol clocks.

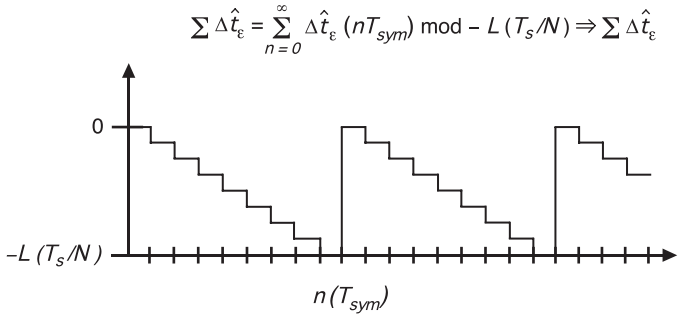


Fig. 6-92. Time delay.

designed for any delay that is a ratio of *L*/*N* times (T_S). It will be demonstrated that for certain parallel processing architectures specific delays result in much more simple overall implementations. To gain more insight into these issues consider the example in Fig. 6-93.

- Illustration (a) is a received 2-PPM signal with 4 samples per slot $(D = 4)$ and no dynamics.
- Illustration (b) is a conceptual depiction of the effects of dynamics on the signal where the amount of Doppler is greatly exaggerated to illustrate key concepts. In reality, the timing error in illustration 2 would typically be thousands or millions of times smaller per slot period.
- Illustration (c) demonstrates how the error is used to change *u*, thus choosing a filter with different subsample timing delay. For this example, assume that the slot filter is a perfect interpolation filter, so incrementing *u* creates a time-quantized delay, quantized to $T_S/2$, in the signal given in illustration two of the figure. Incrementing *u* by one is equivalent to incrementing one-half of one sample period. The maximum timing error due to time-quantization with this filter bank is $T_S/(2N) = T_S/4$. The total delay in the filter bank is chosen to be $-DT_S - T_S$, which is one and one-quarter slot periods (recall there are $D = 4$ samples per slot). To create a process that results in repetition of an entire symbol when the filter rolls-over or wraps, the filter bank design method of Figure 6-83 had to be extended by $-4T_S$.
- In illustration (d), when the timing error wraps, it instantaneously moves the output signal back in time by $4f_S$ (one slot period). The wrapping of the filter bank index *u* results in samples that have already been output of the filter to be output again. These samples must be discarded, and the slot clock period associated with them must also be discarded. The period of the clock following the filter wrapping is instantaneously doubled, and the slot clock returns to its nominal period (T_{slot}). This

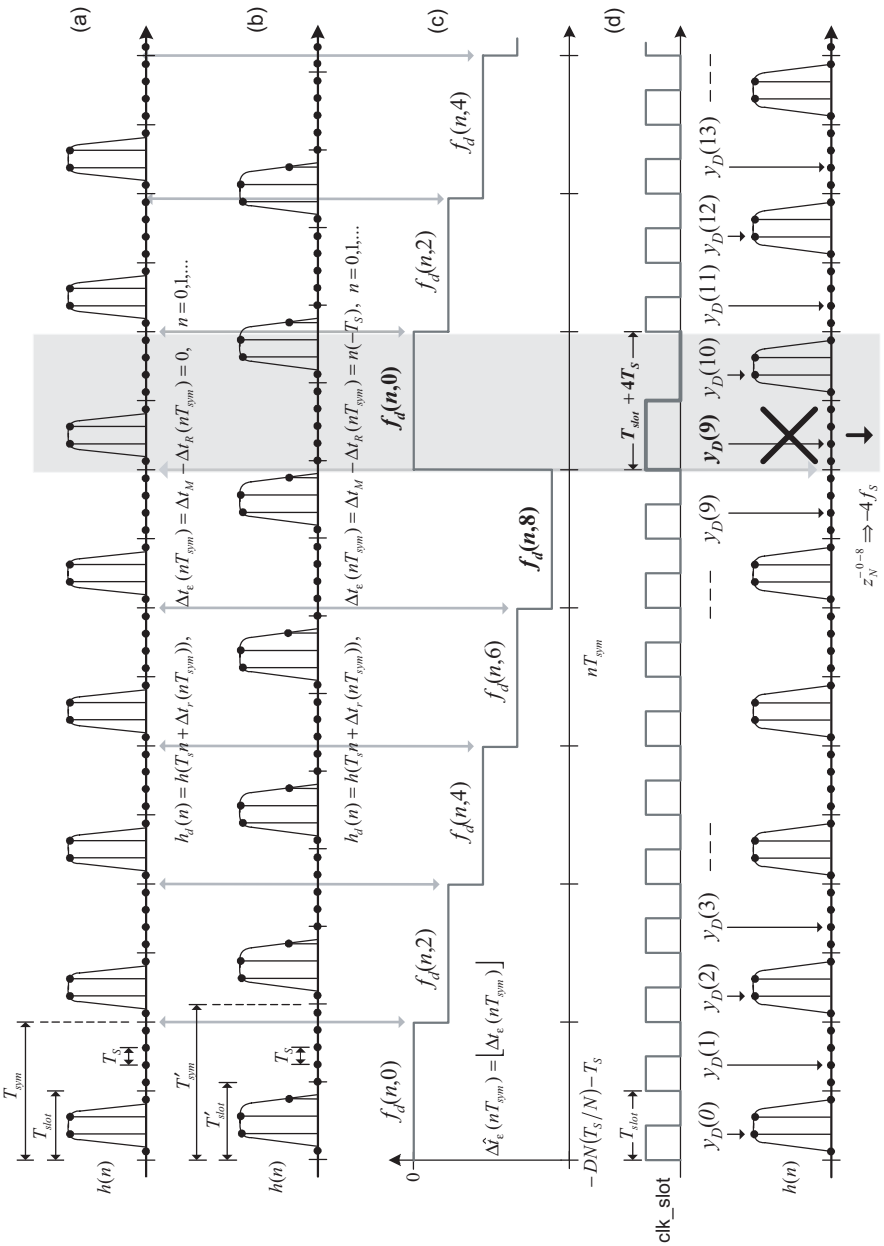


Fig. 6-93. Conceptual illustration of time-varying phase tracking and multirate processing.

processing results in a slot clock with an average frequency that is equal to the recovered slot rate. In this way the timing estimates are made and accumulated to adjust the clock period, which is readily accomplished in digital circuitry. There is obviously a need for a roll-over detector to indicate when the clock period should be adjusted; this is also readily accomplished with digital circuits. This roll-over detector may be designed to further average out or remove with thresholding the variations of the accumulated phase error estimates due to noise and other distortions on the input signal, and thereby avoid multiple roll-overs due to jitter on the error signal.

Although the memory length of the filter in Fig. 6-93 was chosen to be seemingly arbitrary to illustrate specific concepts, the number of samples that are repeated and then discarded after filter index roll-over is a significant consideration in the design of the filter bank. It will be shown that it is very desirable to create the filter bank to rollover on slot boundaries and this requires a modification to the earlier filter bank design method given in Fig. 6-83; consider Fig. 6-94.

In Figure 6-83, the prototype filter, $h_{dN}(n)$ or $v_{dN}(n)$ is zero padded, the filter is appended with a number of zero samples. The number of filters in the bank is now L, setting L equal to an integer number of slots yields:

$$\frac{L}{N}T_S = \frac{RDN}{N}T_S = RDT_S \quad \Rightarrow \quad L = RDN \tag{6.3-19}$$

In this case the total time delay from the 0 indexed filter to the $(L-1)$ indexed filter is the number of samples per slot (D) times the number of slots (R) times the sample period, and the total number of filters is equal to RDN. Figure 6-95 illustrates a conceptual filter bank with $R=1$.

The result is the ability to correct larger timing error before index roll-over. This increased filter memory however results in more complex filter banks, but

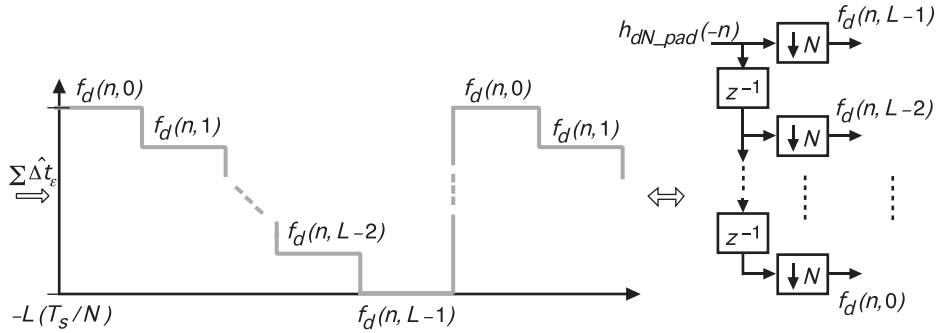


Fig. 6-94. Filter bank for correcting *R* slot times before rollover: prototype filter length $L = R \times D \times N$.

it allows control over the jump in time and corresponding number of samples discarded when the index to the filter bank rolls over. There is an equivalent multirate structure for the filter bank in Fig 6-94 that results in significant simplifications as illustrated in Fig 6-96.

The details of this simplified multirate filter bank are not presented here, but the structure incorporates a variable delay block before the filter bank. The delay is modulo $k-1$, where $L = kN$. With this structure the sub-sample errors are corrected by the filter bank while timing corrections that are integer multiples of a sample period are corrected by the variable delay. In this structure, the complexity of the filter bank is similar to that of Fig. 6-83, but it is increased by the addition of the variable delay. The error estimate $\Delta\hat{t}_\varepsilon$ is decomposed into two components, one quantized to integer sample time error that is used to control the variable delay and another component quantized to sub-sample timing error used to control the filter bank. Note that if $k = D$, then the filter bank will "roll-over" on slot boundaries.

The motivation for the filter index to roll-over on slot boundaries is provided in Fig. 6-97. Consider the filter bank of the parallel architecture of Fig. 6-90, and note that after being processed by the filter bank, the signal is input to the slot synchronization algorithm. The current synchronization algorithms introduced earlier cannot have discontinuities in their input; if the signal is instantaneously (in the discrete-time domain this is effectively possible) shifted in time, it must be done on a slot boundary or some modified slot synchronization algorithm normalizing for the time shifts in the input must be developed. Note that the phase (timing error) wrapping can only occur across the vector boundary as this is when error estimates are updated in the parallel system. Error estimates are not updated within a vector block, only across vector boundaries, and when the roll-over occurs it is known precisely

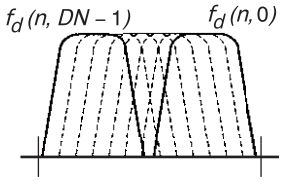


Fig. 6-95. Filter bank.

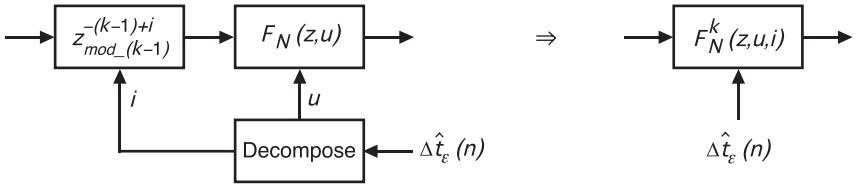


Fig. 6-96. Conceptual simplified filter bank.

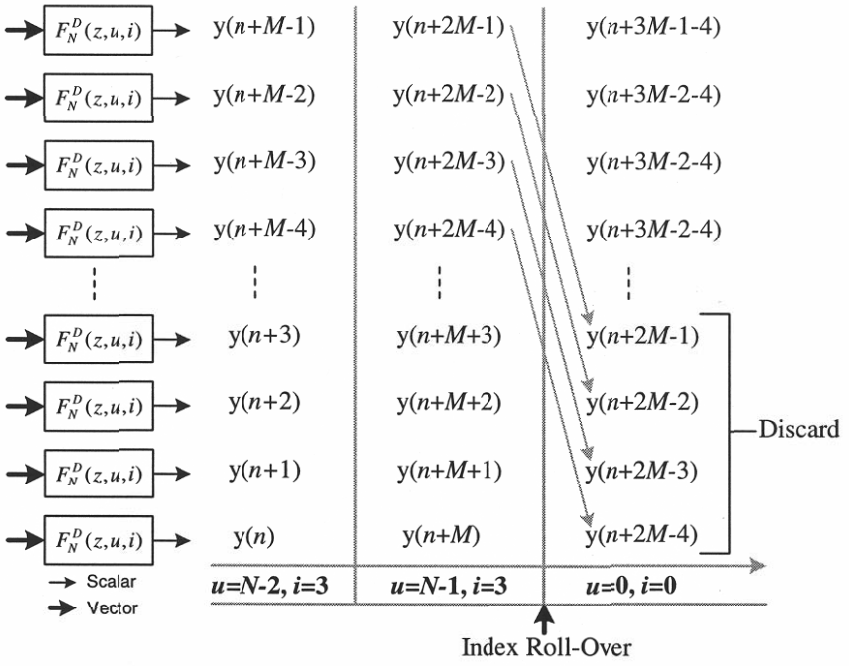


Fig. 6-97. Roll-over on slot boundaries of parallel/vector system.

which data samples are repeated due to the roll-over detector circuit, refer to Fig. 6-97.

There is another model of the time-varying filter bank that is useful for building further insight useful for formal analysis beyond the scope of this document. The filter bank may be modeled as a numerically controlled oscillator (NCO) [113,1,2,3]. The variables *u* and *i* are updated every symbol and may be thought of as the error signal to the NCO. The phase of the NCO is changed by the error $\Delta\hat{t}_R(nT_{sym})$, and frequency is determined by the derivative of this estimate.

Figure 6-98 summarizes the results of this section from a high level. The asynchronous FIFO circuit is a critical subsystem of the demodulator. The functions of this circuit are to order the parallel vector appropriately for the parallel symbol synchronization and detection in the presence of the filter bank index "roll-over" and re-clock the data out at the appropriate recovered rate. The details of the asynchronous FIFO (such as memory depth, overflow and underflow flags, and many other elements of this circuit) are not presented here. There are many possible implementations to accomplish the functions of this circuit.

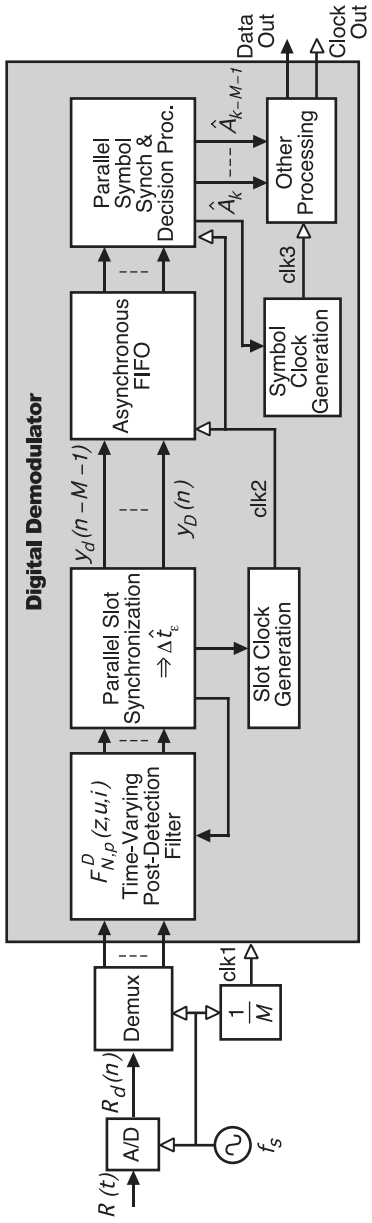


Fig. 6-98. Parallel digital demodulator.

In conclusion, the elements of the architecture presented in this section are largely generic. They accommodate a time-variable post-detection filter for which the purpose is two fold: filtering and the critical operation of correcting sample and sub-sample timing errors. This filter may be utilized with any number of slot-timing error estimation algorithms and in a closed loop system as in Fig. 6-73 through Fig. 6-75 and may approximate the high sample per slot A/D with continuous-time phase adjustments. It was inferred in this development and stated here explicitly that in addition to the challenges of designing the post-detection processing and slot-synchronization algorithms from a communications theoretical perspective, designing architectures for realizing these function in discrete-time at very high rates with Nyquist or near-Nyquist rate sampling entails many challenges. Overcoming these challenges requires developing and synthesizing discrete-time signal processing methods and algorithms for realizing the fundamental algorithms while overcoming the challenges of asynchronous and parallel digital design and implementation given the current state-of-the art in digital VLSI circuits. The discrete-time structures for realizing decision processing and symbol synchronization also pose many challenges; however, these challenges are generally more readily overcome than those for post-detection filtering and slot-synchronization, namely sub-sample and asynchronous real-time processing at the maximum sample rate (A/D rate), as such these architectures were not presented here in detail. With this as context the next section develops more complete parallel discrete-time demodulator architectures.

6.3.8 Parallel Discrete-Time Demodulator Architectures

6.3.8.1 Simple Example Architecture. The purpose of this example is to illustrate further key considerations when synthesizing concepts developed earlier. A greatly simplified list of hypothetical requirements for a communication system and hardware limitations is given as follows:

1) Slot period, T_{slot} = 2 ns (500 MHz bandwidth)
2) PPM pulse period = 1 ns (1 GHz bandwidth)
3) 4-64 PPM modulation

Hardware limitations:

1) VLSI implementation in a FPGA with a maximum clock rate of 250 MHz
2) 4 giga-samples per second A/D converter

Even from this short-list of requirements numerous design parameters may be determined. First the Nyquist sampling rate of the PPM pulse bandwidth, and hence the system, must be >1 GHz × 2 = 2 GHz. It should be clear given the earlier discussions that the architectures development assume $D = 2k$, where *k* is some integer. Although not stated explicitly this assumption leads to

multiple simplifications in the parallel architecture development and facilitates many further simplifications in both the architecture and in the hardware design beyond the scope of this work. There is also additional motivation for sampling above the Nyquist rate, design margin. Given that the A/D can accommodate a 4-GHz sample rate, D is set equal to 4. Given that the VLSI device implementing the demodulator architecture operates at 1/16th the rate of the A/D converter, the rate reduction or number of parallel paths M is set equal to 16.

Assume that some interpolation filter is used as the prototype filter in developing the filter bank $F_N^4(z,u,i)$. Determining N, the number of time-shifted filters in the filter bank, is generally not trivial. From a hardware complexity perspective N should be made as small as possible to minimize complexity and transistor count. Yet, from a pure performance perspective, N should be made as large as possible to minimize residual sampling offset and synthesize as closely as possible the phase/time adjustments of the continuous-time system. Assume $N=128$; the residual sampling offset is then $0.25\,\text{ns}/(2\times128)\cong0.0009766$ ns or approximately 1 picosecond (ps). This number may or may not be a good design choice depending on many other characteristics in the system both in the transmitter and in other subsystems of the receiver. For example, if the A/D oscillator phase jitter is significantly greater than 1 ps, it may not make sense to have the resolution of $N=128$. Another perspective to consider is the performance, both tracking and acquisition, of the closed-loop tracking system of Figs. 6-73 through 6-75 (or other slot synchronization algorithms) with time-quantized error correction. Much previous work exists in analyzing these and other issues; furthermore, a loop similar to that proposed has been analyzed, tested, and validated in hardware with $N=128$ [106]. However, detailed analysis of the performance of the type of loop with the optical channel has not been completed to date. Figure 6-99 is a block diagram of a parallel architecture satisfying the simplified requirements that includes a structure for a non-time-varying pulse-equalizer $B_1(z)$.

Note that details of parallel symbol synchronization and decision processing are not given here, but are generally significantly less complex to realize than post-detection filtering and parallel slot-synchronization. The challenging aspect of these former functions lies in the fact that they are parameterized on PPM order, but with the flexibility of a digital implementation, such challenges are readily overcome. It should be noted that the number of parallel paths and the number of samples per slot in the slot synchronization processing and filter $F_N^D(z,u,i)$ is readily scalable by powers of two, and the amount of parallelization in the functions of symbol synchronization and decision processing is largely independent of these. Note

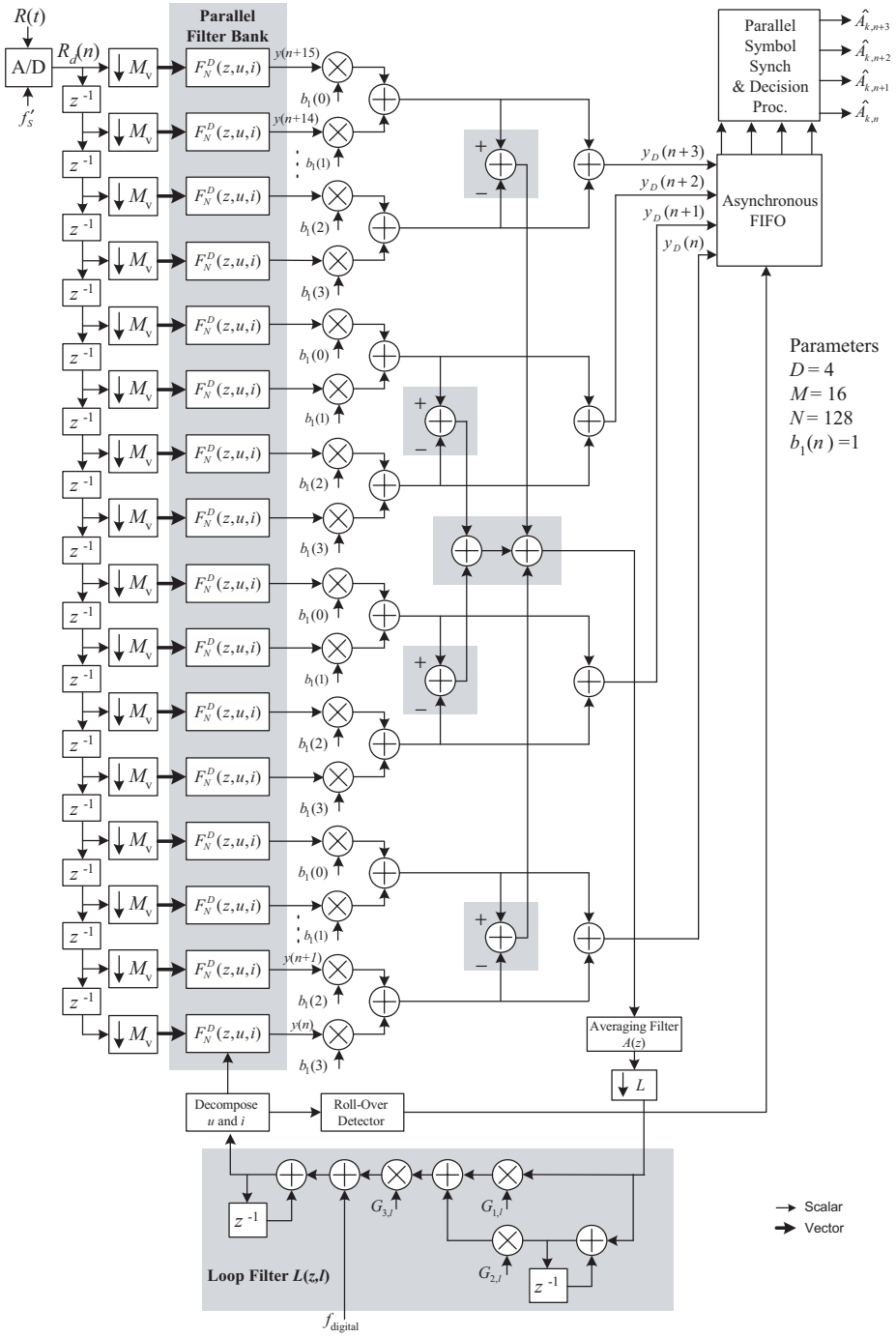


Fig. 6-99. 1/16th rate parallel architecture.

that, due to the large overlap of vectors input to the vector downsample and parallel filter banks of Fig. 6-99, significant simplifications are possible in such parallel architectures. However, derivations of such simplifications are beyond the scope of this work.

As stated earlier, determining the performance analytically or pseudo-analytically of an architecture such as that in Fig. 6-99 for most realistic communications channels is a significant undertaking. However, given that loss-less or perfect reconstruction parallelization techniques were used in its development, there is no need to determine the performance of this specific architecture. Figure 6-100 presents a serial equivalent architecture for which performance can be determined or estimated more readily although this determination is still non-trivial with realistic channel models. This serial architecture is equivalent from a communications systems performance perspective not from a processing rate or complexity perspective that has been a focus of this work. The delay Q_1 is the total delay added by the parallel architecture in the feedback. For example, to model the actual delay in the implementation Q_1 must include the pipeline delays in all multipliers and adders in the feedback path.

6.3.8.2 Performance with a Simple Optical Channel Model. A basic model of the optical PPM symbol detection uses photon-counting detection in the presence of background noise [97,98]. Thermal noise is negligible when using photon counting. This analysis assumes ideal slot and symbol synchronization. The number of photons in a coherent-state optical field generated by lasers is Poisson-distributed, i.e., the probability of detecting *k* photons, $k \geq 0$, is given by the expression

$$P(k) = \frac{K_s^k}{k!} e^{-K_s} \tag{6.3-20}$$

K_s is the average number of signal photons per slot. The probability of correct detection for *M* order PPM is

$$P_M(C) = 1 - \frac{M-1}{M} e^{-K_s} \tag{6.3-21}$$

The probability of symbol error is related to this quantity as

$$P_M(SE) = 1 - P_M(C) = \frac{M-1}{M} e^{-K_s} \tag{6.3-22}$$

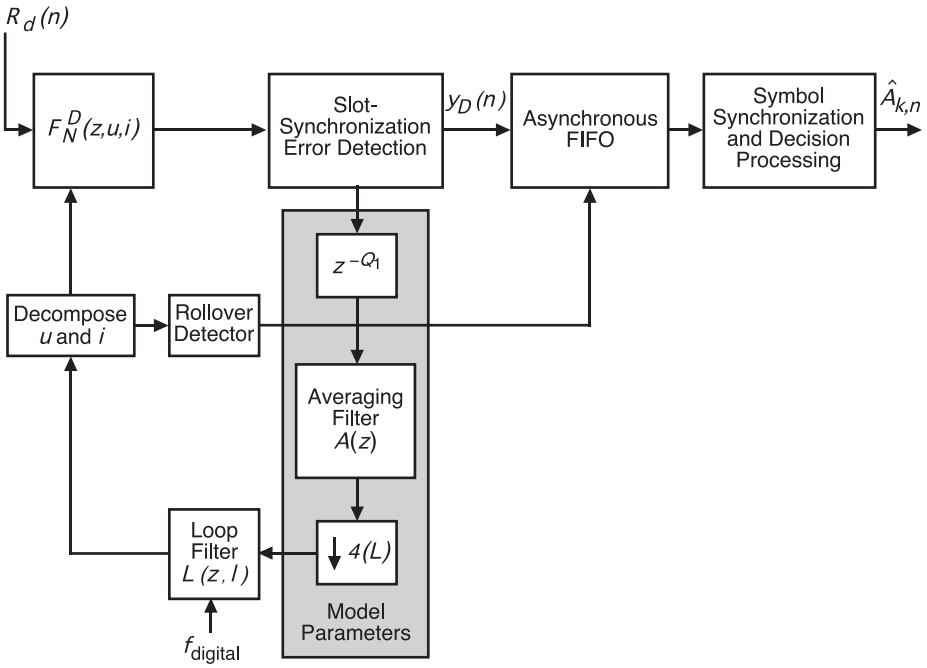


Fig. 6-100. Serial model of the parallel architecture.

When significant amounts of background light enter the receiver along with the signal, it is possible for the receiver to make an error, even if one or more signal photons are detected, because a noise slot may occasionally produce a greater count than the signal slot. Let the average number of background photons be K_b. The probability of correct detection for maximum-likelihood detection of PPM symbols in the presence of noise has been derived in [97] and shown to be:

$$P_M(C) = \left\{ \sum_{r=0}^{M-1} \left(\frac{1}{r+1} \right) \binom{M-1}{r} \sum_{k=1}^{\infty} \frac{(K_s + K_b)^k}{k!} e^{-(K_s+K_b)} \left[\frac{(K_b)^k}{k!} e^{-K_b} \right]^r \right.$$
$$\left. \times \left[\sum_{j=0}^{k-1} \frac{(K_b)^j}{j!} e^{-K_b} \right]^{M-1-r} \right\}$$
$$+ M^{-1} e^{-(K_s + MK_b)}. \tag{6.3-23}$$

The corresponding error probability is given by

$$P_M(E) = 1 - P_M(C) \tag{6.3-24}$$

Software simulations of the system in Fig. 6-100, with 4 PPM, $N = 128$, $D = 4$ samples/slot, and a normalized loop filter bandwidth of approximately 0.005 were performed using Matlab Simulink. Figure 6-101 is a basic block diagram of the simulation model. The simulation assumed a fixed slot phase offset, thus requiring the slot and symbol synchronization loops to acquire and then track.

Using the mathematical model described in Eq. (6.3-1), the transmitted pulse, $p(t)$, was a perfect square pulse that occupied 50 percent of the slot period. No intensity variations other than those from the photo-detector were modeled. The optical channel models only a simplified photon counting type (possibly PMT) photo-detector as described above. The output of the channel is a Poisson distribution given by,

$$y(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \text{ with parameter } \lambda = \begin{cases} K_s + K_b & \text{signal slot} \\ K_b & \text{no signal slot} \end{cases} \tag{6.3-25}$$

The software utilized a Poisson channel model, and the background mean photon count (K_b) remained constant at 1. The signal mean photon count vs. signal error rate (SER) simulation results are plotted in Fig. 6-102. These simulation results vary little from the theoretically predicted results given by Eq. (6.3-24); however, it should be emphasized that this is a simplified optical channel model.

6.3.8.3 Evolved Parallel Architectures. The architecture of Fig. 6-99 is not a unique or final architecture design for the high-rate optical PPM demodulator; it is a starting point from which to evolve. A core processing element is the time-varying post-detection filter bank. This processing or some close variation is envisioned as being a key element of any all-digital demodulator; however, numerous simplifying signal processing structures for accomplishing this processing are possible. Other signal-processing algorithms might be fundamentally improved or changed depending on the optical channel and advancing detection and estimation theoretical techniques derived for that channel. For example, consider that the optimal PPM slot synchronization signal processing algorithm has currently not been derived for the optical

Fig. 6-101. Software model block diagram.

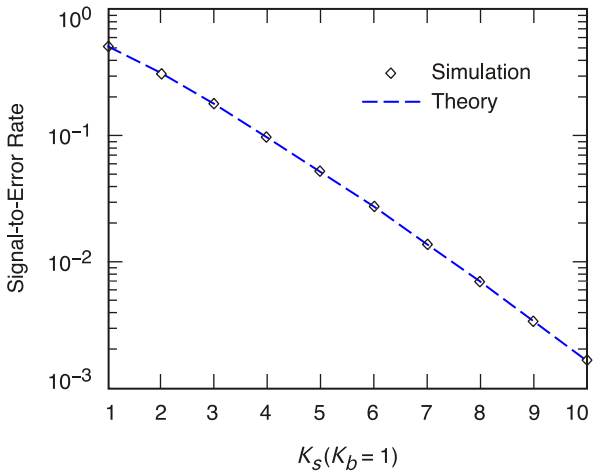


Fig. 6-102. Performance of the software model.

channel described in [99]. In addition to deriving such processing algorithms, improvements to existing synchronization algorithms might include a symbol decision-directed slot-synchronization loop for the optical channel or pulse equalization. Indeed the list of theoretical improvements or channel-specific designs for synchronization, post-detection filtering, and decision processing as well as other processing subsystems is quite long. In addition, as the state-of-the-art hardware implementation devices improve, more computationally complex signal processing may be feasibly implemented. Future algorithm improvements cannot be predicted with a high degree of accuracy. However, there is a set of architecture evolutions or improvements that are largely possible to predict. These are the signal processing functions that while not the focus of this work, are a necessary part of a ground receiver or demodulator. The demodulator architecture will evolve to contain these functions:

- Master controller (for real-time autonomous demodulator monitoring and control)
- Forward error correction preprocessing
- Power estimation and SNR estimation
- Slot synchronization and symbol synchronization lock detection
- Digital automatic gain control

Figure 6-103 illustrates an evolved parallel-demodulator architecture incorporating additional functions. Many additional functions have been added to the base-line architecture of Fig. 6-99, and the slot-synchronization loop incorporates the option of being decision directed. The base-line parallel processing and subsample timing error correction of Fig. 6-99 remains intact. Examples of other architecture improvements and alternatives are presented next. We will demonstrate advantages in having the filter bank for timing-error

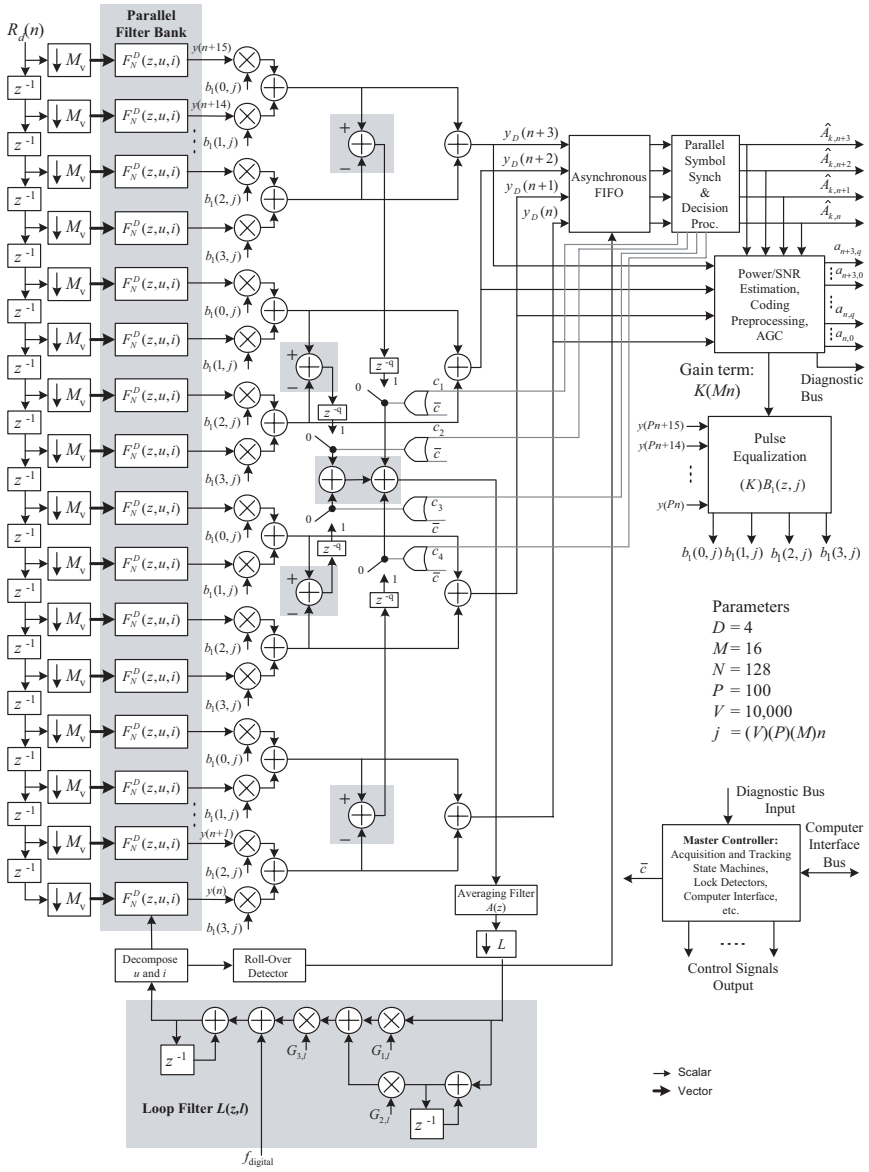


Fig. 6-103. A more comprehensive architecture; core processing of Fig. 6-101 retained.

estimation to be designed independently of the post-detection filter bank used for filtering and timing-error correction on the signal used for decision processing. The process of synchronization and post-detection filtering may be separated or deconstructed. It will become clear that there are several advantages to deconstructing the processing this way. Part of the motivation for this is that different filter-bank designs and processing rates may be used for

synchronization and post-detection filtering, and this may facilitate further advantages in the design and implementation of the synchronization filter.

Consider the architecture in Fig. 6-104; this architecture may be designed to be mathematically equivalent to the slot synchronization architecture given in Fig. 6-99.

Consider the architecture in Fig. 6-105; this architecture is a straightforward modification to that of Fig. 6-104 and incorporates a simple averaging filter with all-ones coefficients that is downsampled by a rate that is equal to the number of coefficients that are in the filter (L). Very large filter orders are possible with the simple feedback structure as illustrated in the figure. The simple averaging filter depicted is often referred to as the integrate-and-dump filter. The input to this averaging filter may be switched with control signals c_i to be either zero ($c_i = 0$) or the sampled input signal ($c_i = 1$). This is done to allow feedback control from the decision processing such that only samples estimated to contain a PPM pulse are input to the slot synchronization algorithm. Note that due to the large overlap of input vectors in Figs. 6-99 and 6-105, significant further simplifications are possible in the realization of these architectures.

Each output line of the vector downsample block in Fig. 6-105 is a vector, so there are actually a vector of switches and averaging filters operating at rate f_s, but the output of these filters is clocked at rate $f_s/(ML)$; this is the input rate to the time-varying filter bank. However, the filter bank $\overline{F}^D_{N_1,p}(z,u,i)$ operates at rate Mf_S/L, that is M times faster than its input because this one filter structure is used to process all inputs. This is referred to as *hardware reuse*. This architecture requires 1/16th the number of concurrent multipliers of that required by Fig. 6-104 for the same prototype filter order (same signal processing performance). This may result in a significant savings in concurrent multipliers but with some added complexity of register delays and other controlling logic that are required in the hardware reuse architecture. This complexity reduction may result in a filter bank design $\overline{F}^D_{N_1,p}(z,u,i)$ with more multipliers and/or more time-shifted filters (larger N) than would otherwise be feasible to implement. *This complexity reduction increases the design options of the filter $\overline{F}^D_{N_1,p}(z,u,i)$, and thus, the design and performance options of the slot-synchronization algorithm that are feasible.* It should be emphasized that a full-rate filter bank $F^D_{N,p}(z,u,i)$ is still required to filter and correct the timing error of the input signal that is then input to the decision process (recall this is the process of estimating the transmitted symbol). Figure 6-106 illustrates the architecture with deconstructed slot-synchronization (timing error estimation) and timing-error correction in the post-detection filtering with some pulse equalizer $B_2(z)$.

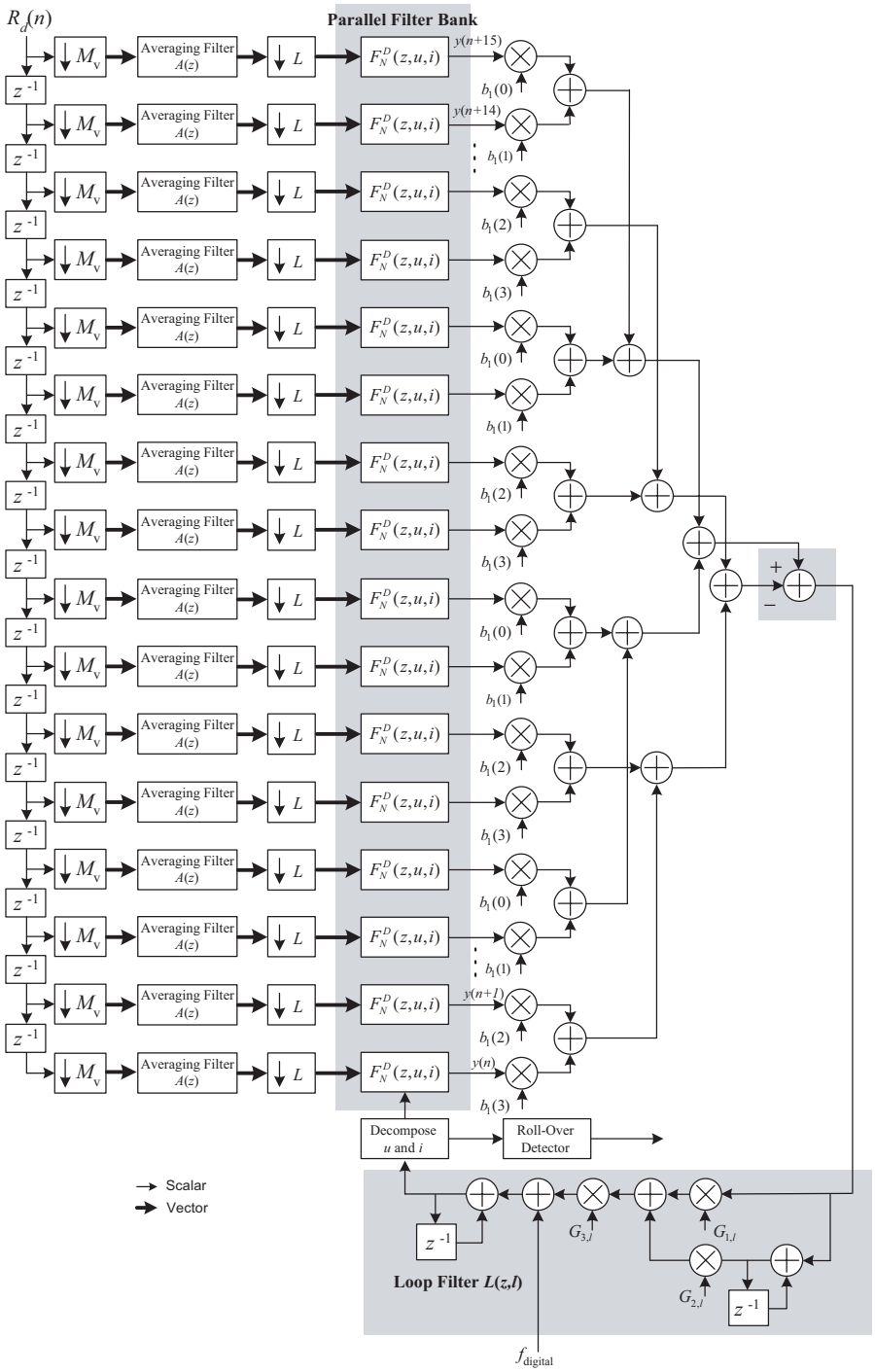


Fig. 6-104. Alternative slot-synchronization architecture equivalent to that of Fig. 6-103.

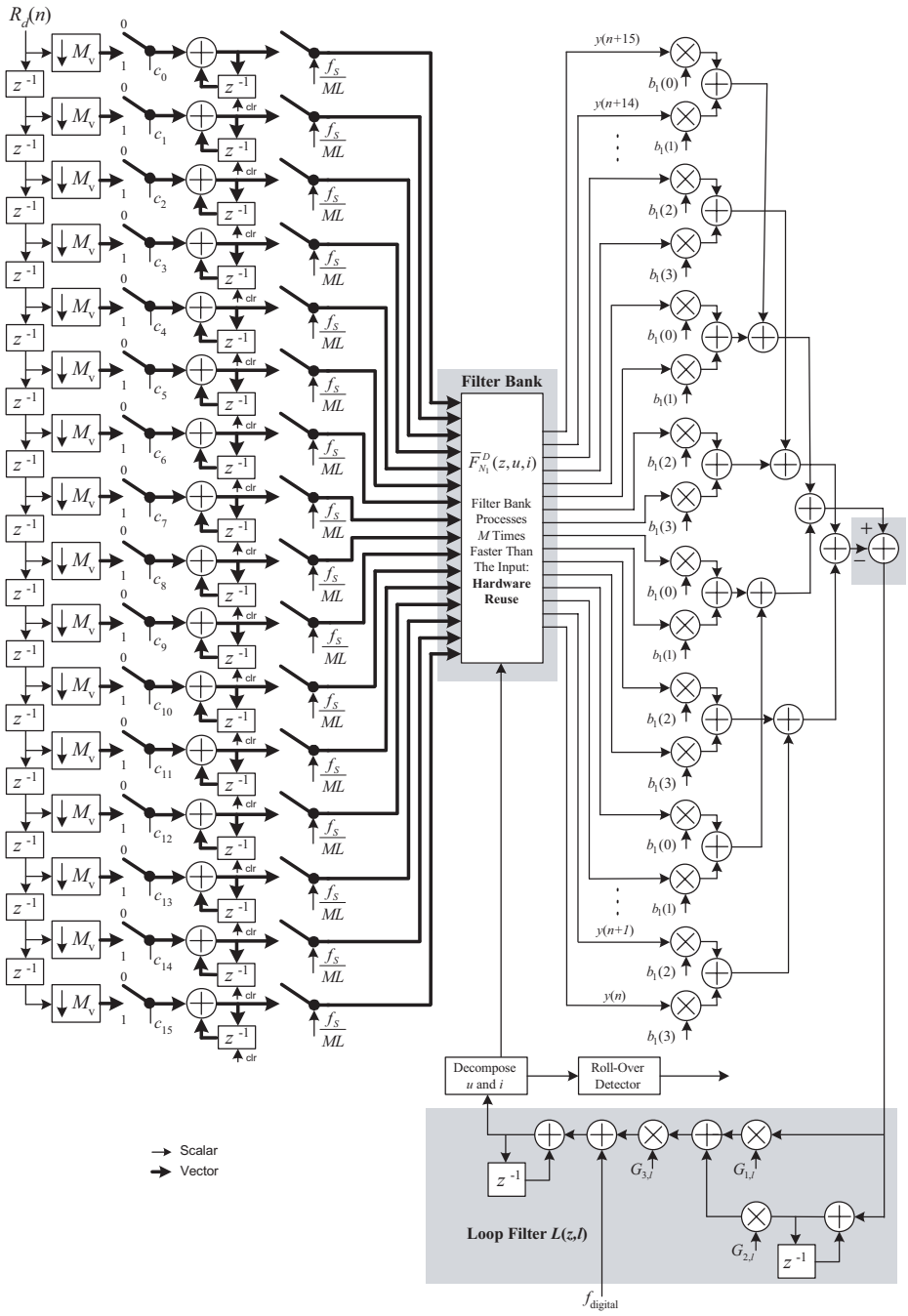


Fig. 6-105. Simplified alternative slot-synchronization architecture; filter rate = $M \times$ (the parallel input rate).

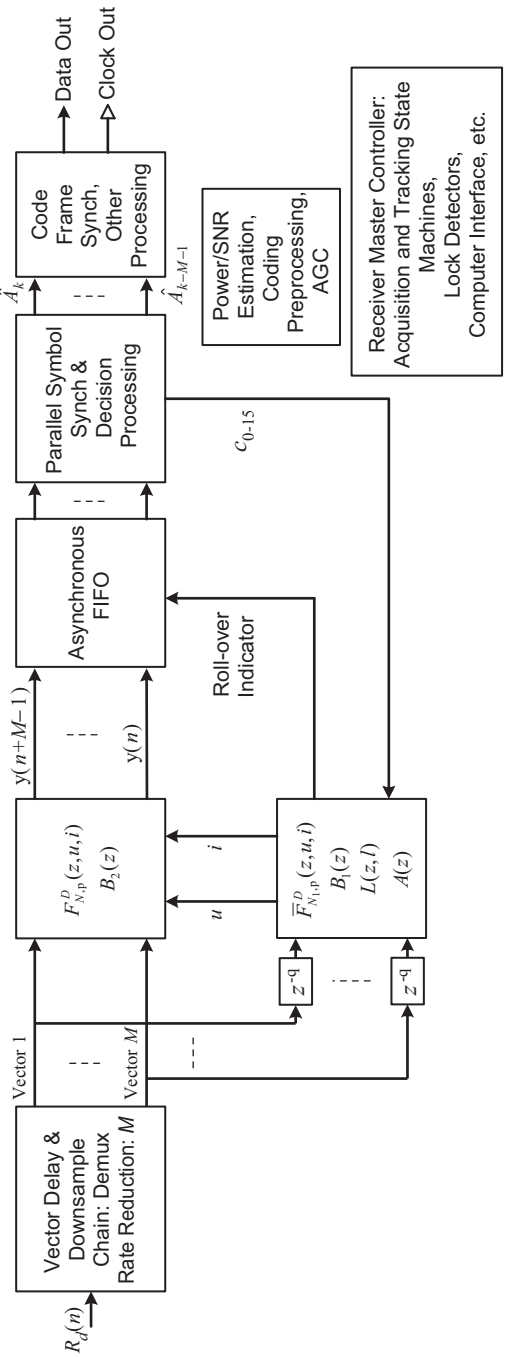


Fig. 6-106. Alternative parallel demodulator architecture: slot-synchronization error estimation with filter bank $\overline{F}^{D}_{N1,p}(z,u,i)$ and equalizer filter $B_1(z)$ separated from the post-detection filter bank $F^{D}_{N,p}(z,u,i)$ and detection equalizer $B_2(z)$; the processing rate of the filter $\overline{F}^{D}_{N1,p}(z,u,i) = M(f_s)/L$ and filter $F^{D}_{N,p}(z,u,i) = f_s/M$.

The processing rate of the filter $\overline{F}^{D}_{N_1,p}(z,u,i) = M(f_S)/L$, the rate of filter the $F^{D}_{N,p}(z,u,i)$ is f_S/M. The time-varying filter $\overline{F}^{D}_{N_1,p}(z,u,i)$ and pulse equalizer $B_1(z)$ for slot-timing error estimation may be designed and implemented independently of the time-varying post-detection filter $F^{D}_{N,p}(z,u,i)$ for filtering and signal timing error correction and the equalizer $B_2(z)$ for pulse filtering (possibly for pulse distortion mitigation). Note the signal delay (z^{-q}) between the slot-synchronization processing and the signal path containing the post-detection filter bank and the decision processing. This is to facilitate appropriate feedback from symbol synchronization to slot synchronization if such a decision-directed system is desired. This delay is not in the feedback path of the closed-loop synchronization algorithm. This is not the case for the architecture of Fig. 6-103, which required a delay in the feedback path in order to be decision-directed. This delay may be small, but any delay in the feedback has the potential to degrade synchronization performance.

Summarizing the advantages of the architecture in Fig. 6-106, first filter bank $\overline{F}^{D}_{N_1,p}(z,u,i)$ may be designed and optimized for synchronization estimation performance independently of the filter bank $F^{D}_{N,p}(z,u,i)$ used for correction timing errors and filtering for the decision process. The filter bank $\overline{F}^{D}_{N_1,p}(z,u,i)$ may incorporate many more design choices and may be many times more computationally intensive, that is a larger number of coefficients and/or larger resolution, and/or larger *N* than would otherwise be possible. The architecture of Fig. 6-106 does not require an additional delay in the synchronization feedback path in order to be decision-directed. This architecture may or may not be a good choice over the architectures of Fig. 6-99 or Fig. 6-103. Depending on the system, desired performance and regions of operation, and the resulting filter bank design(s), the architecture of Fig. 6-106 may be more or less computationally intensive than the architectures of Fig. 6-99 or Fig. 6-103.

The filter bank $F^{D}_{N,p}(z,u,i)$ may be designed for specific pulse shapes and jitter conditions. Although the on-chip memory requirements likely preclude one of numerous banks of filters to be chosen on-the-fly without reconfiguration, there is another solution. As an example of the flexibility of this architecture when targeting reconfigurable (FPGA) VLSI implementation multiple filter banks, *k* filter banks, $F^{D}_{k,N,p}(z,u,i)$, may be developed and stored off-line and reconfigured for field tests and experiments very rapidly (seconds or less).

Finally, the architectures presented in this section are examples of those that may be developed given the methods and framework presented in this work along with other multirate discrete-time signal processing theorems and communications demodulator algorithms found in the literature. Many further discrete-time PPM demodulator architecture variations may be developed with varying degrees of complexity and performance. The conclusion includes a discussion of the high speed digital platform suitable for implementing such architectures and the development process that facilitates an evolving architecture development incorporating software modeling and simulation, analysis, laboratory experimentation, and rapid prototyping. We now present the key design equations for understanding the architecture from a frequency (clock rate) and transmission line perspective; these are useful in understanding the complexity and challenges of the high-speed platform that is used to implement the class of architectures presented here.

6.3.9 Primary System Models and Parameters

There are several aspects of the discrete-time architecture illustrated in Fig. 6-99 that must be considered in the development of the high-speed digital platform. Two of the most significant are the frequency (clock rate) requirements and the number of required data transmission lines. These requirements are largely determined by the sample rate of the system along with the hardware limitations of the digital devices used to realize the discrete-time or digital architectures that together determine M and the number of parallel input/output signals.

The primary clock frequencies are represented in the first four illustrations in Fig. 6-107, the maximum frequencies are in the right-most shaded column. These clock relationships are critical for such things as high-speed platform design, VLSI circuit design, and floor-planning. The relationship between the clock frequency on a given axis and the clock frequency on one axis above is given in the left-most shaded column. The fastest digital clock rate is the sample frequency (also referred to as the system frequency) on the top axis; this rate is reduced by the serial-to-parallel conversion, resulting in the divide by M. The slot rate clock is represented on the second axis, the symbol frequency is given on the third axis, and the bit clock frequency is given in the fourth axis. The slot synchronization loop bandwidth is illustrated on axes as a fraction of the symbol rate.

Certain fundamental design equations of the architecture given in Fig. 6-99 are now summarized. The variables in these equations fall into two categories: *design parameters* and *operational parameters*. It is a key step in the development of complex VLSI systems to assign all parameters to one of these two categories. Such categorization is essential for establishing priorities and the associated schedule for establishing parameter values. This, in turn, allows

development resources to be focused appropriately for various development strategies; one of which is summarized in the conclusion of this work. The two categories are defined as follows.

1) Design parameters determine the physical construction of the prototype board layout and design, the VLSI device, the package, and numerous other significant elements of the design. These physical parameters are generally either not reconfigurable after manufacture or require significant redesign. These parameters should be determined before significant design work is undertaken.

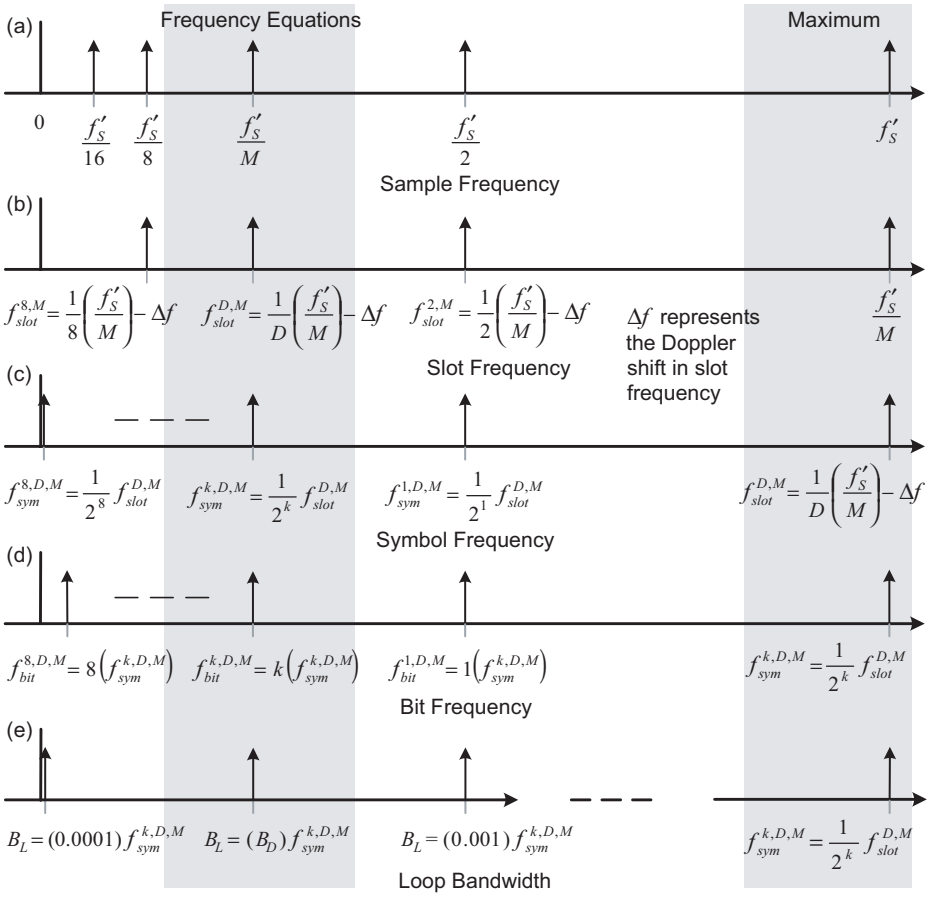


Fig. 6-107. Frequency model.

2) Operational parameters may take on a variety of values depending on the requirements of the system. The operational parameters are either programmable or reconfigurable without significant redesign. The values of these parameters generally do not need to be known precisely before design or even during hardware implementation and test. However, the range, the maximum and minimum values, should be established or estimated as early as possible in the development process.

The top-level design equations and parameters for the architecture in Fig. 6-99 are given in Table 6-6. This is not a comprehensive list of equations describing the digital demodulator, but the most essential ones in early phase hardware development.

Figure 6-108 illustrates a simple digital input/output model of the architecture in Fig. 6-99. This model indicates the approximate number of digital input/output transmission lines required between major signal processing sub-systems of the demodulator. Of primary concern is the number of transmission lines at the output of the parallel-to-serial converter. The number of lines cannot exceed the input capability of the device used to implement the rest of the demodulator. At the same time, the number of lines must support the rate reduction required.

6.3.10 Conclusion and Future Work

Discrete-time demodulator architectures for broadband free-space optical PPM, which are capable of Nyquist or near Nyquist data rates were presented. While these architectures do not represent unique solutions, methods, or designs, they were developed within a framework that encompasses a large body of theoretical work in optical communications, synchronization, and multirate discrete-time signal processing; and they were constrained by many of the limitations of the state-of-the-art digital hardware. The primary focus of this work was on the development of discrete-time processing of the most fundamental algorithms and processing required in PPM demodulators; those necessary for post-detection filtering, synchronization, and decision processing. Numerous other processing subsystems are required in an operational receiver. Integrated in the development and discussion were numerous considerations of modern VLSI devices as well as platform design. The most fundamental design decisions, those globally impacting platform design or fundamentally limiting performance, are summarized as follows:

- The development targets an all-digital approach as opposed to a hybrid analog-digital or primarily analog implementation.

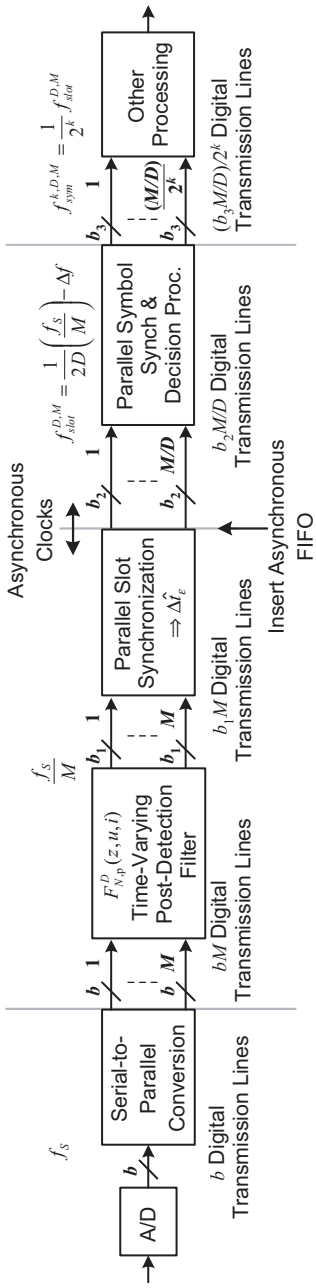
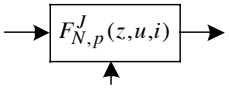
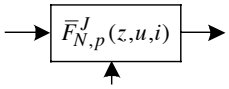
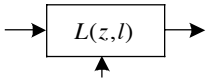


Fig. 6-108. Data I/O model.

Table 6-6. List and description of primary design functions.

Design Function	Parameter Description
$f_{bit}^{k,D,M} = \frac{k}{2^k DM}\left(\frac{f_S}{M}\right)$ $= \underbrace{\left(\frac{k}{2^k}\right)}_{\text{Operational Parameter}} \underbrace{(f_s)\ \left(\frac{1}{D}\right)\ \left(\frac{1}{M}\right)}_{\text{Design Parameters}}$	*Operational:* k = The number of bits per symbol *Design:* D = Nominal number of samples per slot M = Processing rate reduction f_S = Ideal sample rate, not actual sample rate
$f_{sym}^{k,D,M} = \frac{1}{2^k} f_{slot}^{D,M}$	*Operational:* k = bits per symbol 2^k = Number of slots per symbol
$\frac{D}{T_{slot} + T_{slot}\left(\frac{v_{\min}}{c}\right)} \le f_s'$	*Design:* f_S' = The actual sample rate must be greater than the Doppler-shifted slot rate times D
Parallel Time-Varying Post-Detection Filter  i = sample index u = sub-sample index $h_{dN}(n)$ or $v(n)$ prototype filter	*Operational:* N = Prototype FIR filter oversample rate, maximum residual sample offset = $T_S/(2N)$, there are N filters in the filter bank RDN = Number of coefficients in the prototype filter $J(T_S)$ = Timing error accumulated before roll-over
Parallel Time-Varying Slot-Synch Filter  i = sample index u = sub-sample index $h_{dN}(n)$ or $v(n)$ prototype filter	*Operational:* N = Prototype FIR filter oversample rate, maximum residual sample offset = $T_S/(2N)$, there are N filters in the filter bank RDN = Number of coefficients in the prototype filter $J(T_S)$ = Timing error accumulated before roll-over
Slot Synch Loop Filter  l = filter tracking/acquisition index $B_{LD} = (B_N) f_{sym}^{k,D,M}$	*Operational:* B_{LD} = Discrete-time bandwidth of IIR loop filter: $G_{1,l}$, $G_{2,l}$, $G_{3,l}$ B_N = Discrete-time bandwidth normalized as a fraction of the symbol rate (B_L = Prototype analog loop filter bandwidth)
$\frac{1}{T_{slot}} \le f_{\text{digital}} \frac{1}{T_{slot} + T_{slot}\left(\frac{v_{\min}}{c}\right)}$	*Operational:* f_{digital} = Normalized digital slot frequency rate

- The all-digital design or lack of feedback from digital-to-analog processing results in tremendous flexibility without redesign; the many orders of magnitude (typically greater than 8 with current CMOS) possible with digital implementation can be achieved without the difficulties and limitations of the digital-to-analog feedback approach.
- The fixed-sample per slot design results in a simplified demodulator development. As the symbol rate changes, the number of samples per slot remains constant. Once one data rate has been tested and verified, a large subset of characteristics and parameters for all data rates has been verified; thus, the end result is a more simple and time-efficient development. The downside of this approach is that the anti-aliasing filter before the A/D may require a variable bandwidth, and these issues are discussed further in [105,106]. This latter is one motivation for not using this fixed number of samples per slot approach, and there are others. It should be noted that the design process and the reconfigurable platform do not fundamentally limit the choice of alternative designs for different data rates, or for a different number of samples per slot. With an FPGA approach and an appropriately designed platform, the decision to use a different number of samples per slot may be made at virtually any phase of the development, although architecture iteration would be required.

There are numerous other design decisions that were made here and even more that are yet to be made during the development. Many of these decisions require further analysis, such as precisely which slot-synchronization algorithm to use and its parameter values for various regions of operation (SNR and signal dynamics), and many are decisions made by digital designers that have no direct impact on communications performance and are not dependent on regions of operation. It must be understood that a great many of these open-design questions cannot be answered by considering and optimizing one or even a small number of factors. Answering design questions about VLSI systems as complex as those illustrated in Figs. 6-99, 6-103, and 6-106 are usually not about optimality of a criterion but trading off many criteria against one another. One of the motives of Section 6.3.8 was to illustrate some of these trades. If the core problem posed there (asynchronous parallel processing) had been done so and analyzed in purely mathematical terms, the most critical design questions would have been largely obscured. The asynchronous filter bank design is a very complex undertaking with numerous tradeoffs made between hardware complexity, rate reduction, impact of filter design on communications systems performance (synchronization and decision processing), and commercially available methods for processing asynchronous data streams (many of which are device dependent).

The focus in this work was not on the analysis of performance of optical communications systems; although the architectures presented are capable of

closely approximating theoretical performance of certain continuous-time systems. This assertion is given without mathematical proof, but basic simulation results of a software model were presented. In addition there are numerous examples of communications algorithms and theoretical constructs based on similar signal processing theories and structures [105,106]. The performance analysis of the parallel architectures presented here is a significant undertaking and must incorporate the numerous characteristics of the deep-space optical communications channel including the primary transmitter and front-end detector characteristics. Figure 6-109 illustrates the development path of a digital receiver from a high level.

Figure 6-110 illustrates the development path in greater detail. The state-of-the art hardware development process is given with emphasis on VLSI circuit realization in FPGA devices. These devices have revolutionized modern hardware design. In addition to the obvious advantages of reconfigurable hardware, the stringent risk reduction and management techniques required by historical methods of VLSI implementation (analog and digital), particularly application-specific integrated circuits, are significantly relaxed. This results in a significantly altered design process. In addition, the inclusion of a development feedback path is possible as depicted in the figure; such paths were often not feasible in VLSI design and implementation before the advent of the FPGA.

This process, if used correctly and in conjunction with a well designed hardware platform, facilitates rapid prototypes to be generated targeting specific applications. If the application changes, the design can be reiterated. *Arguably the largest advantage of this process targeting reconfigurable devices is the ability to make forward progress without a full set of requirements or without establishing values for all operational parameters. This is often the case with new or untested complex systems where a combination of analysis*

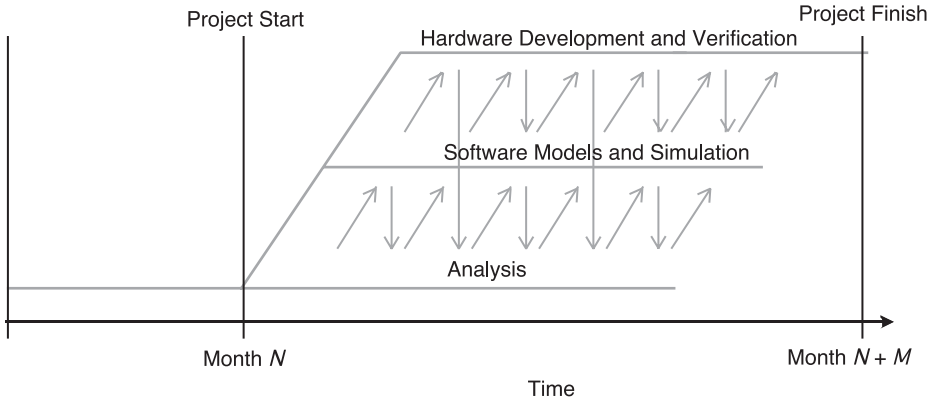


Fig. 6-109. Parallel development model.

and experimental physical work must be accomplished in conjunction to advance the state-of-the art. However, a certain set of design parameters must be established to properly design the hardware platform; if not commercially available, this platform design itself requires another design process with many steps whose description is beyond the scope of this work. Certainly such things as A/D selection, FPGA selection, determination of maximum clock rates and clock rate ranges, and a host of other parameters must be determined before the platform design and implementation or commercial purchase is made in the event that a commercial solution is available.

The development in this work was presented at a sufficiently high level to address the most significant challenges of the all-digital design and

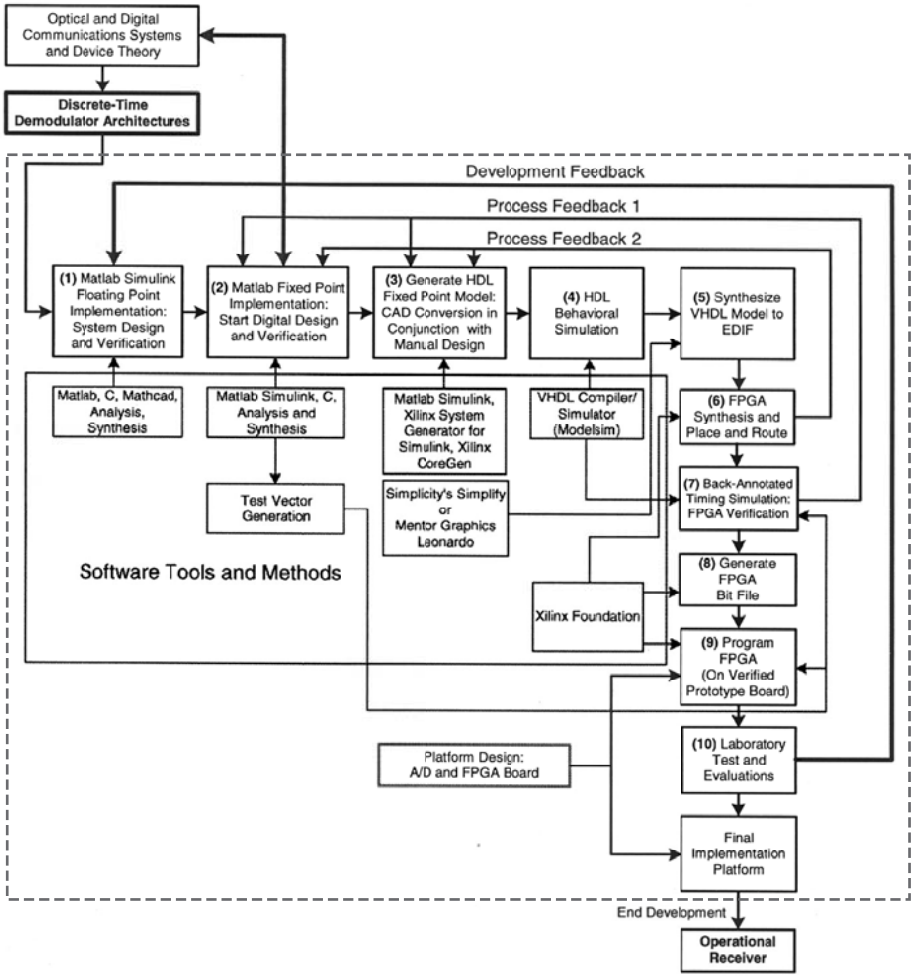


Fig. 6-110. State-of-the-art design process: the path from concept to operational digital demodulator.

implementation. There were many omissions in the work presented here that must be addressed in the development of a complete receiver. The primary issues of the demodulator development in the pre-hardware laboratory test phase, many of which should be addressed targeting a specific set of requirements, are summarized below.

- The architectures presented may be further developed to incorporate numerous simplifications; many of these simplifications occur in a complex trade-space of performance and implementation device limitations. Of particular interest is minimizing the complexity of the filter bank $F_{N,p}^{D}(z,u,i)$.
- The function of automatic gain control (AGC) was largely ignored in this work. This circuit may be accomplished in analog or digital circuits or both. In many systems a "course" AGC circuit is used to control the amplitude of the signal input to the A/D to avoid clipping and underflow, while a "fine" AGC is implemented digitally to provide tight control of amplitude for such signal processing systems as the slot synchronization loop whose stability, bandwidth, damping factor, and overall performance depend on amplitude statistics.
- The completed discrete-time demodulator architecture must be converted to include quantized amplitude. This typically involves varying degrees of analysis and simulation.
- The performance and type, open or closed loop, of the discrete-time slot synchronization algorithm must be established for various receiver regions of operation using analysis techniques outlined in references presented in this work.
- When a closed-loop system derived from variations of Fig. 6-74 and the references given in Section 6.3.3.2 the required analysis may be broken down into a number of subtasks:
  - The discrete-time loop design equation must be established; and gain margin, bandwidth, damping factor, and many other characteristics must all be determined.
  - Characterization of synchronization performance with the large signal amplitude fluctuation characteristic of the optical channel.
  - Tracking versus acquisition performance must be established for the operational scenario or scenarios. This will include development of an acquisition algorithm. Among other things this algorithm might determine when to switch loop filter bandwidths autonomously.
  - Slot lock detection circuits and performance must be established. Probability of false-lock and false-alarm must be established for these circuits.
  - Performance in the presence of transmitter and receiver oscillator jitter must be established.

- A simple interpolation filter may be used with the system of Figs. 6-101 through 6-106. However, the slot or interpolation filter might be designed to achieve specified performance of the PPM slot synchronization loop. The design might be further refined to incorporate criteria derived from the optimal detection given the actual channel.

- Symbol synchronization performance and lock detection must be established. Blind or pilot sequence system or both may be implemented.
- Complete end-to-end system Monte-Carlo or pseudo-Monte-Carlo performance modeling and simulation of the demodulator must be performed. This type of comprehensive validation is accomplished in phases, starting with basic subsystem validation and progressively including more and more subsystems and channel parameters. This type of testing and validation accomplishes two goals, first system performance may be determined or approximated when doing so analytically is not possible or not feasible, and second cost and schedule risks of validating an operational system in the field are reduced. To first order, the resources that are spent on these end-to-end simulations is determined by the level of performance prediction required, the number of operational scenarios, and the risk tolerance of the project. As an example, before a system as complex as that of Fig. 6-103 is field tested, thousands of hours of simulations will be run to accomplish these goals even with moderate risk reduction goals over a moderate number of operational scenarios (a few dozen); many times that number may be utilized for more thorough validation across a more complex set of operating scenarios. These test and verification plans are integrated into the process of Fig. 6-108.
- More optimal slot synchronization algorithms might be developed. In addition, the decision-directed slot and symbol loops, different slot synchronization algorithms might be derived based on open-loop optimality criterion. It is envisioned that such algorithms may be implemented in the architecture given in Figs. 6-99, 6-103, or 6-106 with design evolution (it is envisioned that the majority of the architecture will remain unchanged).
- Modern forward error correction (FEC) decoders may require preprocessing; the performance of this preprocessing must be established.

This is a grossly simplified list of tasks that must be completed as part of the development of the parallel digital demodulator for free-space optical PPM. Before comprehensive laboratory testing and validation of the complete demodulator is undertaken the majority of these tasks should be completed. However, the process of Fig. 6-108 is flexible and gives project managers many

options for trading development schedule and progress with risk. It is possible and in many cases highly desirable to largely complete certain stand-alone signal processing subsystems, that is step them through the entire process of Fig. 6-108, and verify that subsystem with a laboratory hardware test. Such "front-runner" efforts serve a valuable process-risk reduction role by validating the applicability and suitability of various vendor tools and software tool upgrades, as well as hardware devices and platforms. Such "front-runner" activities also serve to demonstrate and validate the end-to-end concept-to-hardware process.

References

[1] H. P. Stahl, G. H. Rowell, G. Reese, A. Byberg, "Multivariable Parametric Cost Model for Ground-Based Telescopes," *Proceedings of the SPIE*, vol. 5497, *Modeling and Systems Engineering for Astronomy*, S. C. Craig and M. J. Cullum, editors, pp. 173–180, September 2004.

[2] J. E. Carey, C. H. Crouch, and E. Mazur, "Femtosecond Laser-Assisted Microstructuring of Silicon Surfaces," *Optics and Photonics News*, vol. 14, no. 2, pp. 32–36, February 2003.

[3] D. L. Robinson and J. R. Lesh, "A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes," *Proceedings of SPIE OE Lase*, Paper 756-20, pp. 130–134, SPIE, Bellingham, Washington, January 1987.

[4] K. Shaik, M. Wilhelm, and R. Horttor, *Ground Based Advanced Technology Study (GBATS)*, JPL D-11000 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 5, 1994.

[5] E. L. Kerr, "Strawman Optical Reception Development Antenna (SORDA)," *The Telecommunications and Data Acquisition Progress Report 42-93, January–March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 97–110, May 15, 1988.
http://ipnpr.jpl.nasa.gov/progress_report/

[6] D. Wonica, "Economical 10-m Class Ground-Based Receiver for Deep-Space Optical Communications," *Proceedings of the SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI***,** G. S. Mecherle, Editor, pp. 487–497, August 1988.

[7] C. Kramer, C. G. Degiacomi, U. U. Graf, R. E. Hills, M. Miller, R. T. Schieder, N. M. Schneider, J. Stutzki, and G. F. Winnewisser, "New KOSMA 3-m Telescope," *Proceedings of the SPIE*, vol. 3357, *Advanced Technology MMW, Radio, and Terahertz Telescopes*, T. G. Phillips, Editor, pp. 711–720, July 1998.

[8] J. A. Booth, M. J. Wolf, J. R. Fowler, M. T. Adams, J. M. Good, P. W. Kelton, E. S. Barker, P. Palunas, F. N. Bash, L. W. Ramsey, G. J. Hill, P. J. MacQueen, M. E. Cornell, and E. L. Robinson, "The Hobby Ebberly Completion Project," *Proceedings of the SPIE*, vol. 4837, *Large Ground-based Telescopes*, J. M. Oschmann, L. M. Stepp, Editors, pp. 919–933, February 2003.

[9] E. Kerr "An Integral Sunshade for Optical Receiving Antennas," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp.180–195, November 15 1988.
http://ipnpr.jpl.nasa.gov/progress_report/

[10] E. L. Kerr and C. W. DeVore, "Shutters and Slats for the Integral Sunshade of an Optical Reception Antenna," *The Telecommunications and Data Acquisition Progress Report 42-95, July–September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 196–201, November 15, 1988.

[11] B. Clymer "Reflected Sunlight Reduction and Characterization for a Deep Space Optical Receiver Antenna (DSORA)" *The Telecommunications and Data Acquisition Progress Report 42-103, July–September 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 110–117, November 15 1990.
http://ipnpr.jpl.nasa.gov/progress_report/

[12] K. Wilson, *Trip Repot to Duisburg, Germany and Zermatt, Switzerland*, JPL D-32056 (JPL Internal document), Jet Propulsion Laboratory, Pasadena, California, November 5, 1997.

[13] J. V. Sandusky, D. J. Hoppe, and M. J. Britcliff, "Deep Space Optical Reception Antenna (DSORA): Aperture Versus Quality," *The Telecommunications and Mission Operations Progress Report 42-143, July–September 2000*, pp. 1–11, November 15, 2000.
http://ipnpr.jpl.nasa.gov/progress_report/

[14] Hextek Corporation web site, accessed June 7, 2005. http://www.hextek.com

[15] S. Piazzolla and S. Slobin, "Statistics of Link Blockage Due to Cloud Cover for Free-Space Optical Communication Using NCDC Surface Weather Observation Data," SPIE Meeting, San Jose, California, 2002, *Proceedings of the SPIE*, vol. 4635, *Free-Space Laser Communication Technologies XVI*, G. Stephen Mecherle, editor, pp. 138–149, April 2002.

[16] R. P. Link, R. J. Alliss, and M. E. Craddock, "Mitigating the Effect of Clouds on Optical Communications," *Proceedings of SPIE*, vol. 5338, *Free-Space Laser Communication Technologies XVI*, G. S. Mecherle, C. Y. Young, J. S. Stryjewski, editors, pp. 223–232, June 2004.

[17] K. Wilson M. Troy, M. Srinivasan, B. Platt, V. Vilnrotter, M. Wright, V. Garkanian, H. Hemmati, "Daytime Adaptive Optics For Deep Space Optical Communications" *Space Activities and Cooperation Contributing to All Pacific Basin Countries: Tenth International Space Conference of Pacific Basin Countries, ISCOPS*, Tokyo, Japan, December 10–12, editors: P. M. Bainum, L. Furong, T. Nakajima, American Astronomical Society, pp. 481–492, 2003.

[18] G. S. Mecherle, W. Akle, C. J. Starkus, J. E. Klein, and G. W. Holleman, "Coherent Detection Optical Relay Satellite for Deep-Space Communication," *Proceedings of SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI*, G. S. Mecherle, editor, pp. 218–236, August 1994.

[19] G. S. Mecherle, W. Akle, C. J. Starkus, and J. E. Klein, "Direct Detection Optical Relay Satellite for Deep-Space Communication," *Proceedings of SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI*, G. S. Mecherle, editor, pp. 134–155, August 1994.

[20] K.E. Wilson, K. Shea, J. Ceniceros, M. Wright, and R. Cesarone, "Cost and Performance Comparison of an Earth-Orbiting Optical-Comm Relay Transceiver and a Ground-based Optical Receiver Subnet," *The Interplanetary Network Progress Report 42-153, January–March 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–12, May 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[21] S. C. Casey, "The SOFIA Program: Astronomers Return to the Stratosphere," *Advances in Space Research*, vol. 34, pp. 560–567, 2004.

[22] K. Wilson, H. Hemmati, and S. Lee, *Mars Optical Communications Study Report*, JPL D-28438 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 26, 2001.

[23] *NMP ST6 Final Briefing Acquisition, Pointing and Tracking for Interplanetary Optical Communication*, Ball Aerospace and Technologies Corporation and Jet Propulsion Laboratory Optical Communications Group, Ball Aerospace, Boulder, Colorado, August 2001.

[24] J. Berg and R. St. Piere, *Solid State Lasers OTRW Presentation to JPL, July 1, 1997*, JPL D-32057 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1997.

[25] N. Savage, "Fiber Lasers Power up," *Optical Engineering Magazine*, vol. 3, p. 53, August 2003.

[26] N. Peyghambarian and A. Schülzen, "High Power Devices in Compact Packages," *Optics and Photonics News*, pp. 36–41, June 2005.

[27] D. L. Fried, "Aperture Averaging of Scintillation," *Journal of the Optical Society of America*, vol. 57, no. 2, pp. 169–175, 1967.

[28] J. Churnside, "Aperture Averaging of Optical Scintillations in the Turbulent Atmosphere," *Applied Optics*, vol. 30, no. 15, pp. 1982–1994, 1991.

[29] J. C. Christou, B. Ellerbroek, R. Q. Fugate, D. Bonaccini, and R. Stanga, "Rayleigh Beacon Adaptive Optics Imaging of ADS 9731: Measurements of the Isoplanatic Field of View," *Astrophysical Journal*, vol. 450, pp. 369–379, September 1, 1995.

[30] P. Wizinowich, D. S. Acton, C. Shelton, J. Gathright, K. Ho, W. Lupton, K. Tsubota, O. Lai, C. Max, J. Brase, J. An, K. Avicola, S. Olivier, D. Gavel, B. McIntosh, A. Ghez, and J. Larkin, "First Light Adaptive Optics Images from the Keck II Telescope: A New Era of Angular Resolution Imagery," *Publications of the Astronomical Society of the Pacific* (PASP), vol. 112, pp. 315–319, March 2000.

[31] J. M. Spinhirne and G. A. Ameer, "Adaptive Optics Using the 3.5-m Starfire Optical Range Telescope," Proceedings of SPIE, vol. 3126, Adaptive Optics and Applications, R. K. Tyson and R. Q. Fugate, editors, pp. 257–268, October 1997.

[32] P. M. Metzen "High Reliable Copper Vapor Lasers," *Proceedings of the ESO Workshop Laser Technology for Guide Star Adaptive Optics Astronomy*, Garching, Germany June 23–26, 1997.

[33] E. Kibblewhite, "The Design and Performance of Laser Systems for Generating Sodium Beacons," *Proceedings of NATO Advanced Studies Institute on Laser Guide Star Adaptive Optics for Astronomy*, vol. 551, September 20–October 10 , 1997, edited by N. Ageorges and C. Dainty, Kluwer Academic Publishers, pp. 67–87, 1997.

[34] S. S. Olivier, J. A. K. Avicola, H. D. Bissinger, J. M. Brase, H. W. Friedman, D. T. Gavel, C. E. Max, J. T. Salmon, and K. E. Waltjen, "Performance of Laser Guide Star Adaptive Optics at Lick Observatory," *Proceedings of the SPIE*, vol. 2534, San Diego, California, pp. 26–37, August 1995.

[35] S. S. Oliver, D. T. Gavel, "Tip-tilt Compensation for Astronomical Imaging," *Journal of the Optical Society of America* A, vol. 11, no. 1, pp. 368–374, January 1994.

[36] R. Foy, J.-P. Pique, A. Petit, P. Chevrou, V. Michau, G. Grynberg, A. Migus, N. Ageorges, V. Bellanger, F. Biraben, R. Deron, H. Fews, F. Foy, C. Högemann, M. Laubscher, D. Mül1er, C. d'Orgeville, O. Peillet, M. Redfern, M. Schöck, P. Segonds, R. Soden, M. Tallon, E. Thiébaut, A. Tokovinin, J. Vaillant, and J.-M. Weulersse, "Polychromatic Guide Star Feasibility Study," *Proceedings of the SPIE*, vol. 4065, *High-Power Laser Ablation III*, C. R. Phipps, editor, pp. 312–323, August 2000.

[37] M. Reyes, A. Comeron, A. Alonso, A. Rodriguez, J.A. Rubio,V. Fedrico Dios, S. Chueca, and S. Zodnik, "Ground-to-Satellite Bidirectional Laser Links For Validation of Atmospheric Turbulence Model," *Proceedings of the SPIE*, vol. 4821, Seattle, Washington, pp. 33–43, July 2002.

[38] C. Higgs, H. Barclay, D. Murphy, and C. A. Primmerman, "Multibeam Illumination," *Lincoln Laboratory Journal*, vol. 11, pp. 8–22, 1998.

[39] M. Jeganathan, K. E. Wilson, and J. R. Lesh, "Preliminary Analysis of Fluctuations in the Received Uplink-Beacon-Power Data Obtained From the GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report 42-124, October–December 1995*, pp. 20–32, Jet Propulsion Laboratory, Pasadena, California, February 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[40] K. E. Wilson, A. Biswas, S. Bloom, and V. Chan "Effect of Aperture Averaging on a 570 Mbps 42 km Horizontal Path Optical Link," *Proceedings of the SPIE*, vol. 2471, pp. 244–253, April 1995.

[41] H. T. Yura and W. G. McKinley, "Optical Scintillation Statistics for IR Ground-to-Space Laser Communications," *Applied Optics*, vol. 22, no. 21, pp. 3353–3358, November 1, 1983.

[42] K. Wilson, W. T. Roberts, V. Garkanian, F. Battle, R. Leblanc, H. Hemmati, and P. Robles "Plan for Safe Laser Beam Propagation from the Optical Communications Telescope Laboratory," *The Telecommunications and Data Acquisition Progress Report, 142-152, October–December 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, February 15, 2003.

[43] K. E. Wilson and J. R. Lesh, "An Overview of the Galileo Optical Experiment (GOPEX)," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 192–204, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[44] *RCA - Primus - Honeywell - Sperry Primus-40 WXD Digital Radar Maintenance Manual*, IB8029050, Rev-3, p. 4-3, October 1978.

[45] M. Skolnik, *Introduction to Radar Systems*, Second Edition, McGraw–Hill (New York, St. Louis), p. 15, 1980.

[46] L. D. Smullin and G. Fiocco, "Project Luna See," Institute of Electrical and Electronics Engineers Proceedings, vol. 50, pp. 1703–1704, 1962.

[47] L. H. Allen, M. S. Schumate, and J. W. Young, *The Surveyor-VII Laser Pointing Experiment Table Mountain Station Description*, JPL D-32055 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1968.

[48] K. E. Wilson and J. R. Lesh, "An Overview of the Galileo Optical Experiment (GOPEX)," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, pp. 192–204, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[49] R. Q. Fugate, "GOPEX at the Starfire Optical Range," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 255–279, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[50] W. M. Owen, Jr., "Telescope Pointing for GOPEX," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 230–235, August 15, 1993. http://ipnpr.jpl.nasa.gov/progress_report/

[51] B. M. Levine, K. S. Shaik, and T.-Y. Yan, "Data Analysis for GOPEX Image Frames," *The Telecommunications and Data Acquisition Progress Report 42-114, April–June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 213–229, August 15, 1993.

[52] B. M. Levine and K. Kiasaleh, "Intensity Fluctuations in the Compensated Earth-Moon-Earth Laser Link (CEMERLL) Experiment," *Proceedings of the SPIE*, vol. 2123, *Free-Space Laser Communication Technologies VI*, G. S. Mecherle, editor**,** pp. 409-422, January 1994.

[53] P. J. Shelus, "MLRS: a Lunar/Artificial Satellite Laser Ranging Facility at the McDonald Observatory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-234, pp. 385–390, 1985.

[54] J. E. Faller, "The Apollo Retroreflector Arrays and A New Multi-lensed Receiver Telescope," *Space Research XII*, Akademie-Verlag, Berlin, pp. 235–246, 1972.

[55] K. E. Wilson, P. R. Leatherman, R. Cleis, J. Spinhirne, and R. Q. Fugate, "Results of the Compensated Earth-Moon-Earth Retroreflector Laser Link (CEMERLL Experiment," *The Telecommunications and Data Acquisition Progress Report 42-131, July–September 1997*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, November 15, 1997.

[56] K. Araki, Y. Arimoto, M. Toyoda, M. Toyoshima, M. Shikatani, T. Takahaski, T. Fukuzawa, H. Okazawa, Y. Suzuki, and T. Aruga, "Laser Communications Experiment Using ETS-VI Satellite," *CRL International Topical Workshop on Space Laser Communications*, March 10–11, 1997 Communications Research Laboratory, Tokyo, Japan, pp. 34–47, 1997.

[57] K. Wilson, J. R. Lesh, K. Araki, and Y. Arimoto, "Preliminary Results of the Ground/Orbiter Lasercom Demonstration Experiment Between Table Mountain and the ETS-VI Satellite," *Proceedings of the SPIE*, vol. 2699, *Free-Space Laser Communication Technologies VIII*, G. S. Mecherle, editor**,** pp. 121–132, April 1996.

[58] K. Wilson, M. Jeganathan, J. Lesh, J. James, and G. Xu, "Results for Phase 1 and Phase 2 GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report 142-128, October–December 1996*, Jet Propulsion Laboratory Pasadena, California, pp. 1–11, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[59] M. Owen, S. D. Gillam, J. W. Young, and D. Mayes, "Ephemeris Generation for the ETS-VI and Its Effects on Pointing Strategies Adopted for Daytime Acquisition and Tracking," *The Telecommunications and Data Acquisition Progress Report 142-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[60] M. Toyoshima, K. Araki, Y. Arimoto, M.Toyoda, M. Jeganathan, K. Wilson, and J. Lesh, "Reduction of ETS-VI laser Communication Equipment Optical-Downlink Telemetry Collected During GOLD," *The Telecommunications and Data Acquisition Progress Report 142-128, October–December 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–9, February 15, 1997.
http://ipnpr.jpl.nasa.gov/progress_report/

[61] M. Jeganathan, M., K. E. Wilson, and J. R. Lesh, "Preliminary Analysis of Fluctuations in the Received Uplink-Beacon-Power Data Obtained From the GOLD Experiments," *The Telecommunications and Data Acquisition Progress Report PR 42-124, October–December 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 20–32, February 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[62] S. D. Gillam, J. W. Young, and D. R. Sidwell, "JPL Table Mountain Facility Support of the Ground/Orbiter Lasercomm Demonstration," *The Telecommunications and Data Acquisition Progress Report 42-125, January–March 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, May 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[63] C. Pasqualino "Characterization of Avalanche Photo Diode Detector Used for Optical Communications Experiments with the Japanese ETS-VI" *The Telecommunications and Data Acquisition Progress Report 142-128, October–December 1996*, Jet Propulsion Laboratory Pasadena, California, pp. 1–8, February 15, 1997. http://ipnpr.jpl.nasa.gov/progress_report/

[64] R. Gagliardi and S. Karp, *Optical Communications*, New York: John Wiley and Sons, 1976.

[65] W. Pratt, *Laser Communications Systems*, New York: John Wiley and Sons, 1969.

[66] J. W. Goodman, *Introduction to Fourier Optics*, New York: McGraw-Hill, 1968.

[67] D. Petersen and D. Middleton, "Sampling and Reconstruction of Wavenumber-Limited Functions in N-Dimensional Euclidian Spaces," *Information and Control*, vol.5, pp. 279–323, 1962.

[68] V.Vilnrotter, *Focal-Plane Processing for Scattered Optical Fields*, Ph.D. Dissertation, University of Southern California, Los Angeles, California, June 1978.

[69] V. Vilnrotter and M. Srinivasan, "Adaptive Detector Arrays for Optical Communications Receivers," *IEEE Transactions on Communications*, vol. 50, no. 7, pp. 1091–1097, July 2002.

[70] P. Negrete-Regagnon, "Practical Aspects of Image Recovery by Means of the Bispectrum," *Journal of the Optical Society of America*, vol. 13, no. 7, pp.1557–1576, July 1996.

[71] C.-C. Chen, "*Figure of Merit for Direct-Detection Optical Channels," The Telecommunications and Data Acquisition Progress Report 42-109, January–March 1992*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, pp. 136–151, May 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[72] D. Boroson, A. Biswas, and B. L. Edwards, "MLCD: Overview of NASA's Mars Laser Communications Demonstration System," *Proceedings of SPIE: Free-Space Laser Communication Technologies XVI*, [Ed. S. Mecherle], vol. 5338, pp. 16–28, 2004.

[73] G. G. Ortiz, J. V. Sandusky, and A. Biswas, "Design of an Opto-Electronic Receiver for Deep-Space Optical Communications," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 136–151, August 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[74] D. M. Boroson, R. S. Bondurant, and J. J. Scozzafava, "Overview of High Rate Deep Space Laser communications Options," *Proceedings of SPIE: Free-Space Laser Communication Technologies XVI*, [Ed. S. Mecherle], vol. 5338, pp. 37–49, 2004.

[75] B. Moision and J. Hamkins, "Deep-Space Optical Communications Downlink Budget: Modulation and Coding," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–28, August 15, 2003.
http://ipnpr.jpl.nasa.gov/progress_report/

[76] A. Biswas and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–38, August 15, 2003.
http://ipnpr.jpl.nasa.gov/progress_report/

[77] M. Britcliffe, D. Hoppe, W. Roberts, and N. Page, "A ten-meter Ground-Station Telescope for Deep-Space Optical Communications: A Preliminary Design," *The Interplanetary Network Progress Report 42-147, July–September 2001*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, pp. 1–17, November 15, 2001.
http://ipnpr.jpl.nasa.gov/progress_report/

[78] V. Vilnrotter, C.-W. Lau, M. Srinivasan, R. Mukai, and K. Andrews, "An Optical Array Receiver for Deep-Space Communication through Atmospheric Turbulence," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–21, August 15, 2003.
http://ipnpr.jpl.nasa.gov/progress_report/

[79] D. M. Boroson, R. S. Bondurant, and D.V. Murphy, "LDORA: A Novel Laser Communications Receiver Array Architecture," *Proceedings of SPIE: Free-Space Laser Communication Technologies XVI*, [Ed. S. Mecherle], vol. 5338, pp. 56–64, 2004.

[80] J. R. Lesh, J. Katz, H. H. Tan, and D. Zwillinger, "2.5-Bit/detected Photon Demonstration Program: Description, Analysis and Phase I Results," *The Telecommunications and Data Acquisition Progress Report 42-66, September–October 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 115-132, December 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[81] J. Katz, "2.5 Bit/Detected Photon Demonstration Program: Phase II and III Experimental Results," *The Telecommunications and Data Acquisition Progress report 42-70 May–June 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 95–104, August 15, 1982.
http://ipnpr.jpl.nasa.gov/progress_report/

[82] A. Biswas and W. H. Farr, "Laboratory Characterization and Modeling of a Near-Infrared Enhanced Photomultiplier Tube," *The Interplanetary Network Progress Report 42-152, October–December 2002,* Jet Propulsion Laboratory, Pasadena, California, pp.1–14, February 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[83] R. A. Larue, K. A. Costello, G. A. Davis, J. P. Edgecumbe, and V. W. Aebi, "Photon Counting III-V Hybrid Photomultipliers Using Transmission Mode Cathodes," *IEEE Transactions on Electronic Devices*, vol. 44, pp. 672–678, 1997.

[84] R. A. Larue, G. A. Davis, D. Pudvay, K. A. Costello, and V. W. Aebi, "Photon Counting 1060 nm Hybrid Photomultiplier with High Quantum Efficiency," *IEEE Electron Device Letters*, vol. 20, pp. 126–128, 1999.

[85] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics,* J. Wiley and Sons, Inc. (New York, New York), pp. 666–673, 1991.

[86] S. B. Alexander, *Optical Communication Receiver Design,* International Society for Optical Engineers (Bellingham, Washington) and The Institution of Electrical Engineers, pp. 102–109, 1997.

[87] A. Biswas, B. Madden-Woods, M. Srinivasan, V. A. Vilnrotter, and W. H. Farr, "Ground Detectors for Optical Communications from Deep Space," *Proceedings of SPIE: Free-Space Laser Communication Technologies XIV*, [Ed. S. Mecherle], vol. 4635, pp. 72–83, 2002.

[88] S. Takeuchi, J. Kim, Y. Yamamato, and H. Hoag, "Development of a High-Quantum Efficiency Single-photon Counting System, "*Applied Physics Letters*, vol. 74, pp. 1063–1065, 1999.

[89] K. A. McIntosh, J. P. Donnelly, D. C. Oakley, A. Napoleone, S. D. Calawa, L. J. Mahoney, K. M. Molvar, E. K. Duerr, S. H. Groves, and D. C. Shaver, "InGaAsP/InP Avalanche Photodiodes for Photon Counting at 1.06 μm," *Applied Physics Letters*, vol. 81, pp. 2505–2507, 2002.

[90] A. Biswas and W. H. Farr, "Detectors for Ground-Based Reception of Optical Communications from Mars," Paper MH3, IEEE Lasers and Electro-Optics Society (LEOS) Annual Meeting, November 7–11, 2004, Rio Grande, Puerto Rico, United States, 2004.

[91] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Optical Engineering Press, Bellingham, Washington, 1998.

[92] D. L. Snyder, *Random Point Processes*, John Wiley & Sons, New York, New York, 1975.

[93] E. V. Hoversten, R. O. Harger, and S. J. Halme, "Communications Theory for the Turbulent Atmosphere," *Proceedings of the IEEE*, pp. 1626–1650, October 1970.

[94] W. M. Hubbard, "Binary Detection in an Optical Twin Channel Receiver," *IEEE Transactions on Communications Technology*, vol. 19, pp. 221–223, April 1971.

[95] V. A. Vilnrotter, *Optical Receivers Using Rough Reflectors*, JPL Publication 85-25, Jet Propulsion Laboratory, Pasadena, California, May 1, 1985.

[96] P. Negrete-Regagnon, "Practical Aspects of Image Recovery by Means of the Bispectrum," *Journal of the Optical Society of America*, vol. 13, no. 7, July 1996.

[97] R. Gagliardi and S. Karp, Optical Communications, Second Edition, John Wiley & Sons, Inc., (New York, New York), 1995.

[98] V. Vilnrotter and N. Srinivasin, "Adaptive Detector Array for Optical Communication Receivers," *Telecommunications and Mission Operations Progress Report 42-141*, *January–March 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–22, May 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[99] J. Proakis, *Digital Communications*, McGraw-Hill, Inc., New York, New York, 1995.

[100] M. Simon, S. Hinedi, and W. Lindsey, *Digital Communication Techniques*, PTR Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[101] N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[102] R. Sadr, P. P. Vaidyanathan, D. Raphaeli, and S. Hinedi, *Parallel Digital Modem Using Multirate Filter Banks*, JPL Publication 94-20, Jet Propulsion Laboratory, Pasadena, California, August 1994.

[103] A. Gray, *Very Large Scale Integration Architecture for Nyquist Rate Digital Communications Receivers*, Ph. D. Dissertation, University of Southern California, Los Angles, California, May 2000.

[104] V. Vilnrotter, A. Biswas, W. Farr, D. Fort, and E. Sigman, "Design and Analysis of a First Generation Optical Pulse-Position Modulation Receiver," *The Interplanetary Network Progress Report 42-148*, *October–December 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, February 15, 2002.
http://ipnpr.jpl.nasa.gov/progress_report/

[105] R. McDonough and A. Whalen, *Detection of Signals in Noise*, Academic Press, New York, 1995.

[106] C. W. Helstrom, *Elements of Signal Detection and Estimation*, PTR Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[107] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I, John Wiley and Sons, Inc., New York, 1968.

[108] R. Sadr and W. J. Hurd, "Detection of Signals by the Digital Integrate-and-dump Filter With Offset Sampling," *The Telecommunications and Data Acquisition Progress Report 42-91*, July–September *1987*, Jet Propulsion Laboratory, Pasadena, California, pp. 158–173, November 15, 1987. http://ipnpr.jpl.nasa.gov/progress_report/

[109] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[110] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[111] H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Wiley Series in Telecommunications, John Wiley and Sons, New York, New York, 1990.

[112] C. N. Georghiades, "Detecting Random PPM Sequences in the Absence of Sequences in the Absence of Synchronization," *Proceedings of the 22nd Conference on Information Science Syst*ems, Princeton, New Jersey, pp. 159–163, March 1988.

[113] X. Sun and F. Davidson, "Timing Recovery in Free Space Direct Detection Optical Communication Systems with PPM Signaling," Communications, 1989. *ICC 89, BOSTONICC/89, IEEE International Conference on World Prosperity Through Communications*, vol. 1, pp. 428–432, June 1989.

[114] F. Davidson and X. Sun, "Slot Clock Recovery in Optical PPM Communication Systems with Avalanche Photodiode Photodetectors," *IEEE Transactions on Communications*, vol. 37, no. 11, pp. 1164–1172, November 1989.

[115] X. Sun and F. Davidson, "Word Timing Recovery in Direct Detection Optical PPM Communication Systems with Avalanche Photodiodes Using a Phase Lock Loop," *IEEE Transactions on Communications*, vol. 38, no. 5, pp. 666–673 May 1990.

[116] S. Chauchin, L.-Y. Huang, J.-J. Lee, and C.-K. Wang, "A Frame-Based Symbol Timing Recovery for Large Pull-in Range and Small Steady State Variation," *The First IEEE Asia Pacific Conference on ASICs, AP-ASIC '99*, pp. 75–78, August 1999.

[117] G. L. Lui, "PPM Symbol Synchronization in Random Data," *IEEE Military Communications Conference, MILCOM '91, Military Communications in a Changing World*, vol. 3, pp 1047–1053, November 1991.

[118] C. N. Georghiades, "On the Synchronizability and Detectability of Random PPM Sequences," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 146–156, January 1989.

[119] C. N. Georghiades, "On PPM Sequences with Good Autocorrelation Properties," *IEEE Transactions on Information Theory*, vol. 34, no. 3, pp. 571–576, May 1988.

[120] R, Velidi and C. N. Georghiades, "Symbol Synchronization for Optical Multi-Pulse Pulse Position Modulation Systems," *IEEE International Conference on Personal Wireless Communications*, pp. 182–184, August 1994.

[121] R, Velidi and C. N. Georghiades, "Optimal and Suboptimal Frame Synchronizers for Optical Multi-Pulse PPM," *IEEE International Symposium on Information Theory*, p. 79, July 1994.

[122] K. Sato, T. Ohtsuki, I. Sasase, and S. Mori, "Performance Analysis of (m, 2) MPPM with Imperfect Slot Synchronization," *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, vol. 2, pp. 765–768, May 1993.

[123] M. Asano, T. Ohtsuki, H. Uehara, and I. Sasase, "A Novel Frame Synchronization Rule for Optical DPPM Systems with Restriction of Frame Length," *IEEE Global Telecommunications Conference, Communications: The Key to Global Prosperity*, vol. 2, pp. 923–927, November 1996.

[124] D. Jinsong, I. Oka, and C. Fujiwara, "Symbol Error Probability of Time Spread PPM Signals in the Presence of Interference," *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, '10 Years PACRIM 1987-1997 - Networking the Pacific Rim*, vol. 1, pp. 1–4, August 1997.

[125] C. Georghiades, "Optimum Joint Slot and Symbol Synchronization for the Optical PPM Channel," *IEEE Transactions on Communications*, vol. 35, no. 6, pp. 632–636, June 1987.

[126] K. Sato, T. Ohtsuki, H. Uehara, and I. Sasase, "Communication Systems with PPM Signaling Lightwave Technology," *Journal of Lightwave Technology*, vol. 14, no. 9, pp. 1963–1969, September 1996.

[127] J. M. H. Elmirghani and R. A. Cryan, "Jitter Implications on Optical Fibre PPM Performance," *ICC 94, SUPERCOMM/ICC '94, Conference Record, IEEE International Conference on Serving Humanity Through Communications*, vol. 2, pp. 665–669, May 1994.

[128] V. A. Vilnrotter, E. R. Rodemich, and H. H. Tan, "A Synchronization Technique for Optical PPM Signals," *The Telecommunications and Data Acquisition Progress Report 42-86, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 24–31, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[129] S. Aguirre, W. J. Hurd, R. Kumar, and J. Statman, "A Comparison of Methods for DPLL Loop Filter Design," *The Telecommunications and Data Acquisition Progress Report 42-86, July–September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 114–124, November 15, 1986. http://ipnpr.jpl.nasa.gov/progress_report/

[130] S. A. Stephens and J. B. Thomas, "Controlled-Root Formulation for Digital Phase-Locked Loops," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 1, January 1995.

[131] A. Gray, "Parallel Sub-Convolution Filter Bank Architectures," *IEEE International Symposium on Circuits and Systems*, Bangkok, Thailand, May 2003, vol. 4, pp. IV-528–IV-531, 2003.

[132] W. Zhuo and E. Micheli-Tzanakou, "A High Performance Continuous Data Flow Filter Using Sliding Discrete Fourier Transform (DFT) and One Point Inverse DFT," *IEEE International Conference on Information Technology Applications in Biomedicine*, pp. 51–56, May 1998.

[133] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, vol. 9, pp. 14–37, January 1992.

[134] M. A. Soderstrand and R. J. Miller, "High-Speed Convolution Using GQRNS Arithmetic Hardware," *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 1556–1558, 1989.

[135] W. S. Gan and Y. K. Chong, "Parallel Implementation of the Frequency Bin Adaptive Filter for Acoustical Echo Cancellation," *International Conference on Information, Communications and Signal Processing*, vol. 2, pp. 754–757, September 1997.

[136] G. E. Johnson, R. A. Muir, J. M. Scanlan, and W. M. Steedly, "The Sliding Frequency-Domain Adaptive Filter Algorithm Amenable to Parallel Implementation," *IEEE Proceedings of ASILOMAR*-29 (Asilomar, California), vol. 2, pp. 1126–1130, 1996.

[137] M. E. Deisher and A. S. Spanias "Real-Time Implementation of a Frequency-Domain Adaptive Filter on a Fixed-Point Signal Processor," *International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 2013–2016, 1991.

[138] K. Eneman, and M. Moonen, "Filter bank Constraints for Subband and Frequency-Domain Adaptive Filters," *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, p. 4, October 1997.

[139] D. R. Bungard, L. Lau, and T. L. Rorabaugh, "Programmable FFT Processors for Wide-Bandwidth HF Spread-Spectrum Communications and Radar Signal Processing," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1357–1359, 1989.

[140] Y. Elcherif, S. Mashali, and M. El-Serif, "Semi-Systolic Bit Level Arrays for High Speed Convolution," *IEEE International Conference on Systems Engineering*, pp. 391–394, 1989.

[141] H. A. Chow, H. Alnuweiri, and S. Casselman, "FPGA-Based Transformable Computer for Fast Digital Signal Processing," *IEEE Symposium on FPGAs for Custom Computing Machines*, pp. 197–203, 1995.

[142] Yuan-Pei Lin and P. P. Vaidyanathan, "Application of DFT Filter Banks and Cosine Modulated Filter Banks in Filtering," *IEEE Asia-Pacific Conference on Circuits and Systems*, pp. 254–259, 1994.

[143] Y. C. Lim, and B. Farhang-Boroujeny, "Fast Filter Bank (FFB)," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, no. 5, pp. 316–318, May 1992.

[144] A. Klouche-Djedid, "IIR Stable, Causal and Perfect Reconstruction Uniform DFT Filter Banks with a Real or Complex Prototype Filter," *IEEE Colloquium on Digital Filters: An Enabling Technology*, pp. 13/1–13/6, 1998.

[145] N. Fliege, "Polyphase FFT Filter Bank for QAM Data Transmission," *IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 654–657, May 1990.

[146] S. Weiss, L. Lampe, and R. W. Stewart, "Efficient Subband Adaptive Filtering with Oversampled GDFT Filter Banks," *IEEE Colloquium on Adaptive Signal Processing for Mobile Communication Systems*, pp. 411–419, October 1997.

[147] A. P. R. Rodrigues, M. R. Petraglia, and J. Szezupak, "An Adaptive Filtering Subband Structure with Critical Sampling," *IEEE International Symposium on Circuits and Systems, ISCAS '96*, vol. 2, pp. 368–371, 1996.

[148] K. Eneman and M. Moonen, "A Relation Between Subband and Frequency-Domain Adaptive Filtering," *13th International Conference on Digital Signal Processing Proceedings*, vol. 1, pp. 25–28, July 1997.

[149] J. Vollmer, "Analysis and Design of Numerically Controlled Oscillators Based on Linear Time-Variant Systems," *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 453–456, October 1998.

[150] M. M. Al-Ibrahim, "A Multifrequency Range Digital Sinusoidal Oscillator with High Resolution and Uniform Frequency Spacing," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48 no. 9, pp. 872–876, September 2001.

[151] M. M. Al-Ibrahim, "A Simple Recursive Digital Sinusoidal Oscillator with Uniform Frequency Spacing," *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 689–692, May 2001.

Chapter 7
Future Prospects and Applications

Hamid Hemmati and Abhijit Biswas

7.1 Current and Upcoming Projects in the United States, Europe, and Japan

The concept of free-space optical communications was conceived shortly after the invention of lasers. Strides have been made in developing and demonstrating the technology ever since. Early experiments that targeted terrestrial point-to-point, air-to-ground, and space-to-ground links were not fully successful because the technology was immature. Most of these demonstrations were government-funded, both for civilian and military applications.

The promise of laser communication, high data-rate delivery with significantly reduced aperture size for the flight terminal, led to the continued funding for the successful experiments and provided the incentive for further demonstrations. Table 7-1 presents a chronological summary of major successful laser-communication technology demonstrations [1–5] to or from air or space. Plans for additional major experiments are discussed below.

7.1.1 LUCE (Laser Utilizing Communications Experiment)

The Optical Inter-orbit Communications Engineering Test Satellite (OICETS) carrying the LUCE payload is planned for launch into low Earth orbit (LEO) in 2005 (Fig. 7-1). LUCE has an aperture diameter of 26 cm and is equipped with 200-mW 847-nm diode lasers for 50 megabits per second (Mbps) transmission to the European Space Agency's (ESA's) Advanced Relay

Table 7-1. Summary of major accomplishments in laser-communications technology.

Year	Experiment	Performing Organizations*
1980	AFTS (Airborne Flight Test System) 1 gigabit per second (Gbps) link from aircraft [1]	McDonnell Douglas (U.S. DoD)
1990	RME (Relay Mirror Experiment) Precision laser beam pointing [2]	Ball Aerospace (U.S. DoD)
1992	GOPEX (Galileo Optical Experiment) Uplink from Earth to deep-space [3]	JPL (NASA)
1995	LCE (Laser Communication Experiment) Bi-directional link from terminal at geosynchronous Earth orbit (GEO) [4]	CRL (Japan) and JPL (NASA)
1998	GEOLite (Geosynchronous Lightweight Technology Experiment) Multi-Gbps link from GEO orbit [5]	Lincoln Laboratory (U.S. DoD)
1998	SILEX (Semiconductor Intersatellite Link Experiment) Low Earth orbit (LEO) to GEO and LEO- and GEO-to-ground [6]	ESA

* Organizations Listed: Communication Research Laboratory (CRL), U.S. Department of Defense (DoD), European Space Agency (ESA), Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA)

and Technology Mission Satellite (ARTEMIS). It is capable of receiving 2.048-Mbps links from ARTEMIS at 819 nanometers (nm) [7–8].

7.1.2 Mars Laser-Communication Demonstrator (MLCD)

NASA is planning for the first deep-space laser-communication downlink in the 2009–2011 time frame from Mars distances utilizing the Mars Laser Terminal (MLT) being built by the Massachusetts Institute of Technology (MIT) Lincoln Laboratory flying aboard the Mars Telecom Orbiter spacecraft [9]. MLCD will demonstrate data rates on the order of 1 to 80 Mbps from the longest distance (about 2.4 astronomical units [AU]) to the shortest distance (about 0.67 AU), assuming a 5-equivalent-diameter aperture. This data rate is at least an order of magnitude higher than state-of-the-art RF Mars communication systems.

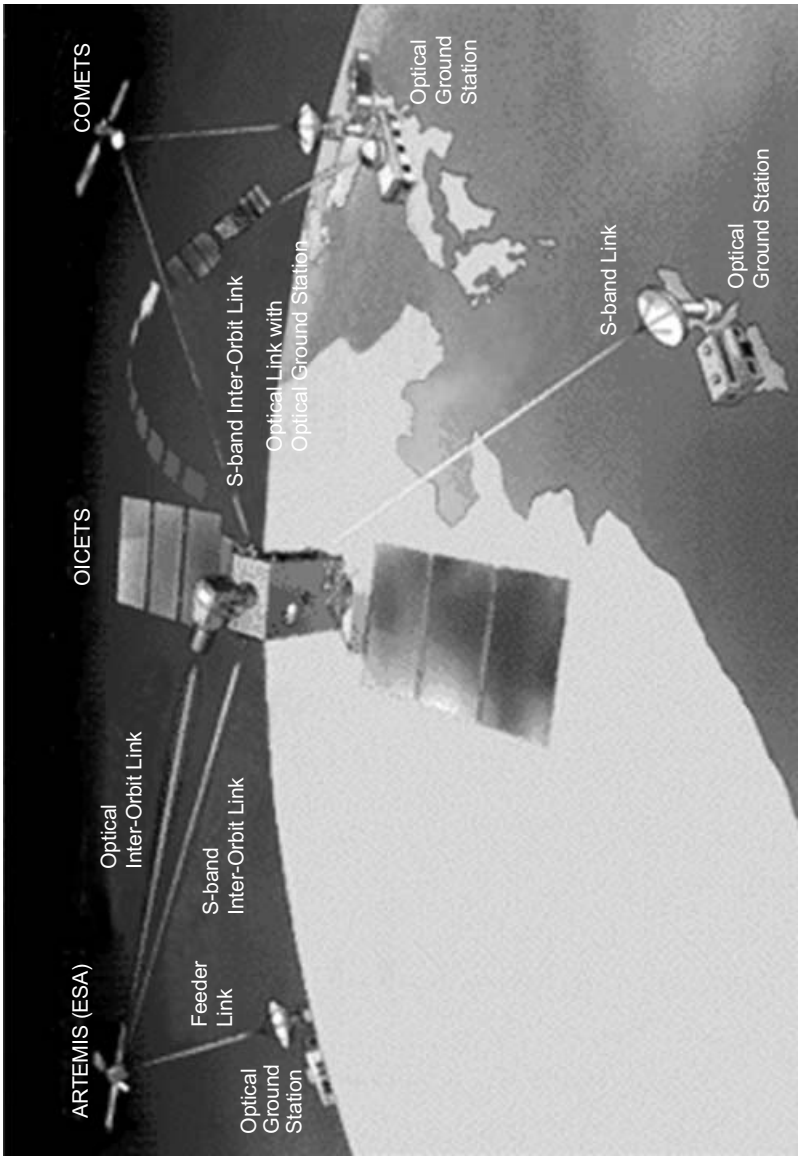


Fig. 7-1. Artist concept of the laser communications terminal on Japan's OICETs spacecraft communicating with the European Space Agency ARTEMIS and COMETS spacecraft (COMETS = Communications and Broadcast Engineering and Test Satellite).

7.2 Airborne and Spaceborne Receivers

Use of ground-based receivers was discussed extensively in Section 5.2. Here, we briefly discuss the merits of airborne and spaceborne receivers. These alternative options will offer significant advantages over ground-based systems when made practical through technology development and validation to the extent where cost, reliability, and redundancy against single point failure (for spaceborne receivers) become attractive. A more quantitative description of the advantages is provided below.

7.2.1 Advantages of Airborne and Spaceborne Receivers

The main advantage of airborne and spaceborne receivers is that they are above the clouds and most if not all of the atmosphere. This increases link availability and removes the atmospheric-turbulence-imposed limitation of operating many times diffraction limited for both transmission and reception. Dramatically reduced sky background contributed by scattering of sunlight from atmospheric constituents also benefits the data-receiving function. As a result, with such a platform, the required equivalent aperture size is considerably less than that required for ground-based terminals. The collection area is a function of the platform altitude. For example, relative to a ground-based terminal, a nearly 35 percent reduction in the required aperture diameter is expected with an airborne terminal located at a 20-km altitude. Similarly, a reduction of nearly 50 percent in aperture diameter can be expected when utilizing a spaceborne platform for the receiver.

Optical receivers also need the capability to transmit laser signals to the deep-space assets. The laser transmissions are needed to provide beacon-pointing reference sources and/or provide uplink commands. Intuitively it appears that airborne and spaceborne assets utilizing a single platform for both receiving and transmitting would be cost effective; however, issues relative to transmit–receive isolation and the point-ahead required will need to be carefully considered. One of the overriding advantages of placing the transmitting lasers above the clouds and the majority of the atmosphere will be removal of the severe limitations induced by turbulence on transmitting lasers from the ground. The benefits could be of the order of 30 dB.

In general, taking most if not all the atmosphere out of the optical channel involved in a deep-space communication link opens the possibility of near-diffraction-limited performance. The important ramifications of this are the possibility of overcoming performance penalties associated with atmosphere-induced limitations on how small the communications detector field of view can be. This of course has a two-fold effect of increasing background and limiting detector bandwidth.

Allowing near-diffraction-limited operations also opens up the possibility of implementing coherent communications. Coherent communications are immune to background contamination, and even though sky radiance backgrounds are not of concern in this case, the possibility of communicating at very small SEP/SPE (Sun–Earth–probe/Sun–probe–Earth) angles becomes possible. This is significant for a variety of deep-space configurations, reducing outages near solar conjunctions as well as allowing greater tolerance to having stars or planets in the detector field of view. Stabilized lasers used for coherent techniques also open up new possibilities for utilizing the laser communications sources for novel light science investigations since detection becomes sensitive to the optical phase.

The extent to which near-diffraction-limited performance can be embraced will be tempered by the associated stringent pointing control requirement. This, in turn, will be influenced by the stability and quiescence of the airborne/spaceborne platform, as well as the cost of flight-qualified large optics. One can speculate that a work-around may be to limit aperture size but to have many apertures. Use of multiple apertures will likely result in performance inferior to a single aperture; however, the cost versus performance of adopting an array architecture in space needs to be evaluated. The added advantage of deploying arrays in space is the redundancy they provide, eliminating a single point of failure

7.2.2 Disadvantages of Airborne and Spaceborne Receivers

The disadvantages for airborne and spaceborne receivers need to be distinguished from each other. Airborne systems can be re-deployed multiple times and are therefore readily accessible for hardware reconfiguration; however, there are concerns about platform attitude control and stability and their influence on pointing stability, as well as field-of-view blockage. To first order, the pointing stability can be considered as comprised of coarse and fine components. The coarse pointing will need to be a fraction of the attitude uncertainties thought to be of the milliradian (mrad) class, whereas the fine pointing will need to be a fraction of the diffraction-limited spot size, which can be anywhere from tens to thousands of nanoradians (nrad) depending on how many times diffraction limited the terminal design is. In fact, for airborne systems, residual turbulence may still require a 5- to 10-μrad class of communications fields of view requiring approximately 0.1- to 1-μrad-class pointing accuracy.

Conversely, spaceborne platforms require redundancy built in to prevent potential single-point failures. Moreover, there is the question as to where spaceborne receivers need to be deployed in order to provide maximum coverage of deep-space assets. For an operational receiver system, a constellation of orbiting receivers may be required. In addition, it will be

necessary to get the data from these assets back to Earth and/or to provide adequate buffering at the receiver. Placing optical receivers at Lagrange points or the Moon is also viable. Again the design chosen to provide very high performance will dictate the amount of pointing and platform stability control required. Here cost performance trade-offs are needed.

Another cost trade-off that should be carefully evaluated is the comparison of implementing adaptive optics systems on the ground to compensate for atmospheric turbulence versus deploying systems in space. In other words, if adaptive optics that largely compensated for atmosphere-imposed limitations could be implemented, then the only advantage of going airborne or spaceborne would be getting above the clouds. Once again the question needs to be answered as to the cost of implementing a ground network with adaptive optics versus an airborne or spaceborne system. The answer will require further study and evaluation of ongoing and emerging technologies in diverse areas and the costs associated with them. Many of the limitations discussed here are engineering problems and are certain to be alleviated with time, making them viable platforms to host the receive terminal.

7.2.3 Airborne Terminals

The airborne platforms include balloons, airships, and airplanes. The best reception availability is obtained when the airborne terminals are located in the most southerly or most northerly latitudes because line-of-sight blockage by the Earth is minimized.

7.2.3.1 Balloons. The balloon platform instabilities may be overcome with a gimbal-mounted receiver telescope. However, the location of the balloon itself has to be limited for the purpose of data relay with the ground. For example, scientific, free-flyer balloons that are wind driven are not useful to this application. The required number of balloons is largely dependent on the field-of-view restrictions for the optics. Since current balloons accommodate the payload on tethers below the balloon, field of view is often severely restricted. Balloons made of materials that are transparent at the signal wavelength may allow a see-through capability to the payload, albeit with some signal attenuation. Currently available tethered balloons have a very small payload weight capacity and require technology development in order to serve as good candidates for optical communication receiver host platforms.

A 20-km-altitude tethered balloon-based laser communications receiver called Space Relay Communication Link (SPARCL) has been conceptually analyzed [10]. An altitude of 20 km will be sufficient to circumvent much of the atmospheric effects. This concept assumes that the laser-communication payload is mounted on top of the balloon. One of the main challenges identified at the operational altitude is when the balloon encounters high wind regions. In

that case, active control of the balloon was deemed necessary. This study suggests that use of a high-altitude balloon-based receiver concept is technically feasible; however, significant technology development is necessary before the receiver can be used operationally.

7.2.3.2 Airships. Station keeping is a major requirement for an airship serving as the platform for a receiving terminal. For that task, the aerodynamic drag on the airship has to be kept to a minimum. Typically, the airships are designed for minimum drag. Therefore, the terminal has to be located within the airship. Airships could have a lifetime of many years and are capable of landing for recovery. Limited radio-frequency (RF) communication with the airship is provided. Instrument-available electrical DC powers exceeding 1 kW are planned for the airships that are under development [11]. Tethered aerostats can carry heavy payloads. The maximum altitude for current tethered airships is about 6 km [11]. This altitude is too low to fully mitigate the atmospheric effects (e.g., clouds).

For any of the airborne dirigibles that are engine driven, air turbulence effects generated by the engine fans must be taken into account.

7.2.3.3 Airplanes. The altitude of 12.5 to 13.7 km still has 20 percent of the air molecules of sea level so airplanes can still fly, but it is above most water vapor so infrared observations can be made that are impossible for ground-based stations. Consequently, there has been a string of aircraft-borne infrared observatories since the mid 1960s [28]. First, a NASA Convair CV 990 had telescopes pointed out a window. Then, a NASA Lear jet was fitted with a 30-cm telescope in place of an emergency exit window. In 1974, a United States Air Force C-141 Stratolifter had a 91.5-cm telescope mounted in front of the left wing to become the Kuiper Airborne Observatory (KAO). In 2005, operations are set to begin with the joint NASA space agency Stratospheric Observatory for Infrared Observation (SOFIA), a 2.5-m telescope mounted on a Boeing 747SP [29].

None of these airborne facilities were developed with optical deep-space communication in mind. Nevertheless, their stratospheric location mitigates most atmospheric effects, particularly clouds. Moreover, their evolution has proceeded along a learning curve in dealing with the major problems of maintaining pointing accuracy and instrument function despite vibration, wind gusting, and temperatures in the range of –40 deg Celsius.

The limitation for airplane-borne facilities is cost for fueling and maintaining the facility. This is especially so considering that even SOFIA's 2.5-m telescope is at the low end of the size range needed for deep space communication.

7.2.4 Spaceborne Receiver Terminals

Unaffected by atmospheric effects, spaceborne terminals can theoretically provide availability exceeding 98 percent. Spacecraft located at the most favorable GEO and medium Earth orbit (MEO) provide higher availability relative to those in favorable LEO orbits. The required spacecraft-pointing capability, on the order of half of the field of view (approximately ±1 mrad), is well within the capability of current spacecraft platforms.

Several studies have been conducted in the past on the feasibility and costs of spaceborne, Earth-orbiting communication relay satellites. In a 1993 NASA-funded study, TRW and Stanford Telecom conceptually designed and costed a space-based transceiver for optical communications [12,13]. JPL's Advanced Project Design Team studied both direct-detection and coherent-detection configurations for an optical relay satellite. A comparison of the results of these past studies is given in [14]. JPL based its cost estimation on a combination of grass roots estimates and quotes for mission operations, the launch vehicle, and the various spacecraft subsystems. Cost models were used for other mission components, including payload, systems engineering, integration and test, management, and reserves. To reduce the cost, the Next Generation Space Telescope technology development heritage was assumed for the front-end optical signal collection aperture. JPL's Advanced Project Design Team study showed that the most probable cost of a single 7-m direct-detection telescope on the relay satellite was a factor of two less than the previous estimates [14].

This cost now exceeds that of an eight-station ground-based facility [6]. Moreover, a single spaceborne station is limited in coverage, whereas an eight-receiver ground-based station provides full coverage of the spacecraft. The majority of the cost is for the host spacecraft and the launch vehicle. To make the spaceborne receiver attractive relative to the ground-based receivers, innovations in the technology of lightweight, low-cost telescopes (photon collectors) are required to minimize the overall cost per spacecraft.

7.2.5 Alternative Receiver Sites

A third category of receiver/transmitter station sites is the Moon. As with space-based receivers, such sites will avoid the atmospheric effects. However, maintenance and upgrades are cost prohibitive at this time. Moon-based radio telescopes have been studied in some detail [15].

7.3 Light Science

In a manner analogous to traditional radio-science measurements, "light-science" measurements are possible through use of the laser beam transmitted from a spaceborne laser-communication terminal for positional reference and light propagation experiments. Several preliminary studies have been made into

viable scientific applications of laser communications [16,17]. Some of these are (1) light-propagation experiments that include occultation investigations of probe planetary limbs and scattering from the medium throughout interplanetary space; (2) enhanced knowledge of Solar-System body (e.g., planet, moon asteroid, or comet) properties; (3) tests of fundamental theories of physics; and (4) improved knowledge of Solar-System ephemerides. Some of these measurements are unique to optical communication technologies and the application of today's state-of-the-art tracking capability. Many science measurements can be made with incoherent systems (which may include pulsed laser sources), while others require (or are more precise with) coherent systems.

Some of the possible light-science measurements are discussed in further detail below.

7.3.1 Light-Propagation Experiments

Light-propagation experiments include occultation and interplanetary light-scattering. In general these experiments can be designed to detect intensity or phase. Each type of sensing imposes requirements on the laser source used by the laser communications system. For example, frequency-stabilized lasers used for phase-sensitive detection may not be easily useable if the receiver is ground-based and limited by atmosphere-induced turbulence and background. Furthermore, light-propagation experiments from ground-based measurements will need to have reliable independent means of calibrating the atmospheric attenuation. Use of suitable celestial sources in the vicinity of the laser communication terminal is a possibility. However, being able to perform light-science measurements where received signal can be easily correlated to phenomena in the intervening medium, in the absence of atmospheric perturbations, offers clear advantages.

7.3.2 Occultation Experiments to Probe Planetary Atmospheres, Rings, Ionospheres, Magnetic Fields, and the Interplanetary Medium

In occultation experiments, laser light from the flight terminal on one spacecraft is received by a second spacecraft or by an Earth-based terminal as the transmitting spacecraft passes behind the limb, atmosphere, or rings of a planet. As the transmitting spacecraft is occulted, its laser is observed (i.e., received) by the receiving terminal. The laser output from the flight terminal can be either continuous or pulsed at high repetition rates. Detecting intensity alone would involve looking for deterministic changes in received average power that offer a clearly discernible signature representative of attenuation or refractive bending.

7.3.2.1 Atmospheric Occultations. Characteristics of the atmosphere around a planet or other bodies in space can be determined through measurements of a flight-terminal laser's attenuation or complete occultation due to the atmosphere [18,19]. In addition to direct intensity detection, wavelength perturbations due to Doppler could in principle be sensed, allowing for simultaneous tracking of the spacecraft velocity component along the line of sight and of the average laser signal. For example, for a high polar orbit around Mars, a 1000-nm laser would undergo a ±2.5-GHz change due to Doppler variations. While it is non-trivial to sense 1/1000th of a nanometer and smaller changes in wavelength, and would require a laser line width that is much narrower, the interesting possibility of extracting signatures of average laser power fluctuation as a function of position in orbit presents itself. Refractive bending could thereby be extracted.

When two flight optical-communication terminals are used, it becomes possible for one terminal to acquire occultation data through reception of laser light from the second flight terminal. If a flight terminal transmits two or more different coherent wavelengths, in conjunction with a suitable space (or ground) receiver, relatively precise atmospheric data can be extracted from the relative phase delay of each received frequency. The precision of this method can be further enhanced if a second but un-occulted flight terminal in the same vicinity is used as a reference.

In atmospheric-occultation experiments, typical goals are to determine temperature and pressure as functions of altitude in the stratosphere and troposphere, determine composition (e.g., methane and helium abundances), characterize the vertical structure of the ionosphere, and investigate turbulence and other irregularities.

7.3.2.2 Ring-Investigation Experiments. In ring-occultation experiments, typical goals are to determine the size and size distribution of ring particles, radial structure of the ring system, and vertical structure of the rings (e.g., whether ring structures are widely distributed or confined to a plane). As in atmospheric-occultation experiments, this is traditionally done in radio frequencies by transmitting two wavelengths coherently from the spacecraft to the ground. Depending upon the proximity of a laser-communications-bearing spacecraft to such rings, backscattered light intensity patterns and the intensity and angular characteristics could be utilized to investigate the density, size, and shape of particles in the Mie-scattering-sized regime.

7.3.3 Enhanced Knowledge of Solar-System-Object Masses and Gravitational Fields, Sizes, Shapes, and Surface Features

An optical terminal provides means to gather information about gravitational fields and the sizes, shapes, and surface features of intercepted

planets, moons, and other interplanetary objects. Gravitational fields are measured through observing changes in the spacecraft trajectory and velocity. These changes are measured by astrometry techniques, by observing Doppler effects on the received wavelength, and (in the case of a pulsed optical transmitter) by observing the timing of the received laser pulses. The sizes and shapes of moons and other objects, as well as some surface features, can be determined through multiple occultations. Surface features can also be measured through sounding the surface of an object with the flight terminal laser, and then receiving the reflected light with either the same or a different flight terminal receiver. However, practicality of the sounding technique is dependent on surface and atmospheric characteristics of the object observed.

7.3.3.1 Improved Knowledge of Solar-System Body Properties. Precise optical tracking of a spacecraft trajectory during target approach can provide information about the mass and gravity field of the target. Similarly, precise tracking of a spacecraft as it passes through a system of satellites can aid in the calculation of relative center-of-mass locations and velocities. In a hybrid system with both RF and optical capabilities, remote optical tracking data would be a valuable supplement to conventional RF Doppler and very long baseline interferometry (VLBI) data used to determine target masses, positions, and motions. Such optical observations may rival or surpass conventional RF data in their accuracy and usefulness for some measurements.

The capability of precise spatial tracking of a spacecraft, together with improved ephemerides for planets and satellites, would put navigational abilities on a whole new level. Collision avoidance, now a primary concern, would give way to a mode that facilitates closer approaches to planets and other Solar-System objects. This capability would also refine trajectory and encounter sequences to conserve propulsion fuel.

7.3.3.2 Optical Reference-Frame Ties. Precise spatial tracking of a laser-carrying spacecraft might also be used to help tie together optical star reference frames, which typically are more accurate in right ascension than in declination. By tracking the spacecraft as it moves through several values of declination, and assuming a continuous Kepler trajectory, one could calculate the angular distance between separated reference stars.

7.3.4 Tests of the Fundamental Theories: General Relativity, Gravitational Waves, Unified Field Theories, Astrophysics, and Cosmology

A suitably configured optical flight terminal provides tools of unprecedented accuracy for testing the fundamental theories of general

relativity, gravitational waves, unified field theories, astrophysics, and cosmology.

7.3.4.1 Tests of General Relativity and Unified Field Theories, Astrophysics, and Cosmology. Several experiments that would require or benefit from a laser communications capability on spacecraft are possible. These include light-deflection tests of general relativity, gravitational-wave detection, tests of the change with time of the gravitational constant G, and observation of a gravito-magnetic interaction. Some of these tests will be accomplished by precise astrometric measurements [20], while others require precise laser Doppler and/or ranging data.

In the areas of astronomy, astrophysics, and cosmology, there are numerous scientific goals that would benefit from coherent laser communications technologies in space. Several goals currently being pursued with astrometric interferometers [20,21] include refinement of the cosmic distance scale and of the mass estimate for our galaxy, and a search for other planetary systems. These interferometric instruments use laser metrology systems to control systematic errors. Another scientific goal is improved understanding of the composition, concentration, and velocity of interplanetary and cosmic dust. For this, coherent optical Doppler techniques will be useful.

7.3.4.2 Effects of Charged Particles on Electromagnetic Wave Propagation, Including Test of l/f Hypothesis. An important existing problem in spacecraft navigation and orbit determination is inconsistencies between Doppler and range data. The inconsistency between Doppler and range data is attributed to incomplete modeling of non-gravitational accelerations on spacecraft, e.g., solar pressure and non-ideal thruster behavior such as leaks or exhaust-plume impingement on parts of the spacecraft.

While it is probable that the nongravitational accelerations are the cause for most, if not all, of the range–Doppler inconsistencies, by deweighting range, one masks any other effects that might exist and that would also lead to inconsistencies. Such effects could be of considerable scientific interest. One such effect is that the calibration of the effects of free electrons along the transmission path on electromagnetic signals is incorrect. Such free-electron-induced perturbations have a 1/frequency (1/f) dependence. Therefore, testing with higher frequency optical signals will reduce this uncertainty.

7.3.5 Enhanced Solar-System Ephemerides

By comparing actual optical range data with predictions based on current ephemerides, the accuracy of Solar-System ephemerides can be greatly enhanced. One benefit of light science is the potential for making real-time

angular measurements from a single station, in contrast to the long passes and multiple stations required by radio-frequency VLBI. Another benefit is the possibility of direct spacecraft–target tracking with Solar-System bodies.

Candidate optical angle-tracking techniques used in astrometry show promise of nanoradian (about 1/5,000 arcsec) angular accuracy from the ground and picoradian (prad) (1/5,000,000 arcsec) accuracy from space for measurements of relative angular position between a flight-terminal laser and sufficiently bright, point-like targets [18,22,23,27]. For target-relative measurements, the extended discs and non-uniform brightness of planetary bodies may limit achievable tracking accuracies to a few tens of nanoradians—a few tens of kilometers at Saturn, for example. However, the better tracking accuracies possible for measurements between point sources can potentially be exploited for other important mission enhancements, such as remote optical tracking of landers, rovers, and orbiters.

Astrometric telescopes currently in use on the ground have demonstrated night-to-night reproducibility for differential angular measurements that are 20 nrad or better for point-like stars of apparent visual magnitude 11 and brighter [18,24]. For reference, a 5-W visible-wavelength (532-nm) laser at Saturn firing through a 30-cm telescope would have an apparent visual magnitude of about 11.

7.3.5.1 Science Benefits of Remote Optical Tracking: Ephemeris Improvement. Optical measurements of the angular separation between a laser-carrying spacecraft and a Solar-System body (e.g., planet, satellite, or asteroid) made remotely from Earth or Earth orbit would complement traditional RF data types based on range, Doppler, and quasar-relative VLBI measurements [25,26].

An important implication of a remote optical tracking capability is that the target position could be estimated accurately prior to encounter in all three dimensions, two by onboard optical measurements and the third (along the spacecraft trajectory) by a remote measurement.

With a laser on the spacecraft, the angular separation between target and spacecraft might be observed optically from the vicinity of Earth (this assumes an Earth–asteroid separation of about 3 AU, a spacecraft–target relative velocity of about 7 km/s, and a 45-degree angle between the trajectory and the Earth–asteroid line of sight). The remote data and the onboard data are complementary because of the different viewing angles, and they would be of comparable accuracy if ground-based astrometric techniques could achieve accuracies of 25–50 nrad for angle measurements between an asteroid and a laser-carrying spacecraft.

A smaller ephemeris improvement should be expected for a target such as Saturn, because of its extended disk (~100-μrad angular diameter) and

increased distance. To rival the 10-km metric accuracy of onboard optical data, ground-based data must be capable of 10-nrad accuracy for measurements of the angular separation between the spacecraft and Saturn. The primary obstacle to a precise measurement of relative position is calibration of the offset between Saturn's center of mass and its geometric center of brightness.

For large (~0.5-arcsec) and nearly circular targets such as Titan or Saturn, limb-fitting techniques can be used to deduce the geometric center and approximate center of mass. Such techniques have the advantage of being relatively insensitive to albedo variations. Center-of-mass information can then be refined through use of the tracking data obtained prior to and during the encounter(s).

Judged from their performances to date and predicted capabilities, astrometric telescopes that use Ronchi rulings and have fields of view of at least a few milliradians appear to be suitable instruments both for aiding in calibration of offsets between center of brightness and center of mass for extended bodies and for use with target-relative navigation and tracking. Interferometers, on the other hand, are not well suited to measurements on extended sources because the fringe overlap degrades fringe contrast, making a bright, extended body appear quite dim.

An onboard optical communication system, if appropriately configured with imaging capability, can also provide means for optical navigation. More study is needed to determine how accurately remote optical navigation and tracking techniques can be made to work with Solar-System bodies and to determine the optimum techniques and navigation strategies.

7.3.6 Applications of Coherent Laser Communications Technology

Traditional occultation experiments on Solar-System bodies, rings, or atmospheres have used coherent radio-frequency communications systems at two or more wavelengths. For example, the Voyager spacecraft transmitted X- and S-bands to Earth through the atmosphere and rings of other planets. Some of the potential science benefits offered by optical analogs of these coherent radio measurements are briefly covered below.

Coherent laser communication technology is applicable to the development and deployment of optical flight terminals. Development of a flight terminal also presents the opportunity for advances in coherent laser communication technology and its application to light science. The optical flight terminal provides a useful test bed for optical communication and other extremely long-range applications of laser technology.

To perform light science with coherent laser communications payloads will require frequency-stabilized lasers with sub-kilohertz line widths, 1–10 W average power, and phase-matching (or at least frequency-matching)

transponders. The narrowband filtering intrinsic to coherent detection permits communications and tracking under conditions of much higher background light (noise) than what is possible with incoherent detection.

Sensitive range and Doppler data capable of detecting small perturbations to the relative separations of two co-orbiting satellites above Earth (or any Solar-System object) would permit inference of the gravitational field with excellent spatial resolution. For Earth, such measurements would contribute to a better understanding of plate tectonics and continental drift.

7.4 Conclusions

In conclusion, it would be fair to say that optical communication, as an operational capability, holds much promise. Technology demonstrations in the near future will retire the risks of implementing a future communications capability. At that time, exploiting the potential benefits of performing light science will inevitably be explored.

References

[1] R. S. Mason and G. M.Lanham, "Free Beam Laser Communications System Technology Development," (A82-13001 03-35), *Electro-optics/Laser 80 Conference and Exposition*, Boston, Massachusetts, November 19–21, 1980, Chicago, Illinois, Industrial and Scientific Conference Management, Inc., pp. 278–294, 1980.

[2] P. A. Lightsey, "Scintillation in Ground-to-Space and Retro-Reflected Laser Beams," *Optical Engineering*, vol. 33, pp. 2535–2543, August 1994.

[3] K. E. Wilson, J. Lesh, and T. Yan, "GOPEX: A Laser Uplink to the Galileo Spacecraft on its way to Jupiter," *Free-Space Laser Communication Technologies V*, *Proceedings of SPIE*, G. S. Mecherle, ed., vol. 1866, pp. 138–146, August 1993.

[4] K. Araki, M. Toyoshima,, T. Takahashi, T. Fukazawa, M. Toyoda, M. Shikatani, and Y. Arimoto, "Experimental Operations of Laser-Communication Equipment Onboard ETS-VI Satellite," *Free-Space Laser Communication Technologies IX*, *Proceedings of SPIE*, G. S. Mecherle, ed., vol. 2990, pp. 264–275, May 1997.

[5] http://www.wtec.org/loyola/satcom2/03_06.htm

[6] T. Tolker-Nielsen, B. Demelenne, and E. Desplats, "In Orbit Test Results of the First SILEX Terminal," *Free-Space Laser Communication Technologies XI*, *Proceedings of SPIE*, G. S. Mecherle, ed., vol. 3615, pp. 31–42, April 1999.

[7] K. Nakagawa, A. Yamamoto, and M. Toyoda, "Performance Test Result of LUCE (Laser Utilizing Communication Equipment) engineering model," *Free-Space Laser Communication Technologies XII*, *Proceedings of SPIE*, G. S. Mecherle, ed., vol. 3932, pp. 68–76, May 2000.

[8] Y. Suzuki, K. Nakagawa, T. Jono, and A. Yamamoto, "Current Status of OICETS Laser Communication Terminal Development," *Free-Space Laser Communication Technologies IX*, *Proceedings of SPIE*, G. S. Mecherle, ed., vol. 2990, pp. 31–37, May 1997.

[9] D. Boroson, A. Biswas, and B. L. Edwards, "MLCD: Overview of NASA's Mars Laser Communications Demonstration System," *Free Space Laser Communications Technologies XVI, Proceedings of SPIE*, G. S. Mecherle, ed., vol. 5338, pp. 16–22, January 2004.

[10] S.S. Badesha, "SPARCL: A High Altitude Tethered Balloon-Based Optical Space-to-Ground Communication System, *Proceedings of SPIE*, vol. 4821, pp. 181–193, 2002.

[11] http://www.lockheedmartin.com/akron/protech/aeroweb/aeroindx.htm

[12] *DRSS Concept Design and Technology Report*, TR 93062–TR 93068, Stanford Telecom, Reston, Virginia, July 7, 1993.

[13] G. S. Mehrle, W. Akle, C. Starkus, and J. Klein, "Direct detection Optical Relay Satellite for Deep-Space Communication," *Free Space Laser Communication Technologies VI, Proceedings of SPIE*, vol. 2123, pp. 134–155, 1994.

[14] K. E. Wilson, M. Wright, R. Cesarone, J. Ceniceros, and K. Shea, "Cost and Performance Comparison of an Earth-Orbiting Optical Communication Relay Transceiver and a Ground-Based Optical Receiver Subnet," *The Interplanetary Network Progress Report 42-153*, *January–March 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–12, May 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[15] http://radio.astro.gla.ac.uk/antenna.ps

[16] B. Schumaker, *Scientific Applications of Frequency-Stabilized Lasers in Space*, JPL Publication 90-50, Jet Propulsion Laboratory, Pasadena, California, December 1990.

[17] K. Sokolowski and J. Lesh, "Deep Space Optical Communications," *Optical Systems for Space Applications*, *Proceedings of the Meeting*, March 30–April 1, 1987, The Hague, Netherlands, pp. 172–177, January 1987.

[18] G. Gatewood, "The Multichannel Astrometric Photometer and Atmospheric Limitations in the Measurement of Relative Positions," *Astronomical Journal*, vol. 94, pp. 213–224, July 1987.

[19] G. Gatewood, L. Breakiron, R. Goebel, S. Kipp, J. Russel, and J. Stein, "On the Astrometric Detection of Neighboring Planetary Systems. II," *Icarus*, vol. 41, pp. 205–231, February 1980.

[20] R. Reasenberg, R. Babcock, J. Chandler, M. Gorenstein, J. Huchra, M. Pearlman, I. Shapiro, R. Taylor, P. Bender, A. Buffington, B. Carney, J. Hughes, K. Johnston, B. Jones, and L. Matson, "Microarcsecond Optical Astrometry—An instrument and Its Astrophysical Applications," *Astronomical Journal*, vol. 96, pp. 1731–1745, Nov. 1988.

[21] D. Mozurkewich, D. Hutter, K. Johnston, R. Simon, M. Shao, M. Colavita, D. Staelin, B. Hines, J. Hershey, J. Hughes, and G. Kaplan, "Preliminary Measurements of Star Positions with the Mark III Stellar Interferometer," *Astronomical Journal*, vol. 95, pp. 1269–1277, April 1988.

[22] M. Co1avita, M. Shao, and D. Staelin, "Atmospheric Phase Measurements with the Mark III stellar Interferometer," *Applied Optics*, vol. 26, pp. 4106–4112, October 1, 1987

[23] A. Buffington and M. R. Geller, "Measurements of Atmospheric Seeing with a Ronchi Telescope," *Bulletin of the American Astronomical Society*, vol. 18, p. 62, 1986.

[24] D. G. Monet and C. C. Dahn, "CCD Astrometry I. Preliminary Results from the KPNO 4-m/CCD parallax Program," *Astronomical Journal*, vol. 88, pp. 1489–1507, October 1983.

[25] R. Treuhaft and G. Lanyi, "The Effect of the dynamic Wet Troposphere on radio Interferometric Measurements," *Radio Science*, vol. 22, pp. 251–265, March–April 1987.

[26] K. Nordtvedt, "Gravitomagnetic Interaction and Laser Ranging to Earth Satellites," *Physical Review Letters*, vol. 61, issue 23, pp. 2647–2649, December 5, 1988.

[27] M. Co1avita, M. Shao, and D. Staelin, "Two-color method for optical astrometry—Theory and preliminary measurements with the Mark III stellar interferometer," *Applied Optics*, vol. 26, pp. 4113–4122, October 1, 1987.

[28] S. Stephens, "Flying Telescopes," *Mercury*, *Astronomical Society of the Pacific*, vol. 31, pp. 23–31, May/June 2002.

[29] J. Wolf, S. C. Casey, and J. A. Davidson, "Stratospheric Observatory for Infrared Astronomy—SOFIA," *Proceedings of SPIE*, *Infrared Spaceborne Remote Sensing XI, August 6–8, 2003, San Diego, California*, vol. 5152, pp. 61–69, 2003.