Geodesy in the Year 2000

Committee on Geodesy, National Research Council
ISBN: 0-309-53695-2, 186 pages, 8.5 x 11, (1990)

Geodesy in the Year 2000

Committee on Geodesy
Board on Earth Sciences and Resources
Commission on Physical Sciences, Mathematics, and Resources
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.1990

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

The Overview and Recommendations of this report have been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The contributed papers represent the opinions of their authors only, and appearance in this volume does not constitute endorsement of their conclusions by the National Research Council or the Committee on Geodesy.

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

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

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

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

Support for this project was provided by National Geodetic Survey, National Oceanic and Atmospheric Administration; Defense Mapping Agency; Air Force Office of Scientific Research; National Aeronautics and Space Administration; and Department of Energy.

Library of Congress Catalog Card No. 89-63530
International Standard Book Number 0-309-04145-7
Additional copies of this report are available fromNational Academy Press2101 Constitution Avenue, N.W. Washington, D.C. 20418

S055
Printed in the United States of America **First Printing, February 1990Second Printing, December 1990**

COMMITTEE ON GEODESY

JOHN RUNDLE, Sandia National Laboratory, *Chairman*
TIMOTHY H. DIXON, Jet Propulsion Laboratory
CLYDE GOAD, Ohio State University
ERNEST H. METZGER, Bell Aerospace Textron
J. BERNARD MINSTER, University of California, San Diego
RICHARD SAILOR, The Analytic Sciences Corporation
ROSS STEIN, U.S. Geological Survey

Staff

HYMAN ORLIN, Consultant

Liaison Members

WILLIAM H. CHAPMAN, U.S. Geological Survey
MIRIAM BALTUCK, National Aeronautics and Space Administration
DONALD H. ECKHARDT, Air Force Geophysical Laboratory
CHARLES HALL, Defense Mapping Agency
GEORGE KOLSTAD, U.S. Department of Energy
JOHN G. HEACOCK, Office of Naval Research
MICHAEL A. MAYHEW, National Science Foundation
COL. JERRY PERRIZO, Air Force Office of Scientific Research
WILLIAM STRANGE, National Geodetic Survey, NOAA

BOARD ON EARTH SCIENCES AND RESOURCES

COMMISSION ON PHYSICAL SCIENCES, MATHEMATICS, AND RESOURCES

PREFACE

During the course of the efforts required by the Committee on Geodesy to assemble this volume, it became clear that the discipline of geodesy is passing through a unique time of opportunity and challenge, in the scientific and technological areas. The advent of space geodetic technology, beginning more than thirty years ago, together with the recent rapid advances in hardware, data processing methods, and computational and modeling techniques, makes clear that after the year 2000, geodesy will be a fundamentally different science from what it was prior to Sputnik. The rapidity and convenience with which it will be possible to gather highly accurate masses of geodetic data will be unprecedented. As a result, a groundswell of interest from many agencies, institutions, and people is steadily building, which is, paradoxically, the source of a number of current difficulties.

Geodesy is becoming a truly global science, both in the technical, as well as in the political, sense. For example, many of the new observational technologies require the nearly continuous tracking of artificial earth artificial earth satellites. And, since the earth's topography and gravity field are continuous across all political and geographic boundaries, mapping them requires careful integration of data collected by various survey techniques in different countries and physiographic regions. These tasks can only be carried out with cooperation among many countries. Moreover, many of the scientific problems of interest, one of which is the study of earthquakes for the purpose of mitigation of seismic hazard, are international in scope.

It is against the backdrop of an evolving international space technology that this report, Geodesy in the Year 2000, should be viewed. The Committee on Geodesy, recognizing the critical nature of this transitional period, sponsored a session at the fall 1987 American Geophysical Union meeting to highlight the opportunities the immediate future holds for scientific and technical progress. The speakers at that session were asked to contribute papers to this volume. The Overview and Recommendations, prepared by the Committee on Geodesy, are based upon these papers, some of which were updated following the meetings at Erice, Sicily, and Coolfont, West Virginia, described below, and the deliberations of the committee.

Following the 1987 American Geophysical Union meeting, a workshop entitled “The Interdisciplinary Role of Space Geodesy” was organized by the International School of Geodesy of the E. Majorana Centre for Scientific Culture and held in Erice, Sicily, during the period July 23–29, 1988, at which more than 100 representatives from a variety of nations convened to discuss the scientific and technological challenges that confront us. The result of that workshop is a lengthy document (Mueller and Zerbini, 1989), outlining where the international community currently stands, both scientifically and technologically, as well as likely directions for future research.

As a complement to this effort, the National Aeronautics and Space Administration (NASA) organized a similar workshop at Coolfont, West Virginia, during July 1989, to plan the NASA space-geodesy program for the period spanning 1991–2002. This period of time, which immediately follows the termination of the highly successful NASA Crustal Dynamics Project, encompasses a critical transitional period for the disciplines of geodynamics and geology. During this interval, it is expected that the Earth Observing System, and the associated Mission to Planet Earth, will become operational, presaging an era in which the earth will be studied from an integrated systems viewpoint. The result of the Coolfont meeting will be a program plan, budget, and strategy for implementing an integrated NASA program.

JOHN B. RUNDLE, CHAIRMAN

COMMITTEE ON GEODESY

CONTENTS

OVERVIEW AND RECOMMENDATIONS

We stand on the threshold of a technological revolution in geodesy. With the introduction of space-based observational techniques over the past two decades, geodesy has undergone and continues to undergo profound changes. Between now and the end of this century, new technological advances are sure to transform the field even further, by making new generations of measurements possible, with highly increased speed and accuracy. Yet the scientific requirements, in terms of precision, spatial and temporal density of the measurements, and overall coverage of the planet, represent a formidable challenge to even the most advanced geodetic tools.

The central, most important question is this: How are we, as a scientific community, to ensure that very long term geodetic observations will be carried out over the same geodetic networks using the same or comparable standards over durations of many decades? The federal system that has evolved in recent times is currently geared to cycles of funding for research and development extending over a few years at most. Moreover, because of the complexity of space technology, gathering geodetic data is no longer a simple matter of a field party making measurements and an individual serving as “computer” for the network adjustments. Instead, there is a huge variety of tasks to be done, including maintaining continuity over decades of observational techniques, monumentation, network maintenance, education, and survey practice, together with ancillary services such as technology development, data management, orbit determination, and field coordination. With so many agencies and institutions interested in various aspects of space geodesy, and with so much to be done, it is critical that careful thought be devoted to the problem of ensuring that all of these tasks are carried out efficiently and expeditiously. It is hoped that this report, Geodesy in the Year 2000, will stimulate the kind of open, wide-ranging debate within the national and international communities needed to forge a smoothly running operational system. This goal, as well as the scientific and technological issues involved, is the major challenge that currently confronts us.

In the fall of 1987 the Committee on Geodesy sponsored a session at the meeting of the American Geophysical Union to highlight the opportunities the immediate future may hold for scientific and technical progress. Following the Overview and Recommendations are eight of the nine Contributed Papers prepared for that session. The ninth paper was published in 1988 (Paik et al.) The first contributed paper, an historical introduction by Rundle, is a brief account of the land-based methods previously employed in geodesy, and of how these methods often dictated the scientific problems of interest to earlier geodesists. This paper illustrates how much of the precise geodetic work pursued by the U.S. government was at one time vested in one federal agency, the Coast and Geodetic Survey. This is in contrast to the present situation where many organizations are participating in geodetic activities. The second paper, by Minister et al., underscores the critical scientific problems that might most profitably be addressed by new space-geodetic precise positioning technology. In particular, these authors emphasize the need to focus on space-time departures from average plate motions predicted by the global, rigid-plate motion models. Much of the material in the second paper has since been incorporated in the Erice report (Mueller and Zerbini, 1989). The third paper, by McNutt, describes the accuracy requirements for a knowledge of the geoid in the year 2000, and what steps must be taken to achieve this and other scientific goals. Some new technology developments to attain this goal were recently described by Paik et al. (1988). Zlotnicki, in the fourth paper, discusses the need for geodetic control in oceanographic research, focusing particularly on the need for accurate measurement of the oceanic geoid and sea level surface.

The remaining papers describe developments in space-based technology currently in progress. The fifth paper, by Smith, describes a new generation of orbiting lasers for use in precise positioning by ranging to ground retroreflector networks and for use in altimetry. Lasers of this type are already in the planning stages for the Geodynamics Laser Ranging System of the Earth Observing System, and for altimetric purposes on the Mars Observer Mission. Spiess, in the sixth paper, discusses ocean-bottom geodesy, a field that will essentially be created over the next decade. Rodgers details in the seventh paper how a variety of technical improvements over the next decade are expected to improve accuracies in Very Long Baseline Interferometry observations. Finally, Melbourne, in the eighth paper, summarizes the technical improvements that will be possible in Global Positioning System receivers in the near future, leading to inexpensive and user-friendly receivers.

RECOMMENDATIONS

Both the Erice report (Mueller and Zerbini, 1989) and the papers included in the following pages demonstrate the promise of the new technology. In addition, the Erice report makes a variety of specific recommendations as to which of the various scientific problems, and which technology development alternatives, should receive the greatest priority. After consideration of these priorities, and in light of the results presented in the chapters of this volume, the Committee on Geodesy recommends the following priorities be established in support of scientific and technological opportunities in geodesy for the year 2000:

1. <u>The U.S. government should sponsor, as a critical national priority, a vigorous, coordinated program for the development and exploration of modern geodetic techniques, through aggressive pursuit of continued technological advances, and through a long-term commitment to the routine acquisition, reduction, archiving, and distribution of global and regional high-precision geodetic data.</u>

   Observation and monitoring of high-precision networks by space-and ground-based geodetic technology address vital national needs, which cannot be satisfied on the necessary scale by private or educational sectors.

   Moreover, by virtue of the long periods involved, as well as the wide geographic coverage and centralized coordination necessary, federal, rather than state or local leadership, is mandatory. There are critical needs for continuity in observational techniques, monumentation, network maintenance, education, and survey practice, together with ancillary services such as technology development, orbit determination, and field coordination, over periods of decades and longer. Historically, this role was filled in the United States, in large part, by the National Geodetic Survey and its predecessor agencies, which were able to provide continuity in operational procedures and standards over more than 100 years. However, many of the tasks involved in operational geodesy are now spread over a diversity of federal agencies, giving rise to problems of coordination stemming from overlapping responsibilities. Unfortunately, some agencies are no longer able to execute effectively certain programmatic objectives, due to fiscal problems arising from declining budgets. Interagency accords and financial support will therefore be essential for development and routine use of high-precision space geodetic technology. In addition, coordinated international collaboration will be necessary, given the global nature of space geodesy.

2. The federal government should organize and sponsor programs to conduct geodetic instrument research and development; maintain international geodetic observatories; strengthen global positioning networks with stations on stable plate interiors; and deploy long-term, dense, frequently observed local positioning networks at sites of active crustal deformation.

New and refined technologies that will enable positions to be measured frequently or continuously at the level of a few millimeters are a high priority. They will be an important component of the arsenal of techniques to be used in future studies of tectonic and volcanic phenomena as well as global climate change. To address these problems, it is crucial that federal support of geodetic observatories continue for the purposes of developing and calibrating new and existing instruments in a field environment, performing reliability tests, and comparing competing borehole, surface, and space technology systems. This program of testing and development should also include increased research to understand the origin of survey monument motion due to instabilities in the uppermost meters of the surficial layers of the earth, because such 1- to 5-mm motions will soon become a major contributor to measurement uncertainty. Seafloor observatories are also needed for testing and comparing sea floor positioning systems. Global geodetic networks, with fiducial stations sited on extremely stable monuments located in stable plate interiors, should be regularly observed as calibration points for transient deformation at plate boundaries. For example, the “Fiducial Laboratories for an International Natural Science Network” (FLINN), conceived at the Coolfont workshop, would contribute significantly to implementing this committee's recommendation. As proposed, FLINN would build on existing global space geodetic networks to achieve worldwide coverage with intersite spacing on land of approximately 1000 km, and positions to 1-cm precision over short periods (one day) and 1-mm precision over longer time intervals. Dense (0.2–2 monuments/km), highest-precision (1–3 mm), local (10^2–10^4 km^2) networks confined to sites of active seismic, tectonic, or magmatic deformation with the prospect of continuous or frequently repeated observations (one per day) will be essential for short-term analyses of earthquake or volcanic activity.

3. A global topographic data set should be acquired with a vertical accuracy of about 1 m, at a horizontal resolution of about 100 m.

Existing topographic data are largely derived from stereo-photogrammetric techniques, whose data vary greatly in resolution, accuracy, format, and reference level. Acquisition of a coherent, global data set should constitute a major geodetic goal for the year 2000. Over the continents, these data are best obtained from space, by advanced radar and laser altimeters. Radar interferometers may also be feasible, and research into these and other technologies should be vigorously pursued. Improved topographic data are essential for interpretation of the continental gravity field, and would support a variety of tectonic, hydrological, and ecological studies. Monitoring alpine glaciers and the Greenland and Antarctic ice sheets to determine their responses to the possibility of carbon-dioxide-induced global warming is critical. Alpine glaciers serve as sensitive indicators of mean global temperature, while changes in the polar ice caps must be monitored and understood because of their potential for catastrophic impact on mean sea level.

4. The geodetic community strongly encourages the development of space-based techniques to determine variations in the earth's gravity field to a resolution of 100 km or better, with an accuracy of several milligals, together with airborne techniques using precise Global Positioning System (GPS) tracking to obtain regional gravity fields with a resolution of 10 km at an accuracy of better than 1-mgal.

Improved gravity field information is necessary for studies of compensation mechanisms, lithospheric structure, and geodynamical studies of the earth's deep interior. Error analysis studies suggest that resolution of the global gravity field to 100 km can be obtained by the use of satellite-borne gravity gradiometers, one of which is currently planned for the European ARISTOTELES mission, to fly at 200 km, with an accuracy goal of 5 mgal. Enhanced resolution can be obtained by flying a superconducting gravity gradiometer in a drag-free configuration at an altitude of 160 km, for a planned accuracy of 1-mgal. A gradiometer based upon proven airborne gradiometer technology, mounted in a conventional satellite, should be considered as a low-risk alternative. The goal of 100-km resolution may still be inadequate for many tectonic studies. Additional studies suggest that 1-mgal precision, from 300-km to 5-km wavelengths, can be achieved with an airborne gradiometer system under differential GPS control.

Centimeter-level radar altimetry over the oceans, with improved space-time resolution, is of considerable importance. These data will lead to substantial improvement in our knowledge of the oceanic gravity field, have numerous oceanographic applications, and be of value in studies of bathymetry and marine tectonics.

5. To enable the study of complementary geophysical, atmospheric, and oceanographic phenomena, an observational program should be implemented to recover vector earth rotation data, such as polar motion and length of day, at the sub-milliarcsecond level, on time scales from cycles per day to cycles per decades.

The time variation of the earth's instantaneous rotation vector is governed by the planetary angular momentum budget, which in turn depends explicitly on a variety of factors including externally applied torques. Having both long-and short-term measurements permits study of problems including the fluid-outer and solid-inner core interaction, resonances, time variation of the earth's gravity field, and transfer of angular momentum between the atmosphere, oceans, and the solid earth. These data would also support establishment and maintenance of a conventional terrestrial coordinate system, which is necessary for comparison of results obtained by different space geodetic technologies, including very long baseline interferometry, satellite and lunar laser ranging, and the Global Positioning System.

NOTE

The Committee on Geodesy expresses its appreciation to the authors of the eight contributed papers included in this report. However, responsibility for the Overview, including the recommendations, rests with the Committee on Geodesy.

REFERENCES

Mueller, I.I., and S. Zerbini, Proceedings of the International Workshop. The Interdisciplinary Role of Space Geodesy, Erice, Sicily, July 23–29, 1988, Lecture Notes in the Earth Sciences, Springer-Verlag, Berlin, 1989.

National Aeronautics and Space Administration, Solid Earth Science Branch, Earth Science and Applications Division, Office of Science and Applications, Major Emphasis Areas for Solid Earth Science in the 1990s; Report of the NASA Coolfont Workshop, Nov. 27, 1989.

Paik, H.J., J.-S. Leung, S.H. Morgan, and J. Parker, Global Gravity Survey by an Orbiting Gravity Gradiometer, EOS Trans. AGU, 69, 1601–1611, 1988.

CONTRIBUTED PAPERS

Geodesy in the Year 2000: An Historical Perspective

John B. Rundle
Division 6231
Sandia National Laboratories
Albuquerque, NM 87185

INTRODUCTION: GEODESY IN THE YEAR 1900

At 0512 hours Pacific Standard Time on the morning of April 18th, 1906, the city of San Francisco was destroyed by a major earthquake. Subsequent study determined the approximate magnitude to have been in excess of 8, the event having ruptured more than 400 kilometers of the nearby San Andreas fault. In the words of the Carnegie Commission, which was empaneled to investigate the earthquake and its causes (Lawson et al., 1908):

> "The shock was violent in the region about the Bay of San Francisco, and with few exceptions inspired all who felt it with alarm and consternation. In the cities many people were injured or killed, and in some cases persons became mentally deranged, as a result of the disasters which immediately ensued from the commotion of the earth. The manifestations of the earthquake were numerous and varied. It resulted in the general awakening of all people asleep, and many were thrown from their beds. In the zone of maximum disturbance persons who were awake and attending to their affairs were in many cases thrown to the ground. Many persons heard rumbling sounds immediately before feeling the shock. Some who were in the fields report having seen the violent swaying of trees so that their top branches seemed to touch the ground, and others saw the passage of undulations of the soil. Several cases are reported in which persons suffered from nausea as a result of the swaying of the ground. Many cattle were thrown to the ground, and in some instances horses with riders in the saddle were similarly thrown. Animals in general seem to have been affected with terror."

It was well known at the time that earthquakes are caused by the relief of elastic strain in the earth's crust. The Pittsburgh Post of April 19, 1906, page 6, recounts:

> … "it is probable that earthquakes are caused by the same stresses in the earth's crust, partly due to contractions, that tilt and fold rock strata into mountains. The strain to which the rocks are thus subjected, when suddenly relieved by the rocks giving way, produces many earthquakes."

Furthermore, according to the Post:

> "Earthquakes … have usually, when carefully studied, been traced to some line of rock weakness as a fault. Such earthquakes are merely phenomena accompanying rock movements that may in time greatly modify the earth's surface."

While the San Francisco earthquake was not the first to be scientifically investigated, its report had the greatest impact, both upon scientific thought in general, and upon geodesy in particular. Other reports by Robert Mallet on the 1857 Naples earthquake, by R. D. Oldham on the 1897 Assam earthquake, and by John Milne on the 1880 Yokohama earthquake were instrumental in establishing seismology as a scientific discipline. But the 1908–1910 Carnegie Commission reports on the San Francisco earthquake stand alone, because it was there that the elastic rebound theory of earthquakes was first introduced by Harry Fielding Reid (Reid, 1910). The primary data supporting the hypothesis that earthquakes represent a rebound from a state of previously stored elastic strain energy were geodetic survey measurements conducted between 1851 and 1906 (Figure 1, taken from Hayford and Baldwin, 1908). To Reid, these data indicated that the earth's crust had, over the preceding decades, undergone a systematic deformation whose effect was to place the San Andreas fault into a state of disequilibrium. The earthquake was thus a result of forces in the earth's crust returning the system to a state nearer to equilibrium. As Reid showed, the survey data indicated that the sense of sudden motion of monuments near the fault at the time of the earthquake was the reverse of steady motion that occurred prior to the event, indicating the release of stored elastic energy.

The geodetic technology in common use by the Coast Survey in the latter part of the nineteenth, and early part of the twentieth centuries, relied primarily upon triangulation for measurements of horizontal position (Hosmer, 1919). In this method, permanent marks were fixed to the ground in networks of regular triangular patterns, over which were conducted measurements to determine the angles subtended by lines-of-sight between the marks. For the most part, the earlier triangulations were conducted during daylight hours, by sighting on a sun-reflecting heliotrope with a telescope precisely calibrated in angular direction. The most commonly used telescope was of the type known as a "Direction Instrument", first designed in England by Ramsden in 1787. Other telescopes of lesser precision were the "Repeating Instruments" designed in France in the year 1790. Beginning in 1902, triangulations were primarily conducted at night, it being realized that thermal instabilities in the atmosphere produce unacceptable lateral refraction which can be remedied by observing through thermally stable nighttime air. Acetylene lamps were used initially, later supplanted by incandescent electric lights. Generally speaking, observations were obtained by mounting the instruments and the lights at the tops of wooden, and later metal towers, whose heights ranged up to more than one hundred feet.

It was using these instruments and technologies that the data used in the Elastic Rebound hypothesis were obtained. At the time, the combination of standard observing and network adjustment methods were expected to yield first order accuracies in line lengths of about 1 part in 25,000 (Hosmer, 1919). To obtain these accuracies, instructions such as those issued to Clem Garner, chief of a survey party operating near San Francisco in 1922, were typical (Bowie, 1924):

> "You will, therefore, take special precautions against conditions which would cause horizontal refraction and will adopt such an observing program as will secure triangle closing errors with 2.5" as a maximum, and with not more than 1" as a mean. It is recommended that each direction at a station be measured on at least two nights with not less than 12 acceptable directions on each night. One direction at each station may have only 16 acceptable positions observed and these may be on a single night if by so doing a day may be saved and provided further that the closures are within the above limits."

It was realized by the geodetic community at least as early as 1924 (e.g., Bowie, 1924), that the decade-long changes in triangulation networks observed throughout California were the result of earth movements, and that these earth movements were related to the San Andreas fault. In these early papers, one can see a clear emphasis on analyzing changes in triangulation angles between successive surveys as a means of understanding the physical role of fault movements. In modern terminology, we refer to this as relative positioning. But the most important rationale for the establishment of regular survey campaigns, from a political and economic point of view, has always been for cadastral, or boundary determination purposes. The 1850–1900 coast surveys were primarily made for coastal navigation of ships, not for boundary determinations, as shown by the fact that the published maps of the lines of sight also listed sailing instructions in the lower left corner. To improve positional accuracy, astronomic observations of monuments (Laplace Stations) in the network were obtained to control orientation, and to obtain geoid slopes for scientific studies. With proper care, positional accuracy of one part in 100,000 was achieved.

In the early twentieth century astronomical longitudes of points in the field were determined by using a portable transit to measure the time of passage of stars past the local meridian. The principle sources of error were the accuracy with which the exact transit time could be determined by the observer, and undetermined errors in polar motion and universal time (UTl). Moreover, since local transit time is supposed to be taken with reference to the geometrical reference ellipsoid, deflection of the vertical induced by anomalous masses implies a pointing error of the transit telescope, and thus errors in the inferred longitude. Upon comparison of the observed transit time to the precalculated Ephemerides, the longitude could be obtained. Alternately, a comparison could be made to a transmitted time calibration signal. Prior to 1922, these signals were received in the field over telegraph lines, but subsequently were transmitted via radio. Astronomical latitudes were determined by observing the elevation of stars above the horizon using a zenith telescope. Again, a major error source arose from deflection of local vertical. Using these techniques, it was found possible to astronomically measure to about 0.10" in latitude, and 0.003" in longitude, albeit at considerable effort and expense.

With the passage of time, instrumentation for land-based horizontal positioning evolved. Triangulation was still the primary method for obtaining horizontal positions until about 1960, when electro-optical distance measuring instruments were developed, such as the Geodimeter and the Geodolite. When temperature and humidity are measured at the time of ranging, and appropriate corrections for atmospheric refraction are applied, line-lengths can be obtained over networks of monuments with typical spacings of kilometers to tens of kilometers, with accuracies in the range of parts per million or better. Other land-based instruments have followed, including multiwavelength electro-optical distance measuring apparatus, which measure and apply atmospheric refraction corrections automatically.

Although not as important as horizontal triangulation for analyzing motions related to strike-slip faulting, leveling measurements had, by the late nineteenth century, reached a high state of technical accomplishment (Hosmer, 1919). In fact, the techniques and technology used then are, in all essential aspects, basically the same as those in use today. As in triangulation, leveling measurements are made over networks of marks (benchmarks) fixed on the earth's surface, generally along roads, railroad beds, and other gently sloping paths with good access. The major advance in leveling technology occurred with the discovery of an alloy of 35% nickel and 65% steel called “invar”. Discovered originally by C. E. Guillaume, Director of the International Bureau of Weights and Measures near Paris, France, invar is distinct in having an extremely low coefficient of thermal expansion ($\lesssim$.1 ppm/°C), due to a special heat treatment used in its preparation. By 1906, the Coast Survey had begun using invar for measurement tapes and leveling rods, and has continued this practice to the present. The performance of invar has been found to be generally satisfactory, except possibly in leveling measurements of extreme accuracy, when instabilities in material structure may cause unpredictable changes in length of the tapes at the level of parts per million. More important sources of error in precise leveling measurements arise from unequal atmospheric refraction effects over forward and backward sightings, and systematic errors in rod calibration. With more modern self-leveling telescopes, an additional source of error has been found to arise from deflections of the compensator pendulum induced by nearby electromagnetic sources such as power lines. Still, rigorous field tests demonstrate that accuracies achieved in leveling are about 10 mm over 100 kilometer distances, and about 5 mm over 1000 meters of elevation change.

In contrast to positioning, which is the primary geodetic observable of interest in the study of active faulting, gravity is of principal importance in studying the geophysical structure of the earth. Prior to about 1800, it was thought that the matter comprising the earth and its surface features was of roughly uniform density, and that any deformation of the earth was of an essentially elastic nature (Jeffreys, 1976). However, following the 1855 survey of India, J. H. Pratt found that the gravitational attraction of the Himalayas observed by Everest, the Surveyor General of India, was only about one-third as large as it should be, if the mountains were treated as uncompensated masses. Shortly thereafter, G. B. Airy, then the Astronomer Royal, proposed in 1855 that mountains floated on a substratum capable of deforming inelastically in response to the excess gravitational load. Pratt, in 1859, proposed an alternate hypothesis, in which mountains of lower density than the substratum ride passively on the underlying rigid material. Both of these mechanisms are still invoked today, in studying the isostatic compensation of surface features of the earth. The applicability of both mechanisms continues to be a subject of research, in addition to other problems related to the structure and dynamics of the earth.

Early gravity meters were based upon measuring the period of an accurately calibrated pendulum. The first of these, the half-seconds invariable pendulum apparatus, was designed and perfected in 1882 by Sterneck in Austria (Hosmer, 1919). In 1890, T. C. Mendenhall, Superintendent of the Coast and Geodetic Survey modified the design, producing an instrument which was used successfully for many years. The basic design involves the comparison of the locally measured period of the pendulum to a chronometer of known calibration. Due to the nature of the measurement, the amount of time needed for a single observation was typically on the order of 8 to 12 hours. The accuracy of the observations was typically a few ppm, that is to say, a few milligal.

Another instrument also in use at the time, and one whose importance has undergone a resurgence due to studies of the fundamental nature of gravity, was the Eotvos torsion balance. In this instrument, two masses are fixed to the ends of a long, slender rod, which is suspended at the end of a long fiber. While the nature of these masses was not considered important for routine applications, it has since assumed considerable importance, for reasons to be discussed in this volume. Under the action of a spatially varying gravity field, the rod tends to turn into the plane of a great circle oriented perpendicular to the local meridian, called the prime vertical. By measuring the torsion in the fiber, the intensity of the local gravitational field can be deduced. Other designs were also in use at the time, but all were eventually supplanted by the far more durable and portable gravity meters. Accuracies of the torsion balance were similar to those of modern portable gravity meters, a few microgals.

Unlike pendulum meters and torsion balances, gravity meters, which first began to appear in the 1930's, are in essence a mass suspended on the end of a sensitive spring (e.g., Telford et al., 1976). The early gravity meters were stable, that is to say, they had a linear dependence of deflection on deviation of the length of the spring from a relatively large, nonzero value. The major problem with an arrangement of this type is that the signal of interest is the deviation about the mean deflection, thus producing a relatively poor sensitivity. The Gulf and Boliden gravity meters are of this type. Generally speaking, the stable meters have fallen into disuse.

By contrast, the unstable gravity meters have inherently a far superior sensitivity, as small deviations are magnified substantially. The most important of these instruments is the LaCoste-Romberg gravimeter, which utilizes a zero length spring. In 1934, L. J. B. LaCoste found a method of producing a spring-balance system in which the restoring force depends inversely on the length of the spring. The result is a spring-balance system of great sensitivity. In fact, the sensitivity becomes unbounded as the actual length of the spring approaches zero. In practice, such meters are used in a null mode, where rotations of an adjustment screw are used to restore the mass to its original position. The deviation of local gravity from a reference value is then proportional to the number of turns of the screw. Major sources of error are atmospheric pressure changes and temperature variations. For these reasons, the mass and spring assembly are enclosed in a pressure and temperature controlled environment.

Most recently, transportable absolute gravity meters have been developed which are based upon timing the free fall of a mass. In this case, the mass is a corner cube reflector, and its velocity is measured with a laser interferometer system. Accuracies are typically in the range of a few microgals, if several hundred drops of the cube are measured.

CURRENT PROBLEMS: APPROACHING THE YEAR 2000

With the advent of the space age, following the launch of Sputnik in 1957, the science of geodesy entered a new era. As the chapters in this volume describe, a variety of new space-based observational techniques are under development which will allow significantly new approaches to old problems, and provide means for addressing scientific questions which were previously insoluble. As is implied by the foregoing historical discussion, current problems in geodesy revolve principally around positioning and gravity field determination. To this list can be added a new and rapidly evolving topic related to fundamental tests of natural laws, principally General Relativity, and the inverse-square law of classical Newtonian gravitation (see the chapter by Paik).

Precise positioning measurements play a critical role in manygeodynamical problems, some of which are summarized in Walter (1984).

Following the approach of Jordan and Minster (1988), it is possible to classify the most important of these problems into two categories, secular and transient, and to classify the data types needed to address these problems according to whether motions are predominantly vertical or horizontal (see Chapter 2). Of the four distinct combinations, each typically demands its own approach, and to obtain a solution, each typically has its own accuracy standards.

The division of crustal motions into transient and secular reflects the increasing recognition that even now, space-based geodetic measurements are succeeding in defining the time averaged, long-term motion of the plates to subcentimeter accuracies. That is, relative plate motion rates, when measured from the stable interior of plates, and averaged over tens of years, are in most cases essentially identical to rates averaged over millions of years. Thus, the predictions of the global plate motion models (e.g., Minster and Jordan, 1978) are being confirmed with ever-increasing confidence. Where deviations from these rates occur, it seems clear that additional physical processes are at work on a more local scale, which are unrepresented by the global models. As a result, it is becoming possible to address problems related to fluctuations about the long term rates, that is to say, processes which are responsible for producing transient crustal movements. As the long-term motions of arbitrary points becomes known to accuracies of several millimeters per year or better, it will become possible to subtract these motions from the instantaneous rates with confidence that what remains will have physical significance. The observations needed to carry out this program will have the characteristic properties of being frequent, spatially dense observations over geodetic networks of large scale, with positional accuracies of several millimeters.

In terms of dynamical problems related to plate motions, it is thus possible to think in terms of a variety of scales, both spatial and temporal. The average rates given by the global rigid plate motion models are defined over the preceding 2–3 million years, and the typical spatial scales represented are on the order of plate dimensions, that is to say, thousands of kilometers. Physical understanding of the processes responsible for these observations will come from increasingly sophisticated global convection models, which even now are yielding considerable insight into the long term evolution of the earth. As motions on these space-time scales become better defined, many of the frontier-level problems will increasingly be focused on dynamical processes spanning decade and less time scales, over spatial dimensions of tens to hundreds of kilometers. Motions on these time scales have considerable importance in a variety of phenomena, including earthquakes, polar motion and length of day, and post-glacial rebound studies. To the extent that transient dynamical processes of the solid earth are reflected in observed polar motion, the understanding of such diverse phenomena as climatic changes and ocean circulation will be greatly enhanced. Thus, in effect, knowledge of rigid plate motions provides a space-time reference frame, as well as kinematic boundary conditions, for understanding much shorter-lived motions on time scales of hours to hundreds of years.

Transient tectonic motions on spatial scales of hundreds to a thousand or so kilometers are often termed "regional" deformations. It is typically upon these spatial scales that one sees significant departures from the predictions of the rigid plate motion models. These departures are in the form of space-time variations in average motion, due, for example, to earthquakes, transient motions of the asthenosphere, and aseismic slip, as well as in the existence of complex boundary zones of deformation. Examples of the former can be seen at well-studied plate boundaries such as southern California, Japan, and Alaska. Examples of the latter can be seen in the western United States, the Alpide belt, Tibet, and the East African Rift Zone. Of the transient motions, perhaps the most interesting is the apparent migration of stress and strain along plate boundary zones. One of the best examples is the sequence of earthquakes which occurred along the North Anatolian fault zone in Turkey, between 1939–1967. These remarkable events, all of about magnitude 7, appeared to originate with the earthquake of December 26, 1939, which occurred in the northeastern part of the country. Subsequently, in 1942, 1943, 1944, 1953, 1957, and 1967, a sequence of events occurred progressively farther to the west, at an "average" rate of migration of something like 100 meters/day. In addition to these, an event occurred in 1966 to the east of the epicenter of the 1939 event. With the occurrence of these earthquakes, most of the north Anatolian fault, together with the northern segment of the east Anatolian fault, had ruptured. The most fascinating question is whether such a migration effect is a ubiquitous feature of fault zones. In fact, there is an accumulating, but still small, body of evidence which suggests causal relations in space and time between major earthquakes on a variety of plate boundaries. These effects have been observed in the Nankai region of southwest Japan, the Imperial Valley of California, the central Nevada seismic zone, and in recent paleoseismic studies of the San Andreas fault zone of southern California.

In addition to regional-scale events, an important class of local-scale episodic displacements are associated with volcanism. In a number of large caldera structures around the world, such as Long Valley in California, Rabaul in Papua New Guinea, and the Campi Flegrei west of Naples, Italy, displacements of perhaps a meter have occurred over intervals of a few years, without as yet an accompanying eruption. Typically, these caldera structures have areas of perhaps 500–1000 km. In another case, that of Mount Saint Helens, vertical motions of hundreds of meters occurred over about two months, culminating in the devastating eruption of May 18, 1980. These crustal motions occur at such a rapid pace that conventional land-based surveying cannot provide a detailed space-time picture of the course of the inflation, with the consequence that valuable information on the source process is lost. Space geodetic systems, by contrast, have the flexibility to observe large networks of points relatively cheaply, or to operate unmanned in a continuous mode of operation.

The most important and interesting problems in understanding dynamical processes associated with earthquakes and volcanoes are at regional scales, over time intervals of hours to hundreds of years. Slip along fault zones, for example, is accommodated by a combination of seismic and aseismic deformation which can be detected by geodetic methods. Given the fact that both slip in fault zones and volcanic phenomena occur over finite, nonzero regions in space and time, the fundamental dynamical question is how the system becomes organized in a physical sense into the space-time structures observed. Thus, the emerging point of view in the earth sciences is to seek a generic understanding of the dynamics of complex systems. This is a fundamentally process-oriented approach, rather than the more traditional observation-driven approach, and mirrors the same fundamental change occurring in other fields of science. Implementation of this approach emphasizes the need for models to aid in understanding, the need for observations to test the models then following as a logical consequence. These observations will involve spatially dense, temporally frequent positioning and surface gravity change observations.

The evolution in the frontier for current research involving the earth's gravity field has been no less dramatic than for the dynamics of the earth's crust. At the turn of the century, discrimination between differing compensation mechanisms, and study of the the tides and the dynamical figures of the earth and moon received the greatest attention. By contrast, gravity observations now play a critical role in almost all aspects of earth science research. These areas include understanding the structure of the earth's lithosphere and its interaction with the underlying asthenosphere, thermal structure of the continental lithosphere and the nature of the driving forces of plate tectonics, and the composition and rheology of the mantle (NASA, Report of a Gravity Workshop, 1987; see also the chapter by McNutt). To address these questions, accuracy requirements are typically in the microgal range over horizontal distances of tens to hundreds of kilometers. In addition to these structural questions, repeat measurements of gravity at the microgal level are a ready means of measuring, on a temporally continuous basis, small vertical motions of the crust. Typically, these vertical motions are most important in post-glacial rebound studies, where one expects to observe changes of a few microgals per year, and in association with volcanoes or thrust faults, where changes of tens to hundreds of microgals may be observed over time intervals of hours to years.

In studies of the oceanic lithosphere, important questions involve the degree to which structural features formed by dynamical processes, such as mid-oceanic ridges, fracture zones, and seamounts, are isostatically compensated, and upon what time scale this process occurs. Of late, data from several satellite altimeter missions, including

SEASAT, GEOSAT, and GEOS-3, have played an important role in mapping the oceanic geoid. With these missions, uniform mapping of the ocean's gravity field and geoid have been achieved for the first time. However, due to military constraints, these data are usually obtained only in modes where the orbits have been made to repeat every few weeks. Hence, spatial coverage in the along-track direction is far better than in the cross-track direction. By contrast, bathymetry data obtained by shipborne instruments such as Seabeam or SeaMARC are able to provide far more detailed topographic coverage on the local scale of features such as fracture zones and mid-oceanic ridges. Nevertheless, there are problems that can only be addressed by the use of satellite geodetic data, such as the origin of the large gravity anomalies observed at oceanic trenches, and the mechanisms by which mid-plate swells and plateaus are supported.

For the continental lithosphere, gravity data allow models of rifting and continental extension to be systematically tested. For example, locations of continental rift zones are often associated with prominent gravity anomalies. Gravity data can also provide information on the depth and lateral extent of sedimentary basins, as well as upon the nature and extent of the roots of mountain belts. Moreover, gravity data offer important constraints on the deep structure of the continental lithosphere, whose thickness and physical properties are not at present well understood.

But perhaps the most critical role which gravity observations play is in determining the long term dynamics of the interior of the earth. It is clear from a variety of observations that the motions in the mantle are driven by thermal convection, implying the existence of laterally heterogeneous density variations. In concert with recent advances in seismic data analysis techniques, the gravity field of the earth, together with the shape of the geoid, provide the most important constraints on the density contrasts in the earth's deep interior. To the extent that these density variations are physically related to the convective processes which drive the plates over hundred-million year time scales, gravity data play a key role in unraveling the mechanism for long term mantle dynamics. And of course, since it is widely recognized that the source of the geomagnetic dynamo is undoubtedly convective motions within the fluid outer core, geoid and gravitational observations place critical constraints on the generation of the earth's magnetic field.

The most important problem areas in understanding the long term dynamics of the earth involve understanding the depth of penetration of subducted slabs, that is to say, the vertical scale of mantle convection; determining the viscosity structure of the mantle; understanding the possible role of small scale (100 kilometer wavelength) convection in the upper mantle; and investigation of the still-unresolved source of the long-term stability of mantle plumes. In general, these problems require global gravity field coverage, with accuracies on the order of milligals, and resolution of roughly 100 kilometers. In addition, there exists considerable interest in obtaining an improved marine geoid, at the level of 0.1 meter accuracy over wavelengths of 100–200 kilometers. The rationale lies in the search for an improved understanding of general circulation and dynamic topography in the oceans, which is in turn motivated by questions related to atmosphere-ocean interaction, and problems related to global climate change.

As a final note, precise determinations of the earth's gravity field have taken on new importance in light of recent suggestions that classical Newtonian gravitation should be revised to admit shorter range interactions (the popularly termed fifth-and sixth-forces). These suggestions stem from analysis of anomalous kaon decays seen in a few accelerator experiments, new geophysical determinations of G from experiments in mine shafts, from gravity observations in boreholes in icecaps and on towers, and from reanalysis of the classical Eotvos experiments. As yet this controversy is unresolved, but the several-hundred meter wavelengths suggested for the interaction range should be visible in some precise satellite tracking experiments, through precise gravity gradiometry, and by more conventional terrestrial means. In addition, there is considerable interest in using satellite gravity studies to test predictions of General Relativity, such as the rate of precession of a space-borne spinning gyroscope by the Lense-Thirring effect, otherwise known as the “dragging of inertial frames”. The rate of precession depends on the intensity of the local gravitational field.

ACKNOWLEDGEMENTS

I am indebted to my colleagues on the committee on geodesy, namely, C. Goad, T. Dixon, E. Metzger, J. B. Minster, R. Sailor, R. Stein, and H. Orlin, for reviews. The work contained in this paper was supported under contract DE-AC04-76DP00789.

REFERENCES

Bowie, W., Earth Movements in California, Spec. Publ. 106, 273, U.S. Govt. Printing Office, Washington, DC, 1924.

Douglas, N. B., Satellite Laser Ranging and Geologic Constraints on Plate-Tectonic Motion, M. S. Thesis, University of Miami, 1988.

Hayford, J. F. and A. L. Baldwin, The Earth Movements in the California Earthquake of 1906, in Lawson, A. C. and others, The California Earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission, published by the Carnegie Institution of Washington, DC; Volume I, 1908.

Hosmer, G. L., Geodesy, Including Astronomical Observations, Gravity Measurements, and Method of Least Squares, John Wiley & Sons, New York, 1919.

Jeffreys, H., The Earth, its Origin, History and Physical Constitution, Cambridge University Press, Cambridge, 1976.

Jordan, T. H. and J. B. Minster, Beyond Plate Tectonics: Looking at Plate Deformation with Space Geodesy, in The Impact of VLBI on Astrophysics and Geophysics, Proc. IAU Symp. 129, ed. M. J. Reid and J. M. Moran, Reidel, Dordrecht, 1988.

Lawson, A. C. and Others, The California Earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission, published by the Carnegie Institution of Washington, Washington, DC; Volume I, 1908

Minster, J. B. and T. H. Jordan, Present Day Plate Motions, J. Geophys. Res., 83, 5331 – 5354, 1978.

NASA, Geophysical and Geodetic Requirements for Global Gravity Field Measurements, 1987–2000, Report of a Gravity Workshop, Colorado Springs, 1987, published by Geodynamics Branch, Div. Earth Science and Applications, NASA, 1987.

Reid, H. F., The California Earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission, published by the Carnegie Institution of Washington, Washington, DC; Volume II, The Mechanics of the Earthquake, 1910.

Telford, W. M., L. P. Geldart, R. E. Sheriff, and D. A. Keys, Applied Geophysics, Cambridge University Press, Cambridge, 1976.

Walter, L. S., Geodynamics, NASA Conference Publication 2325, published by Earth Science and Applications Division, NASA, 1984.

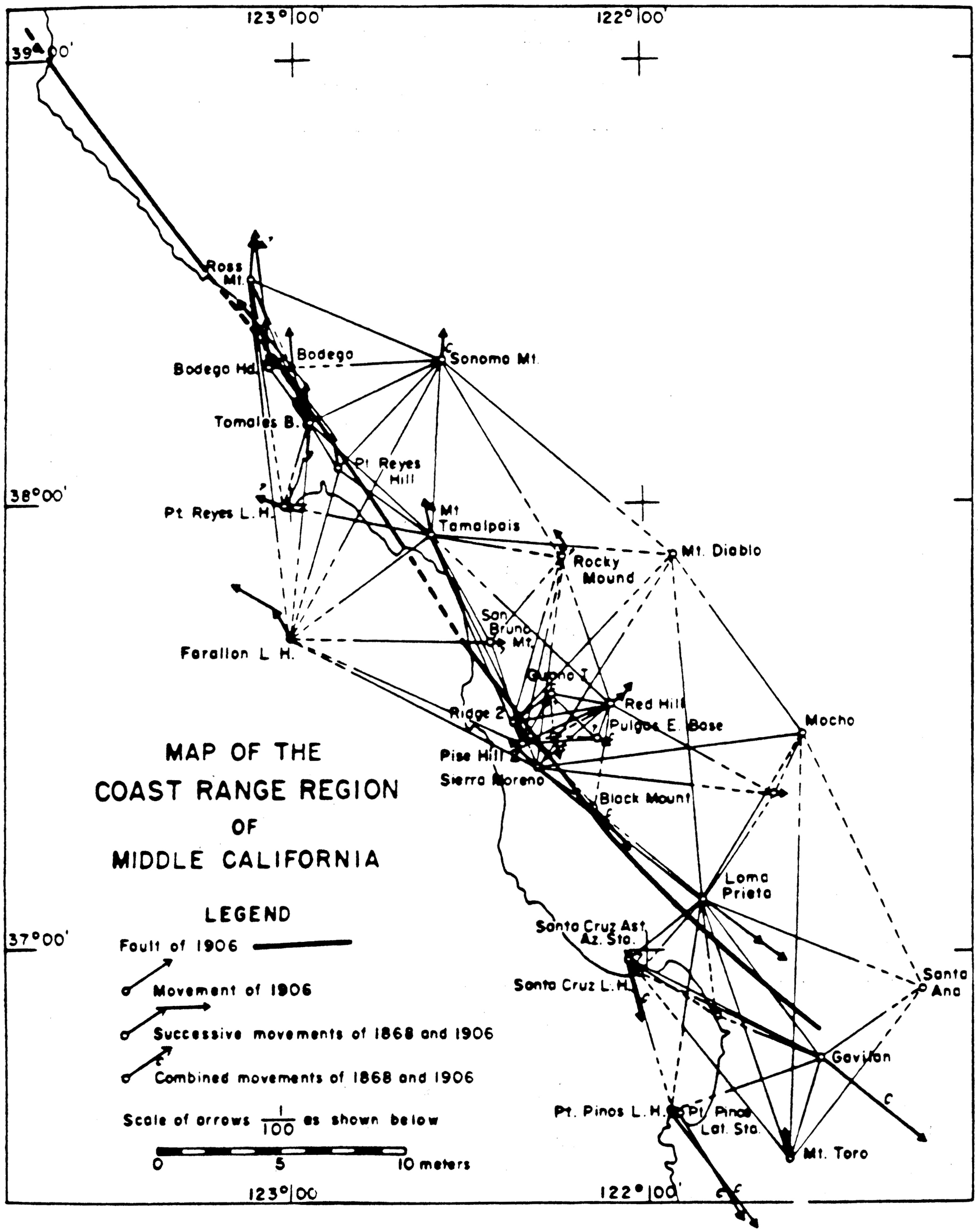


Figure 1. Crustal motions, obtained by first order triangulation near San Francisco, California, during the years 1851–1868, and 1868–1906. Vector motions are due to the earthquakes of 1868 and 1906, as well as to long term interseismic crustal motion.

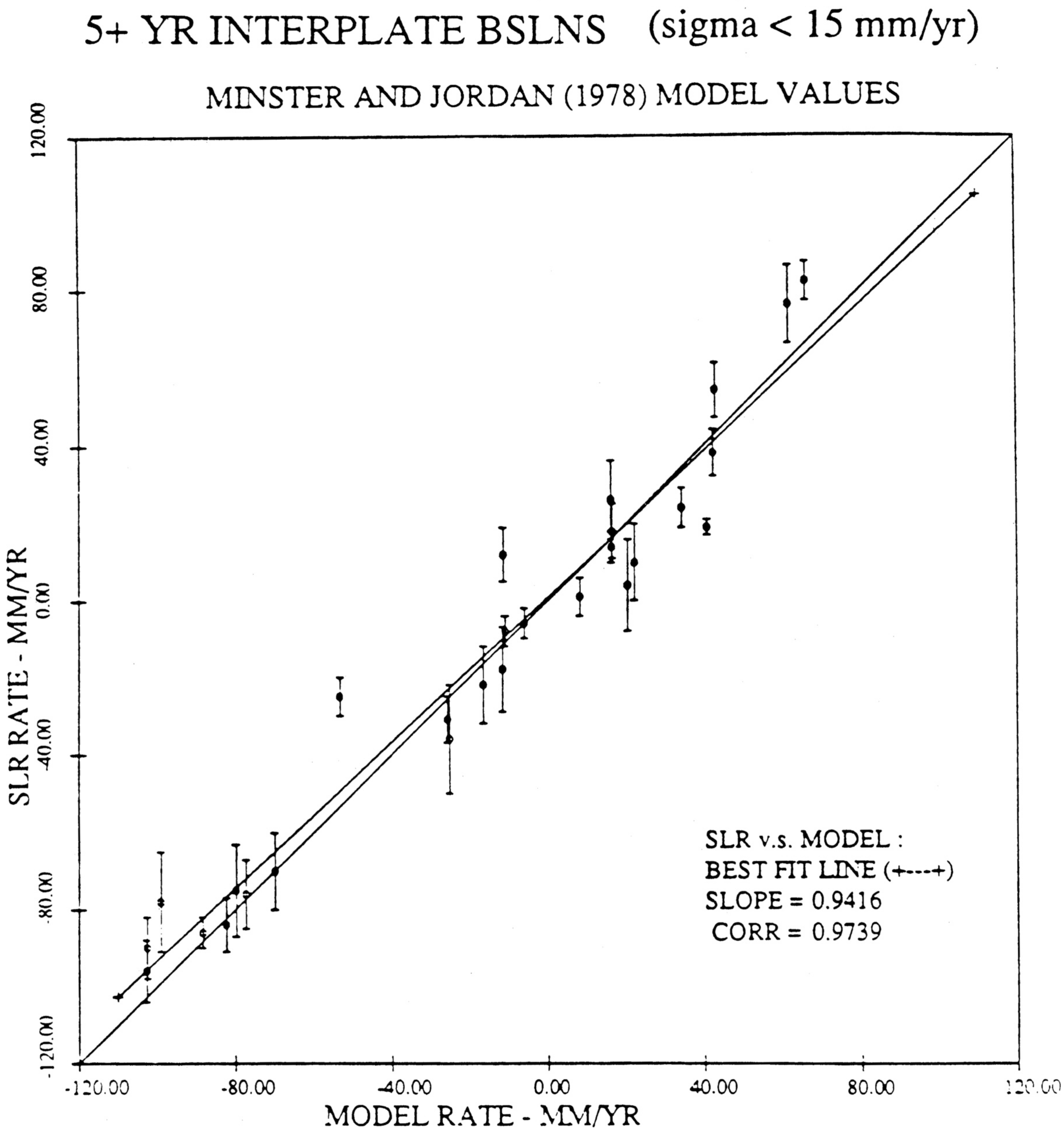


Figure 2. Rates of Baseline length change obtained by Satellite Laser Ranging, compared with rates predicted by the global, rigid plate motion model of Minster and Jordan (1978). Departures of the observed rates from the predicted rates are due to violation of the assumptions inherent in the rigid plate models. Figure taken from Douglas (1988).

Implications of Precise Positioning

Jean-Bernard H. Minster,
Thomas H. Jordan,
Bradford H. Hager,
Duncan C. Agnew,
Leigh H. Royden

INTRODUCTION

One of the most exciting developments in crustal kinematics over the last two decades has been the birth of space-geodetic positioning techniques capable of achieving accuracies of one cm or better. The principal motivation for using these techniques for precise point positioning is to take advantage of the significant increase in technological capabilities that they represent in order to address directly important tectonic problems that cannot be tackled economically at present by ground-based geodetic techniques and classical field geology. The main techniques which have matured over the past decade or so include *Very Long Baseline Interferometry* (VLBI) and *Satellite Laser Ranging* (SLR).

More recently, the less burdensome--from the point of view of field operations--*Global Positioning System* (GPS) has gained considerable popularity for the study of regional deformation problems. Over the technological and scientific horizon, we may count as future candidates the *Geodynamic Laser Ranging System* (GLRS) which is being considered as a facility instrument on the EOS platform, as well as alternatives developed in Europe, such as the French DORIS and the German PRARE systems. Convenience considerations aside, the main advantage of these new techniques, from the geologist's point of view, is that they should permit (in principle) *frequent* resurveys of *dense* networks. This, together with improved control of vertical displacements, represents a completely new, enhanced capability, which will allow geologists to address problems that have so far eluded them.

It may seem difficult to believe that the ability to measure the relative positions of two points on the earth separated by 100 or even 10,000 km has a measurable socio-economic impact. The applications to earthquake prediction and volcanic surveillance are, of course, often invoked, but somehow seem too remote to justify an immediate development effort. Nevertheless, in the past decade, space geodesy has begun to provide useful constraints on the solution of difficult geological problems of immediate importance, such as the distribution of crustal deformation both east and west of the San Andreas Fault in central California (e.g., Jordan and Minster, 1988b). This issue is an interesting one from a scientific point of view, but it also has great practical importance, in view of the high population density and numerous critical facilities found along the Pacific coast of the U.S. In the next two decades, we must extend our experience to other tectonically active areas which have significant human as well as scientific importance.

CRUSTAL MOTIONS AND DEFORMATIONS

Deformation of the earth's lithosphere covers a broad spectrum of temporal and spatial scales, from seconds to aeons and from mineral grains to planetary dimensions. We rely below on the discussion of Jordan and Minster (1988a).

Table 1 categorizes a subset of lithospheric motions that cause geologically and geophysically significant deformations. It is convenient to discriminate secular motions persisting on geological time scales of thousands to millions of years from transients associated with, for example, seismic and volcanic events. Practical research is more concerned with the transients, because they tend to disturb human activities. Secular motions also warrant vigorous study, however, since they provide the kinematical framework for describing transients and understanding their driving mechanisms.

The most significant long-term deformations are those related to plate tectonics. Although local tectonic movements near plate boundaries display large vertical components and time-dependent behavior, the net motions between the stable interiors of large blocks are forced by viscous damping and gravity to be nearly steady and horizontal. The characteristic tangential velocity of the plate system is about 50 mm/yr, which gives rise to displacements easily measured by *geodetic* methods. Horizontal secular motions have been observed both by ground-based networks (e.g., Savage, 1983) and by space-geodetic systems (e.g., Christodoulidis et al., 1985; Herring et al., 1986). Though their application to geodesy is relatively new, space-based techniques have already revolutionized the science of terrestrial distance measurement. They are contributing new information about active tectonics, particularly on the planetary scales previously inaccessible to ground-based surveys (Figure 1).

Rigid-plate motions. In the ocean basins, most of the deformation related to horizontal secular motion occurs in well-defined, narrow zones that are the boundaries of a dozen or so large lithospheric plates. The current plate velocities are constrained by three basic types of data collected along these submerged boundaries: (1) spreading rates on mid-ocean ridges from magnetic anomalies, and directions of relative motion from (2) transform-fault azimuths and (3) earthquake slip vectors. The first self-consistent global models were synthesized soon after the formulation of plate tectonics (LePichon, 1968), and significant refinements were made throughout the next decade (Chase, 1972; Minster et al., 1974). Third generation models were published by 1978 (Chase, 1978; Minister and Jordan, 1978) and are still in use. Work has been recently completed at Northwestern University on an improved fourth-generation plate-motion model name NUVEL-1 (DeMetz et al., 1989), which remedies most of the problems identified with earlier models (see also Gordon et al., 1988).

Since the reference frame is arbitrary, the angular velocity vectors describing the instantaneous relative motions among *M* rigid-plates are specified by 3(*M*-1), independent components, which are derived by least-squares inversion of a carefully selected, globally distributed data set. As shown in Table 2, the increasing sizes of the data sets upon which successive generations of models have been based reflect continued vigorous research activity in geology and geophysics since the advent of plate tectonics.

Table 3 lists the instantaneous rotation vectors ("Euler vectors") that describe the relative motions of the plates in NUVEL-1. It is reproduced from the presentation by R. Gordon and his co-workers at the October 17–21, 1988 NASA Crustal Dynamics Principal Investigators Meeting, held in Munich, Germany (Gordon et al., 1988). The convention adopted in this table is that the first plate moves counterclockwise relative to the second plate. The nomenclature is as follows: af-Africa, an-Antarctica, ar-Arabia, au-Australia, ca-Caribbean, co-Cocos, eu-Eurasia, in-India, na-North American, nz-Nazca, pa-Pacific, sa-South America. Table 3 also lists the one-sigma error ellipses (marginal distributions) attached to the Euler vector estimates. The two-dimensional marginal distribution of the pole position is specified in each case by the angular lengths of the principal axes and the azimuth ξ_{max} of the major axis, and the one-dimensional marginal distribution of the angular rotation rate is specified by its standard deviation σ_w.

The magnetic anomalies employed in the various rigid-plate motion models mentioned above average the rates over the last 2–3 million years, about the shortest time span for which good spreading rates can be obtained on a global basis. Although this interval is hardly "instantaneous" from a geodetic point of view, it is geologically brief, and the small plate displacements that take place during it are well described by infinitesimal (as opposed to finite) rotations. It will probably be some time before the global plate-tectonic models can be significantly improved by space geodesy. Because the geological data sets are large and the inverse problem is strongly overdetermined, the formal uncertainties in the angular velocity components are already quite small, and correspond to formal uncertainties of 1 or 2 mm/yr in the predicted rates of relative motions. More importantly, the fourth generation NUVEL-1 model listed in Table 3 is *consistent,* at the 1–2 mm/yr level, with the hypothesis that major plates behave *rigidly* over a million-year time scale. Moreover, there is growing evidence that the rates-of-change of geodetic baselines spanning plate boundaries are consistent with the geological estimates (e.g., Herring et al., 1986), provided that the endpoints are located within stable plate interiors. (This is comforting, both as a check on the techniques and as corroboration of the geophysical expectation that the instantaneous velocities between points in stable plate interiors are dominated by secular plate motions.) This means that, for those plates whose motions are well constrained by geological observations, direct geodetic measurements will not add significant constraints to the estimate of secular velocities. In any case, given the level of internal consistency of the geological models, to contribute to the improvement of existing models of present-day motion among the major plates, the tangential components of relative velocities on interplate baselines with endpoints located within stable plate interiors must be resolved to an accuracy on the order of 1 mm/yr.

It is therefore clear that geologically-based global plate motion models provide in fact a *kinematic reference frame* in which to analyze short-term geodetic observations, as well as the *kinematic boundary conditions* which must be satisfied by models of plate boundary deformation zones. In other words, the most important and interesting geodetic signals with characteristic time scales ranging from 1 hour to 100 years are detected as *departures* from predictions of million-year average rates based on geological rigid-plate models.

However, there exist examples for which the reliability of currently available global models is difficult to assess, and some improvement could be made from space-geodetic data. For example, Southeast Asia is assumed to be part of the large Eurasian plate, but the active tectonics of China imply it should be moving as a separate entity. Because it is completely surrounded by complex zones of deformation, its motion relative to Eurasia will be difficult to quantify without space-geodetic observations. Similarly, we note that geological rates are lacking across convergent boundaries, such as trenches. This is one reason why the motion of the Philippine plate relative to its neighbors is not well known. Again, space geodesy should provide useful constraints. Closer to home is the case of the Pacific-North America plate pair, whose relative rate of motion can be directly measured only on a tiny ridge segment in the mouth of the Gulf of California. The recent modeling work of DeMetz et al. (1987) (reflected in NUVEL-1) yields a relative plate velocity along that boundary that is 15–20% slower than earlier estimates. Since this rate is critical to models of deformation in the western United States, a geodetic check on the Pacific-North America angular velocity will be very valuable.

With these exceptions and a few others, however, the global networks of VLBI and SLR stations are just too sparse to control plate motions as tightly as the geological observations. The impact of geodetic observations on the development of these plate-motion standards will be relatively minor, at least for the next few years. Of course, this does not say that interplate observations made by space-geodetic methods will not reveal new and interesting phenomena associated with other categories of motion listed in Table 1; particular attention should be focused on time-dependent signals, including the possibility that plate speeds and directions have changed significantly during the 2-My averaging period of the geological data. Our point is to emphasize that *the exciting issues for space geodesy lie beyond the now-classical descriptions of major plate motions.*

Departures from rigid-plate motions. The various types of departures from the predictions of the rigid-plate models fall into two major categories: (1) large scale and regional scale non-rigid behavior (plate deformation and plate boundary zones of deformation), and (2) non-steady motions, including in particular post-seismic strains and aseismic deformations.

Plate deformation. There are conspicuous instances where the ideal rigid-plate model fails to describe adequately the complexities of present-day tectonic interactions, especially within the continents and along their margins. Examples can be found in regions undergoing compression due to plate collision, such as the Alpide Belt, and more particularly, Tibet, or the northern margin of the South American plate, as well as regions dominated by extensional tectonics, such as the African Rift Zone and the western U.S. Such regions are often characterized by spectacular landscapes and have long attracted the attention of geologists.

Perhaps the most outstanding example of large-scale continental deformation is Tibet. Collision of the Indian subcontinent with the Asian continent about 50 million years ago has been followed by roughly 2000–4000 km of convergence across Tibet and the Himalayas, resulting in the most impressive region of young and active deformation on earth (Molnar and Tapponnier, 1975, 1978). Although it is generally accepted that deformation within Tibet has resulted in movement of crustal fragments, with dimensions of tens to hundreds of kilometers, in directions oblique or even orthogonal to the overall direction of convergence between India and Asia, the details of present rates and directions of motion are still sketchy at best (e.g., Lyon-Caen and Molnar, 1985; Molnar et al., 1987). The use of space-geodetic techniques to unravel the complex kinematical picture of this region, even at the reconnaissance level of detail, should certainly be recognized as an important target for the next two decades.

Another young example of active tectonics which space geodesy will help understand better is the Mediterranean region. There we have “back arc” type basins adjacent to zones of coeval subduction and convergence, such as the Aegean-Hellenic systems, the Tyrrhenian-Appennine/Calabrian system, and the Pannonian-Carpathian system (see, for example, Malinverno and Ryan, 1986; McKenzie, 1972, 1978; Mercier, 1977; Royden et al., 1983; Scandone, 1979). These continental systems present an excellent opportunity to study the interaction of active extensional and convergent processes because, unlike most oceanic systems, a reasonably large amount of the region is exposed above sea-level, so that geodetic and geologic field studies are practical. Again the importance of reaching a more precise understanding of the current tectonic evolution of the area is enhanced by the high population density. Accordingly, a substantial long-term multi-national effort, the Wegener-Medlas Project, has been undertaken in 1984 to refine the kinematic picture of the Mediterranean region, and has begun to yield SLR data capable of placing useful constraints on the geological models (e.g., Wilson, 1987). The continuation of this project on a regional scale, and the densification of the network in critical areas (using, for example, GPS campaigns) will remain an important component of tectonic studies of the region in the coming years.

A particularly interesting third example is the western United States, where the interaction between the North American plate and the northwestward moving Pacific plate is spread out over broad zones of deformation, and where the available geodetic data sets are substantially larger. Although California's San Andreas Fault can be identified as a major locus of movement on the Pacific-North America plate boundary, the likelihood that significant crustal deformation is occurring both east and west of the San Andreas has long been recognized by geologists. Geological and geodetic observations of the present rate of slip along the fault in central California (about 34 mm/yr) is significantly lower than the rate predicted from successive generations of rigid-plate motion models, including NUVEL-1 (about 50 mm/yr). This so-called “San Andreas discrepancy” has been analyzed by Minster and Jordan (1984, 1987), using both geological information and geodetic data. Their conclusion was that even the short 4 year record of VLBI observations available to them for the relevant baselines (see Figure 2) was sufficient to place useful constraints on the integrated deformation east of the San Andreas across the Basin and Range, and by vector addition, on the integrated deformation west of the San Andreas across the California margin (see also Weldon and Humphreys, 1985).

Jordan and Minster (1988a) reviewed the use of space-geodetic observations to solve geological problems, with a focus on the various types of *secular horizontal* motions listed in Table 1. They concluded that many geological and geophysical problems related to such motions are indeed currently being addressed by space-geodetic experiments, provided that critical measurements are made at accuracies not feasible by conventional techniques. In particular, they state that measuring the velocities between crustal blocks to ±5 mm/yr can place geologically useful constraints on the integrated deformation rates across continental plate-boundary zones, such as the western U.S., the Mediterranean and Tibet.

However, it must be emphasized that baseline measurements in geologically complicated zones of deformation are useful only to the extent that the relationship of the endpoints to geologically significant crustal blocks is understood. Some antennas have a long history of participation in VLBI experiments, so that their motions in the VLBI reference frame are becoming well known; but they lie within complex zones of faulting, and their motions in kinematical frames fixed to local geology are not at all known. For example, the baseline rates for the Owens Valley Radio Observatory (OVRO) in California relative to the Westford, Massachusetts, and Ft. Davis, Texas, antennas have been measured to a precision of about 2 mm/yr (Figure 2). In their analysis of Basin and Range extension, Minister and Jordan (1987) have assumed OVRO moves with the Sierra Nevada-Great Valley block, but unfortunately, it is separated from the Sierra Nevada by a major system of faults, one of which broke in the great 1872 Owens Valley earthquake. Until the position of OVRO is regularly resurveyed in a local geodetic network which includes stations planted firmly on the Sierra block, the geological implications of the VLBI data will remain in doubt. Consequently, we can recommend that the establishment of frequently (or even continuously) surveyed local geodetic networks of sufficient density, around major geodetic sites in active areas should receive high priority.

The discussion has been focused so far on rather localized deformation, that is, on the occurrence of deformation zones with horizontal scales significantly less than overall plate dimensions. Whether the major tectonic plates, which are found to behave rigidly to an excellent approximation over million-year time scales, are actually undergoing nonlocal steady or episodic deformation on shorter time scales, is another important scientific question which bears directly on our understanding of the mechanical behavior of the lithosphere on a large scale, as well as our models of the force systems that drive plate tectonics. In order to resolve this issue, we will require geodetic coverage on a *global* scale, using techniques capable of delivering mm/yr accuracy for baselines 10,000 km long. Only space geodesy can meet such requirements, and, in the absence of any kind of “ground truth”, we must rely on intercomparisons between independent techniques to evaluate the actual performance of the systems.

Nonsteady motions. The problems of time dependence of the motions and non-rigid behavior of the plates, two issues which lie beyond the now-classical descriptions of major plate motions, are exciting applications of current and future space geodetic systems. The fundamental underlying scientific problem is the *transmission of strain (or stress)* in the lithosphere. This question is intimately related to the problem of coupling between earthquakes, evolution of volcanic eruption, and ultimately to the problem of *predicting* catastrophic events.

The time scales involved range from a fraction of an hour to centuries or longer, and span a range in which the physical phenomena are very poorly understood, primarily because the measurements are sparse, infrequent, and mostly very recent. Thus the evidence for episodic (as opposed to steady) motions along plate boundaries and within plate boundary deformation zones is insufficient at the present time to map the time and spatial scales involved.

A space-geodetic system capable of high sampling rate and dense spatial coverage over large areas will make possible a nearly unprecedented exploration of how crustal deformation varies with time. Much too little is known about this, the only data so far available comes from geodetic measurements that are too infrequent, too insensitive, or too localized to provide conclusive evidence. Space geodesy will provide information that is crucial to understanding the physics of the earthquake process. The motivation for these measurements can be understood in the context of a simple model of the mechanical properties of the crust and upper mantle (Figure 3).

Rocks respond to applied stress in a way very similar to the children's toy, Silly Putty, i.e., they are viscoelastic. They respond elastically to rapid variations in stress, but undergo irreversible deformation by creeping flow under low but sustained applied stresses. If the stresses are high enough, rocks fail brittlely; the result of this sudden failure is an earthquake. Low temperature and low confining pressures favor brittle failure, while high temperatures promote creep by decreasing the effective viscosity of rocks. The rheological behavior also depends upon rock type, crustal rocks being more prone to creep than mantle rocks at the same ambient conditions. Pore fluids and volatiles also have effects. Although the details are poorly understood--indeed, an important goal of geodetic monitoring is to *obtain the observations needed to understand the details*--the general variation in effective strength of the crust and upper mantle is illustrated schematically in Figure 3.

Large earthquakes generally nucleate near one of the maxima in the strength versus depth curve and propagate rapidly through the adjacent brittle region. However, the deformation associated with an earthquake is not limited to this seismic rupture. Aftershocks typically relieve stress on slip-deficient areas of the fault plane and extend the rupture to adjoining regions. Aseismic creep on the fault plane, perhaps also extending into the ductile regime, may increase the total amount of displacement associated with an earthquake. Viscoelastic adjustments of the Earth's crust and mantle cause redistribution of strain after a major earthquake. Stress is relieved by flow in the ductile regions, allowing elastic strain to accumulate in the stronger, more elastic regions.

A general feeling for the interaction of the viscous and elastic properties of rocks in the crust and mantle may be obtained by considering the simple model originally developed by Elsasser (1969) and extended by others (e.g., Melosh, 1976; 1977, 1983; Cohen, 1984; Rundle and Jackson, 1977; Rundle, 1988a,b). This model consists of an elastic layer of thickness h_e overlying a viscous layer of thickness $h_{\&vgr;}$ and a rigid base. An initial sudden displacement of a fault diffuses outward with a diffusivity $\kappa = h_e h_{\&vgr;}$/τ, where τ is the Maxwell time of the system. Repeated jerky offsets on plate boundary faults result in very smooth motion some distance away, corresponding to the steady velocities of plate interiors.

For regional scale problems, such as southern California (Figure 4), the elastic layer can be taken to be the upper crust, the viscouslayer the lower crust, and the rigid substratum is the upper mantle. The displacement from an earthquake on a given fault will diffuse through the crust, causing regional strain and perhaps influencing other faults, on a time scale that depends upon the effective viscosity of the lower crust (e.g., Turcotte et al., 1984). Crustal rheology is poorly constrained, but reasonable estimates indicate that strains may propagate 30 km in 50 years. As can be seen from Figure 4, there have been many earthquakes in southern California in the past 50 years with significant source dimensions. Based on Elsasser's model, these events would be expected to have viscoelastic strain migration associated with them. Observing this strain migration could constrain crustal rheology.

Over a scale of tens of kilometers around the fault, Thatcher (1983) showed geodetic evidence for time-dependent strain rates after great earthquakes on the San Andreas. He showed that his rather sparse data could be fit both by a model in which a viscoelastic asthenosphere underlays an elastic lithosphere, and by a purely elastic model with exponentially-decaying afterslip on the fault. Thus, the change in strain rates could be due to the constitutive properties either of the fault zone or of the asthenosphere, but we do not know which. It is crucial to separate these effects in order to understand better the physics of earthquakes and transmission of stress through the crust. The distribution of postseismic strain in time and space would provide important constraints on the physics of faulting and on the material properties of the fault zone and surrounding Earth. Redistribution of stress and strain to adjacent faults would provide important information related to earthquake prediction.

We discuss below possible time-varying strains that might be observed following earthquakes in the context of specific recent experience in southern California. We then examine another possible source of time-dependent strain not associated with individual earthquakes.

Post-seismic strains. A variety of observations have made it clear that the extremely rapid strain of a fault rupture is followed by strain rates that are high, compared to the long-term average measured for the fault zone, but the data are lacking to determine the physics responsible. To begin closest to the fault, ruptures that break to the surface commonly show substantial slip in the hours and days after the actual earthquake. Does this reflect equal amounts of slip at greater depth, or is the deeper part of the fault stable, this afterslip merely being caused by the gradual propagation of deep slip through the near-surface layers?

A particularly clear example of this kind of ambiguity has recently been provided in southern California by the Superstition Hills earthquake sequence of November 1987. The first large event was the Elmore Ranch earthquake (M_s~6.2), followed 12 hours later by the main Superstition Hills event (M_s~6.6), and of course many aftershocks.

Seismicity patterns and surface rupture showed two conjugate faults, the Elmore Ranch earthquake being on a fault conjugate to the San Jacinto fault zone and the mainshock on a fault (Superstition Hills) parallel to the zone. Figure 4 shows the location of the Pinon Flat Observatory (PFO) relative to these conjugate events (labeled “1987”).

We are fortunate to have available data from the long-base strain instruments at Pinon Flat Observatory (e.g., Wyatt et al., 1982); since these have been properly “anchored” to depth, they give results (out to periods of a year) that are better than any existing geodetic measurement. They thus provide a window, if only at one place, for what we might expect from future space-geodetic systems, and promise to be a source of “ground truth” for these measurements. Strains from these two events were recorded by two of the laser strainmeters at PFO; they are dominated by the coseismic offsets (e.g., Figure 5). These offsets and those recorded by other instruments at PFO are in reasonable agreement with results for a dislocation in a half-space, if seismically-derived parameters are used to define the dislocation. The frequent sampling of these data allows us to look at preseismic and postseismic deformations in some detail, with the best data coming from the fully-anchored NW laser strainmeter. During the interval between the two earthquakes, the largest signal on this record consists of microseisms plus some small residual drift in the instrument. We can certainly rule out any anomalous strains during this time above the level of about 3% of the eventual coseismic offset, and for the final 1000 seconds above about the 0.5% level.

It is hard to believe that the Superstition Hills earthquake was not triggered by the Elmore Ranch earthquake; but it is clear that no simple model of elastic strain and brittle failure can explain the 12-hour delay between events: if the Superstition Hills fault were close to failure, it should have ruptured during the dynamic strains generated by the Elmore Ranch earthquake, or at least soon after it, when the elastic stress changes had been fully imposed. We can think of several models, all speculative, to account for this.

1. All the surrounding material is elastic, and the Superstition Hills fault failed by brittle rupture at a critical stress level. The 12-hour delay was caused by afterslip on the Elmore Ranch fault; only after this had gone far enough was the applied stress sufficient for failure. This model would appear to be ruled out by the PFO data, which show no obvious afterslip (this would appear in proportion to the coseismic strain); any stress changes from this cause could be at most 10% of the stress changes at the time of the first earthquake. It could be that a small additional stress was enough, but this seems unduly *ad hoc.*
2. All parts of the system could have responded elastically, but the fault actually failed by some kind of stress corrosion. The model results of Tse and Rice (1986) show something like this; for a realistic friction law, earthquakes in their fault model begin with slow slip over a very small depth range, rapidly accelerating to seismic

slip, which would probably be consistent with the strainmeter data. However, this assumes a steady increase of applied stress; whether it still would result in inter-earthquake slip undetectable by the strainmeters would require additional modeling.

3. The Superstition Hills fault could have failed brittlely (as in the first hypothesis) with the delay between earthquakes being due to stress diffusion (Elsasser, 1969; Melosh, 1976, 1977, 1983) outward from around the Elmore Ranch fault. The physical model proposed recently by Rundle (1988a,b) may be invoked to account for the effects of interactions between faults (e.g., Rundle and Kanamori, 1987). We may assume that the first earthquake occurred in the brittle upper crust (Figure 4), underlain by a ductile lower crustal "asthenosphere." Just after the earthquake, the stresses in both regions are described by elasticity (for otherwise the coseismic strain step at PFO would not match the half-space solution), but the ductile zone then begins to flow, interacting with the overlying elastic layer to cause a diffusion of stress outward from the fault, and hence increasing the stress on the Superstition Hills fault. For the usual estimates of crustal viscosity, we would expect little change in stress over 12 hours. If, however, the rheology is time-dependent or obeys power-law flow, the effective viscosity close to the fault (where the largest stress changes occur) could be quite low, allowing rapid diffusion of stress, which would tend to slow down as stress levels smooth out. Again, more detailed modeling will be needed to see if such stress diffusion could occur without causing strains in the overlying lithosphere large enough to be detected by the distant strainmeters at PFO.

Space Geodesy could have helped distinguish between these models, if measurements had been collected in the time interval between events. Multiple surveys of an array of monuments around these faults would have provided the best evidence possible on changes of strain between them,; had none been seen, we would have good evidence for some kind of delayed failure on the fault itself, rather than a delay from stress propagation. *It is crucial that planning for space-geodetic systems allow a fast enough response time to make critical observations like these possible.*

Aseismic Deformations. Aseismic deformations, in the form of strain episodes not associated with seismic events, have been hypothesized. One of the classic examples is the now infamous Palmdale Bulge. Repeat leveling of the region near Palmdale on the "Big Bend" segment of the San Andreas fault (see Figure 5) showed an apparent uplift of up to 30 cm, followed by subsidence (Castle et al., 1976). This feature has been interpreted to result from episodic slip on a horizontal slip zone in the lower crust (Thatcher, 1979; Rundle and Thatcher, 1980). It was later discovered that a least part of the inferred bulge was the result of atmospheric refraction errors (Holdahl, 1982), with an additional smaller error due to miscalibration of survey rods (Jackson et al., 1983). The revised amplitude of the uplift is close enough to measurement error to be ambiguous (Stein, 1987), leading some geophysicists to dismiss the entire phenomenon. The controversy still rages, pointing out the problems inherent in interpreting infrequent measurements where the tectonic signal expected is close to the noise level.

There is less controversial evidence from conventional laser trilateration surveys across the San Andreas fault near Palmdale for the existence of a strain event in 1979 (Savage and Gu, 1985). Their paper suggests a five year long period of strain accumulation perpendicular to the San Andreas which was recovered abruptly in a $\sim 10^{-6}$ strain event in 1979. Unfortunately, given the measurement error and the possible effects of aliasing on such sparse (~yearly) occupations, these results by themselves are not definitive. They are tantalizing, however, since completely independent observations suggest that these time-dependent strains may be real. Gravity (Jachens et al., 1983), water level in wells (Merifield et al., 1983), and seismicity (Sauber et al., 1983) in this region all showed unusual changes in 1979. There is a good correlation between variations in these quantities, suggesting that the changes are real, despite the relatively large uncertainties of the individual measurements. There is also intriguing structure in the various (sparse) time series, suggesting that there is time variation on time scales of a month or less that is missed by the rather infrequent measurements, i.e., that aliasing is a serious problem.

Jachens et al. (1983) compared variations in gravity, in dilatational strain, and in elevation (determined by repeat leveling) over several other areas of southern California. Their results show a remarkable tracking of variations in these quantities with relative amplitudes consistent with what is expected based on simple elastic compression, suggesting that they are real, with periods of from months to over two years. Unfortunately, the sampling is sparse enough in time that once again the possible effects of aliasing are a problem. Savage et al. (1987) also found evidence for a strain fluctuation in 1984, but comparisons of geodetic and strainmeter data strongly suggest systematic error in the former. In contrast, the data from PFO have never shown a strain change as large as discussed above, and in indeed the best records from fully-anchored instruments show that strain fluctuations are very small. We thus have several possibilities:

1. Strain fluctuations of sizes seen in 1979 (perhaps?) are fairly common in some regions (although the EDM data for southern California show only one event of this size), but have not been seen at PFO. It may be significant, for example, that Palmdale lies in the Big Bend region, where mantle convergence and downwelling may be occurring (e.g., Bird and Rosenstock, 1984; Humphreys et al., 1984), while PFO is in a simpler strike-slip regime. Or perhaps PFO is atypically quiet, somehow decoupled from its surroundings.
2. Large strain fluctuations occur, but are rare in time and localized over distances shorter than 100 km. Their absence at PFO thus is no more than a consequence of small-sample statistics.

3. The very small changes in strain seen at PFO are typical of most parts of southern California; strain accumulation is mostly steady. The 1979 episode is merely due to larger than expected instrumental error.

There is no way to settle this question without more frequent measurements over more baselines at higher precision, and this is what the next generation of space-geodetic systems will provide. Only with such systems will we be in a position to characterize and understand the *spatial distribution* and the *time dependence* of deformation within tectonic regions, from which constraints on the physics of the deformation process can be inferred.

TYPES OF GEODETIC SIGNALS AND MEASUREMENT TECHNIQUES

The arguments presented so far lead us to conclude that the most significant departures from rigid-plate motions occur in zones of 100 to 1000 km width, within which differential motions are accommodated by a combination of seismic slip and aseismic deformation. However, because direct measurements of the kinematic evolution of geological systems using space-geodetic techniques can only cover time intervals of a few years, we are faced with the need to extrapolate these measurements to geological time scales in order to interpret them. In so doing, it is important to emphasize the space-time organization of geological systems and the processes that control their evolution. In fact, we seek a generic understanding of complex systems and their evolution, with special attention paid to transient behavior. If we are to trust the validity of extrapolations to geological time scales, a *process-oriented* approach appears to be superior to an *observation-driven* approach. Only then can we take full advantage of other data sources. These include, for example:

1. terrestrial geodesy, for which data sets spanning the last century are available (although not everywhere!) with dense spatial coverage (e.g., Snay et al., 1986).
2. classical geological techniques, which provide constraints pertinent to 10^3–10^6 year time scales and more, and spatial scales ranging from local outcrops to continental dimension.
3. geophysical data and models, which help us understand the underlying physics, and formulate testable hypotheses.

In order to examine quantitatively the potential of present and future space-geodetic systems to contribute to the solution of geological problems, we consider on Figure 6 a simple map of various general types of geodetic signals, in terms of the spatial dimensions and temporal scales of the underlying physical phenomena. The time scales of interest range from seconds in the case of brittle seismic fracture to 10^6 years for plate tectonics. Similarly, spatial scales to be considered range from a fraction of the crustal thickness to the dimensions of continents. We note that geological techniques are primarily useful to constrain phenomena mapped in the upper right corner of the graph, whereas the lower left corner is the realm of seismology. The vast domain occupying most of the figure is left for geodesists to study.

To compare this signal map with the capabilities of various geodetic tools, we use a very simple parametric model of the precision limit σ of any given instrumental technique. Specifically, we assume that it is given by $\sigma^2 = \alpha^2 + (bl)^2$, where α is a lower limit independent of the spatial scale (or "wavelength") λ and β is a coefficient of proportionality. Given this model, we examine the likelihood that a single event will be detected *and correctly characterized as to spatial and temporal scales* by selected techniques. The events considered here correspond to geodetic signals, expressed in terms of the displacement of a monument by a fixed fraction γ of the spatial scale λ, over the time scale τ. The model for σ allows us to construct a detection map in the (λ, τ) plane. Estimates of α and β yield the precision of a single observation, and, for simplicity, we take the noise to be Gaussian. In addition, we also truncate the detection maps according to the following considerations:

1. For any given technique, we assign a largest value of λ, λ_{max}, beyond which the technique is inapplicable or inaccurate, and a smallest value λ_{min}, below which *routine* deployment is probably not practical, because of logistical or cost considerations. In practice, we should truncate the map for values of λ smaller than about twice the site spacing, corresponding to the spatial "Nyquist" wavenumber that could be resolved by the observations.
2. Similarly, the frequency of measurements allowed for each technique is not taken to be the highest frequency achievable, but a realistic re-occupation rate for a routinely re-surveyed network. Some notion of logistics and costs is therefore built into the detection map. The detection map is truncated for time scales smaller than that corresponding to the Nyquist frequency of the measurement time series, that is, for time scales smaller than twice the time interval between measurements.
3. Finally, although we allow some degree of noise reduction if repeated measurements are carried out, we limit the improvement to a maximum factor of 3, in view of the empirical observation that further statistical improvements are usually negated by uncontrolled systematic errors.

Figures 7 and 8 show the detection maps of various geodetic techniques for $\gamma = 10^{-7}$, and $\gamma = 10^{-8}$, respectively. These figures include detection maps for several existing systems, as well as two hypothetical ones, examined here for design purposes:

1. VLBI and SLR (for which we have assumed 3 measurements/year, with $\alpha = 1$ cm),
2. 1-color *Electronic Distance Measurement (EDM),* (monthly measurements, $\alpha = 3$ mm and $\beta = 2 \times 10^{-7}$),
3. dedicated 2-color EDM (daily measurements, $\alpha = 0.2$ mm and $\beta = 10^{-7}$),
4. "observatory" measurements, i.e., continuously-recording strainmeters and tiltmeters,
5. a high precision, moderate frequency hypothetical system (weekly measurements, $\alpha = 1$ mm and $\beta = 10^{-8}$, and
6. a moderate precision, quasi-continuous hypothetical system (hourly measurements, $\alpha = 10$ mm and $\beta = 10^{-8}$).

As γ decreases from 10^{-7} to 10^{-8}, most detection curves appear to migrate toward the upper right, with the conspicuous exception of the observatory instruments. These instruments have red noise spectra, which is another way of saying that they drift by larger amounts over longer times; they are thus too unstable to detect small strains with long characteristic time scales. There is, in principle, no restriction in spatial scale, in the sense that an instrument measuring strain over a short distance will also respond to a strain change with much larger characteristic spatial scale. However, there is no way to recognize that a strain event is widespread with a single observatory of this type, so we truncate the curve for λ on the order of the thickness of the elastic portion of the crust.

In both figures, it is clear that the hypothetical systems would fill a niche that no currently available geodetic technique is capable of filling. In particular, *if we can afford them,* either or both would permit investigations of the physics of earthquake coupling, and refinements of physical models of the spatial and temporal distribution of crustal strains that give rise to earthquakes. Whether we should prefer one over the other depends in part on the tradeoff that exists between the spatial and temporal scales we wish to investigate, but also very much on cost and ease of deployment. These diagrams illustrate the fact that space-geodetic techniques already have opened new windows for us to investigate tectonic phenomena, and that these windows will be even wider in the future.

EXTRAPOLATION TO GEOLOGIC CONTEXT

Unlike the situation in oceanic lithosphere, plate boundary activity within the continental lithosphere is commonly distributed over broad zones up to several thousand kilometers wide, which consist of complex networks of faults and folds. In these areas, it is not unusual for extensional, thrust and strike slip faulting to occur contemporaneously within relatively small neighborhoods. Rotation of crustal fragments is common. Even seemingly undeformed areas are probably affected by considerable internal deformation and are only "undeformed" in relation to the more intense deformation occurring along their boundaries. Moreover, not all or even most faults within a given plate boundary deformation zone are active at any one time. Studies in the western U.S. and in China have shown that displacement may shift from one set of faults to another on a time scale of less than one million years, and perhaps as short as 100,000 years, even when rates of displacement on the individual faults are 10 mm/yr or greater.

This degree of spatial complexity and temporal variability limits the insight that even the most accurate geodetic measurements can, by themselves, provide into the nature of deformation of the continental crust. When deployed in such an environment, regional strain nets can at best resolve present-day rates of motion between various parts of a broad plate boundary, and without doubt this has to be a significant achievement in its own right. What regional strain nets cannot provide are the answers to fundamental questions about how strain is in fact accommodated, in what ways and how quickly the spatial distribution of deformation may change with time within a geologic system, and how mechanical and dynamical connections unify disparate types of deformation into a single coherent tectonic picture. However, in combination with other geological and geophysical information, accurate geodetic measurements using carefully positioned stations will provide an extremely powerful tool with which to address specific well-posed questions about crustal deformation in a particular tectonic setting.

Revisiting tectonic examples used earlier, we may raise questions pertaining to the dynamical mantle phenomena underlying the surface kinematics observed by geodesy. For instance, the role of upper mantle flow under Tibet is virtually unknown. It is not known if mantle flow occurs on the same scale as movements of crustal fragments, or if the scale of the mantle flow is comparable to the scale of the entire plate boundary collision zone. In the former case, one would expect zones of local mantle downwelling beneath regions of major crustal shortening in Tibet (such as below the Tien-Shan). In the latter case, movements of crustal fragments must occur within a thin sheet that is largely decoupled from the behavior of the upper mantle beneath it. If mantle flow can be linked spatially to the active movements of sizable crustal fragments, then one might reasonably infer that the same has been true throughout the history of the Tibetan Plateau. This would imply that temporal changes in mantle flow occur over the same time scale as changes in the direction and rate of movements within the crust, which can be dated using traditional geological techniques. Preliminary data from northeastern Tibet show that, on a length scale of tens of kilometers, major deformation has switched from one system of faults to another and from one type of deformation to another over time intervals of less than a million years (Zhang, 1987). If similar temporal variability in fault activity occurs at even larger scales in Tibet, then the time scale over which changes in crustal motion occur may be very short indeed. Alternatively, if mantle flow can be shown to be related to crustal motion only at the scale of the plate boundary as a whole, then little can be said about temporal variability of flow in the mantle from reconstructing crustal movement histories. One might, however, infer that such large scale flow would be likely to vary only over the same time scale as the overall plate boundary activity, roughly tens of millions of years (for example, England et al., 1985).

Similarly, in the case of the Mediterranean, questions aimed at understanding the interaction between active thrusting and extension, and which are amenable to study with geodetic and geologic data might include: How do rates of active convergence vary along the subduction boundaries? How does the rate and direction of active extension in “back arc” regions vary within the overriding plate? Is there a correlation between the rates and directions of extension and the rates and directions of convergence? Do major changes in rate coincide with structural features such as lateral offset or variations in trend of the convergent boundary? The answers to these and other questions about active crustal processes in these and similar systems elsewhere are essential if one seeks to evaluate the interaction between active crustal and mantle processes.

Comparable arguments can clearly be made in the case of the western U.S.; what the mantle flow is beneath southern California or beneath the Basin and Range are outstanding geophysical questions. The point made here is that, through geodesy and through geophysical techniques such as seismic tomography, one obtains but a “snapshot” of the present state of complex dynamical systems. The time scales that govern such systems can be short by geological standards, but are typically long, compared to the lengths of geodetic time series. In order to progress in our understanding of the physics, we must turn to a joint interpretation of geodetic observations, geophysical data and models, *and* geological reconstructions of past history. In fact, the understanding of crustal dynamics depends almost exclusively upon geologic data to characterize how deformation has evolved and been re-distributed over time. Prediction of crustal dynamic activity is not possible without a specific knowledge of regional geological context. Clearly, a broad range of techniques must be brought to bear on large scale tectonic problems. For instance, conventional geologic field studies are crucial for providing a detailed geological context, but such detailed studies cannot be made over large regions. Remote sensing, including structural geologic interpretation, mapping of lithologies and soils, and characterization of weathering surfaces (including relative dating of fault scarps) provides the key to an informed extension of detailed study results to regional scales.

EMERGING LINES AND FUTURE DIRECTIONS

Strategy. Selection of a strategy for the continued development of space-geodetic solutions to outstanding geological problems in the next decade or two should be based on the recognition that resources and capabilities will remain finite, and in some instances, will be in fact somewhat limited. Consequently, we recommend the following guidelines:

1. **Focus on areas with major geological issues that require quantitative answers.** This entails the selection of areas and problems for which ***testable*** **hypotheses can be formulated, together with a preferred focus on recognizable geological problems, for which other data sets of adequate quality are available. In addition, the socio-economic importance of the problem and the potential human impact of eventual solutions should be recognized and taken into consideration.**
2. **Develop affordable technology.** This would have the desirable consequence that it would permit involvement of a greater proportion of the domestic and foreign scientific community. Further, it would allow improved spatial and temporal coverage within controlled costs, and often minimize the logistical burden of field work.
3. **Emphasize easily deployable systems.** In particular, we require a capability for *rapid* field deployment (on time scales of hours). Moreover, such systems would lead to lower personnel requirements, would allow easier access to remote areas, and would support both low resolution reconnaissance work, as well as detailed surveys of dense networks.
4. **Emphasize instrument calibration and measurement validation.** This entails the continued operation of the more burdensome, expensive systems at a level adequate to permit validation of results obtained by “lightweight” field systems. In addition, there will be a continued, and even enhanced need to maintain very high quality global and regional fiducial and reference networks. The systematic intercomparison of independent systems (e.g., VLBI, SLR, GPS) should be retained as an important validation technique, and we strongly recommend a careful examination of monumentation issues (particularly long-term stability), if future space-geodetic systems are used to monitor very large, dense networks.

Point Positioning Objectives. Scientific objectives for Point Positioning activities during the next decade include, for example:

1. to collect and analyze suitable “baseline” data to support the design and implementation of dense space-geodetic networks, to diagnose sources of geodetic noise, and to provide independent calibration and validation data;
2. to develop techniques for the *geological* interpretation of space-based ranging and altimetry data in plate-boundary deformation zones;

3. to formulate and attempt to solve significant problems in crustal deformation resolvable with the limited data (in time) that will be collected within the next decade.

Specifically, we should refine parameterized kinematic models of crustal deformation, and extend the analysis to include constraints from dense, frequently surveyed space-geodetic networks. At the same time, we must improve physical models of time-dependent deformation of the crust and upper mantle, and use these models in the planning of space-geodetic surveys and in the interpretation of the spatio-temporal strain patterns seen by space-geodetic techniques. Of fundamental importance for interpretation purposes will be the systematic comparison of space-geodetic observations to other, independent geophysical and geodetic data, such as observatory measurements of strain along short baselines, seismicity patterns, and other repeated geodetic and gravity surveys, in an effort to achieve a quantitative understanding of the phenomena which control volcanic and seismic cycles. As argued earlier, the elucidation of the time scales relevant to many important aspects of crustal dynamics will require integration of geodetic measurements with interpretations of local and regional geology. Finally, we will need to develop a systematic methodology for exporting approaches proven to be successful in a given region where accessibility is not an issue (e.g., the American west), to inaccessible regions with comparable or very different tectonic regimes (e.g., continental collision zones such as Tibet, trench boundaries such as the west coast of central and south America, strike-slip boundaries such as Turkey and New Zealand, and regions of back-arc deformation, such as the Aegean). Finally, in many instances, some of the critical structures will be submerged, so that a capability to locate precisely points on the seafloor relative to one another, and relative to a land-based geodetic network would be extremely valuable (e.g., Spiess, 1985).

CONCLUSIONS AND RECOMMENDATIONS

Over the next decade and into the next century, the priorities we assign to various aspects of remote sensing applied to the study of the solid earth must acknowledge the tradeoff that exists between the application of recent advances in the technological aspects of space geodesy to solve urgent geological and geophysical problems, and the need to press ahead and develop new and better capabilities. The time scales associated with some of the most critical aspects of the current evolution of our environment are poorly understood, but are thought to be on the order of decades. Consequently, we may not have the luxury to defer practical applications until better technology is at hand, but should in many cases undertake systematic studies on an unprecedented global scale. Important conclusions reached in this section include:

1. To contribute to the improvement of existing models of present-day motion among the major plates, the tangential components of relative velocities on interplate baselines with endpoints located within stable plate interiors must be resolved to an accuracy of about 1 mm/yr.

2. The most important and interesting geodetic signals averaged over 1 hour to 100 years take the form of *departures* from predictions of million-year average rates based on geological rigid-plate models.
3. The most significant departures from rigid-plate motions occur in zones of 100 to 1000 km width, within which differential motions are accommodated by a *combination* of seismic slip and aseismic deformation.
4. Measuring velocities between crustal blocks to ±5 mm/yr will provide geologically useful constraints on the integrated deformation rates across continental plate-boundary zones, such as the western U.S. and Tibet.
5. It is only through the integration of geodetic and geologic field studies that active deformation can be related to earlier activity within an evolving tectonic system. Coordination of geodetic and geologic studies is therefore necessary to establish the temporal dependence of crustal deformation on a geologic time scale (e.g., Royden, 1988).
6. The establishment of frequently (or even continuously) surveyed local geodetic networks of sufficient density, around major geodetic sites in active areas, should receive high priority.
7. The development and systematic deployment of affordable and easily deployable space-geodetic systems with cm to mm precision and high sampling rates--that is, “occupation” frequency ranging from hourly to weekly--will permit investigation of geophysical phenomena, particularly the earthquake cycle, in a range of spatial and temporal scales never explored before.
8. Space-geodetic observations yield constraints on crustal kinematics; to achieve an improved understanding of the dynamics and thus, a better grasp of the underlying physical phenomena, we must rely on a broad combination of geophysical and geological observations as a way to extend the geodetic signals to longer time scales and to extrapolate surface information to crustal and mantle depths.

Based on the discussion held at the Erice workshop, and in view of the conclusions listed above, the panel on the “Long-Term Dynamics of the Solid Earth” formulated the following recommendations concerning the continued development and application of precise point positioning techniques:

Over the next 20 years, major efforts in applying precise positioning techniques should be aimed primarily at:

a. Continued large scale <u>*reconnaissance*</u> *surveys with station spacing on the order of* 10^2 km, to improve our understanding of the *kinematic evolution of extensive, largely unexplored zones of continental deformation.*

b. Sustained, repeated measurements of dense networks at centimeter-level accuracy, to determine the time dependence and spatial distribution of deformation within and across zones of intense tectonic activity. Measurement frequencies should range from daily to annually over a decade or more, with station spacing from 3 to 30 km, and network dimensions from 10 to 1000 km. In regions of complex deformation, geodetic measurements should be complemented by comprehensive tectonic and structural studies and careful estimates of displacements and displacement rates on geologic time scales.

c. Continued improvement of capabilities, to achieve:

--millimeter-level accuracy in both horizontal and vertical components for detailed subaerial studies, system calibration, and ultimately, low-cost routine deployment.

--centimeter-level accuracy in both horizontal and vertical components for sea-bottom systems.

REFERENCES

Bird and Rosenstock, Geol. Soc. Amer. Bull., 95, 946–957, 1984.
Castle et al., Science, 192, 251–253, 1976.
Chase, Geophys. J. R. Astr. Soc., 29, 117–122, 1972.
Chase, Earth Planet. Sci. Lett., 37, 353–368, 1978
Christodoulidis et al., J. Geophys. Res., 90, 9249–9264, 1985.
Cohen, J. Geophys. Res., 89, 4538–4544, 1984.
DeMets et al., Geophys. Res. Lett., 14, 911–914, 1987.
DeMets et al., 1989, (in preparation).
Elsasser, in The Application of Modern Physics to the Earth and Planetary Interiors, S. K. Runcorn, ed. 223–246, 1969.
England et al., J. Geophys. Res., 90, 3551–3557, 1985.
Gordon et al., presented at NASA CDP P.I. Meeting, Munich, Germany, Oct. 17–21, 1988.
Herring et al., J. Geophys. Res., 91, 8341–8347, 1986.
Hodahl, J. Geophys. Res., 87, 9374–9388, 1982.
Humphreys et al., Geophys. Res. Lett., 11, 625–627, 1984.
Jachens et al., Science, 219, 1215–1217, 1983.
Jackson et al., Tectonophysics, 97, 73–83, 1983.
Jordan and Minster, in The Impact of VLBI on Astrophysics and Geophysics, (M. J. Reid and J. M. Moran, eds.) Proc. IAU Symposium 129, 341–350, 1988a.
Jordan and Minister, Scientific American, August 48–58, 1988b.
LePichon, J. Geophys. Res., 73, 3661–3697, 1968.
Lyon-Caen and Molnar, Tectonics, 4, 513–538, 1985.
Malinverno and Ryan, Tectonics, 5, 227–245, 1986.
McKenzie, Geophys. J. Roy. Astr. Soc., 30, 109–185, 1972.
McKenzie, Geophys. J. Roy. Astr. Soc., 55, 217–254, 1978.
Melosh, J. Geophys. Res., 81, 5621–5632, 1976.
Melosh, Pure Appl. Geophys., 15, 429–439, 1977.
Melosh, Geophys. Res. Lett., 10, 47–50, 1983.
Mercier, Bull. Geol. Soc. Fr., 7, 663–672, 1977.
Merifield et al., Tech. Rep. 83-3, Lamar-Merifield Geol. Inc. Santa Monica, CA, 1983.
Minster and Jordan, J. Geophys. Res., 83, 5331–5354, 1978.
Minster and Jordan, Pac. Sec. Soc. Econ. Paleontol. Mineral, 38, 1–16, 1984.
Minister and Jordan, Geophys. Res., 92, 4798–4804, 1987.
Minister et al., Geophys. J. Roy. Astr. Soc., 36, 541–576, 1974.
Molnar and Tapponnier, Science, 189, 419–426, 1975.
Molnar and Tapponnier, Geophys. Res., 83, 5361–5375, 1978.
Molnar et al., Geology, 15, 249–253, 1987.
Royden et al., Tectonics, 2, 63–90, 1983.
Royden, in Geodetic Studies and Crustal Dynamics, U.S. Geodynamics Committee Progress Report, 3.1–3.10, 1988.
Rundle, J. Geophys. Res., 93, 6237–6254. 1988a.
Rundle, J. Geophys. Res., 93, 6255–6274, 1988b.
Rundle and Jackson, Geophys. J. Roy. Astr. Soc., 49, 575–592, 1977.
Rundle and Kanamori, J. Geophys. Res., 92, 2606–2616, 1987.

Rundle and Thatcher, Seismol. Soc. Amer. Bull., 70, 1869–1886, 1980.
Sauber et al., J. Geophys. Res., 88, 2213–2219, 1983.
Savage, Ann. Rev. Earth Planet. Sci., 11, 11–43, 1983.
Savage and Gu, J. Geophys. Res., 90, 10301–10309, 1985.
Savage et al., J. Geophys. Res., 92, 4785–4797, 1987.
Scandone, Bull. Soc. Geol. Ital., 98, 27–34, 1979.
Snay et al., Royal Soc. New Zealand Bull, 24, 131–140, 1986.
Spiess, IEEE Trans. on Geoscience and Remote Sensing GE23(4), 502–510, 1985.
Stein, Rev. Geophys., 25, 855–863, 1987.
Thatcher, J. Geophys. Res, 84, 2351–2370, 1979.
Thatcher, Nature, 299, 12, 1983.
Tse and Rice, J. Geophys. Res., 91, 9452–9472, 1986.
Turcotte et al., J. Geophys. Res., 89, 5801–5816, 1984.
Weldon and Humphreys, Tectonics, 5, 33–48, 1985.
Wilson, Geo journal, 14.2, 143–161, 1987.
Wyatt et al., Bull. Seismol. Soc. Amer., 72, 1701–1715, 1982.
Zhang, Rate, Amount, and Style of late Cenozoic Deformation of Southern Ningxia, Northeastern Margin of Tibetan Plateau, Ph.D. Thesis, Mass. Inst. of Tech., Cambridge, MA, 1987.

TABLE 1. Types of motions at the earth's surface.

	SECULAR	TRANSIENT
HORIZONTAL	Plate motions Boundary-zone tectonics Intraplate deformation	(Pre,co,post)-seismic Fault creep Stress redistribution
VERTICAL	Tectonic motions Thermal subsidence Diapirism Crustal loading Post-glacial rebound Cratonic epeirogeny	(Pre,co,post)-seismic Magma inflation Tidal loading

Table 2. Data sets used in successive generations0 of plate monon models

	Magnetic rates	Transform faults	Slip vectors	Total data
2nd generation, e.g. RM1[(1)], 1974	68	62	106	236
3rd generation, e.g. RM2[(2)], 1978	110	78	142	330
4th generation, NUVEL-1[(3)], 1988	277	121	724	1122

[(1)] Minster at al., (1974). [(2)] Minster and Jordan, (1978). [(3)] Gordon et al., (1988).

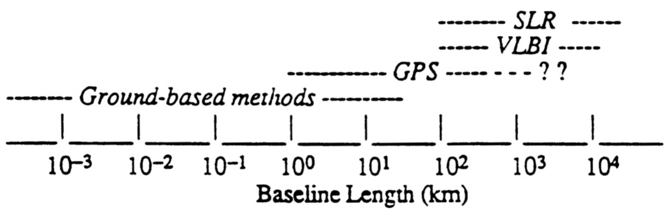


Figure 1 Spatial scales sampled by various geodetic methods.

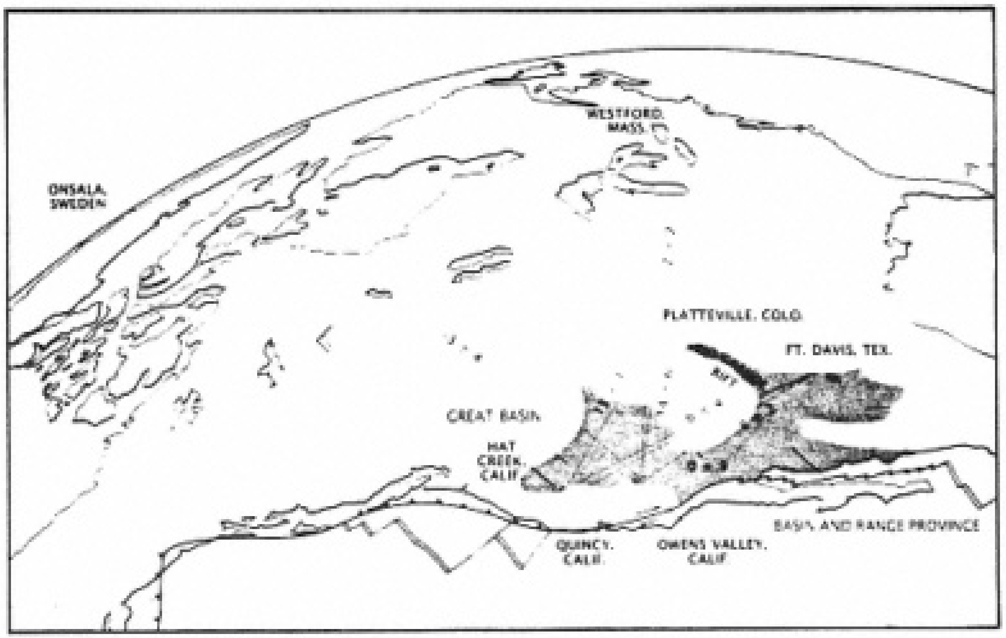


Figure 2. The San Andreas fault in central California is one element of the western U.S. plate boundary zone. In addition, we must account for contributions from crustal deformation both east and west of the fault. The integrated deformation west of the fault, which consists of NW-SE extension across the Great Basin, can be measured directly from the rates-of-change of VLBI baselines monitored by NASA's Crustal Dynamics Project. Combining this estimate with the geologically and geodetically observed rate of slip on the San Andreas, and the rigid-plate estimate of total motion between the Pacific and North America plates, we can derive geologically useful constraints on the integrated rate of deformation across the California margin, west of the fault. (From Jordan and Minster, 1988b. Reproduced with permission and by courtesy of Scientific American.)

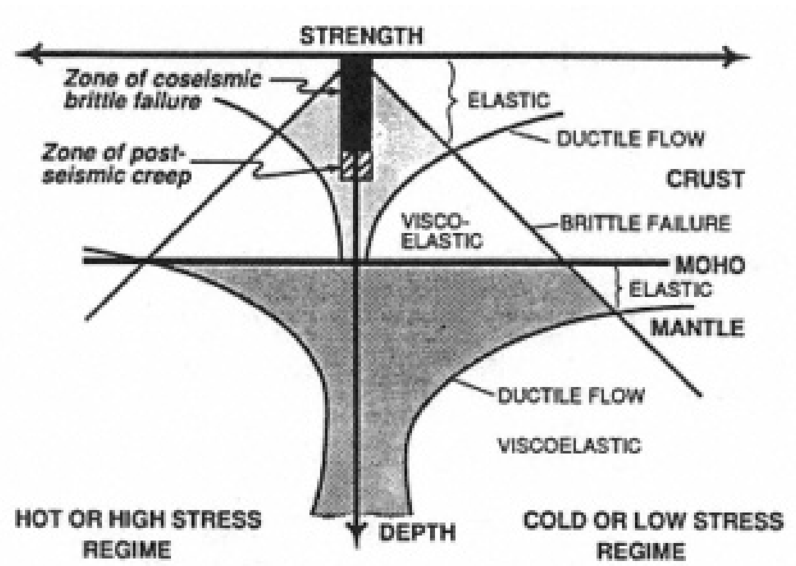


Figure 3. Schematic diagram of the strength and mode of deformation of the crust and upper mantle. Postseismic deformation occurs as the result of creep on the extension of the fault plane or stress relaxation in the viscoelastic regions.

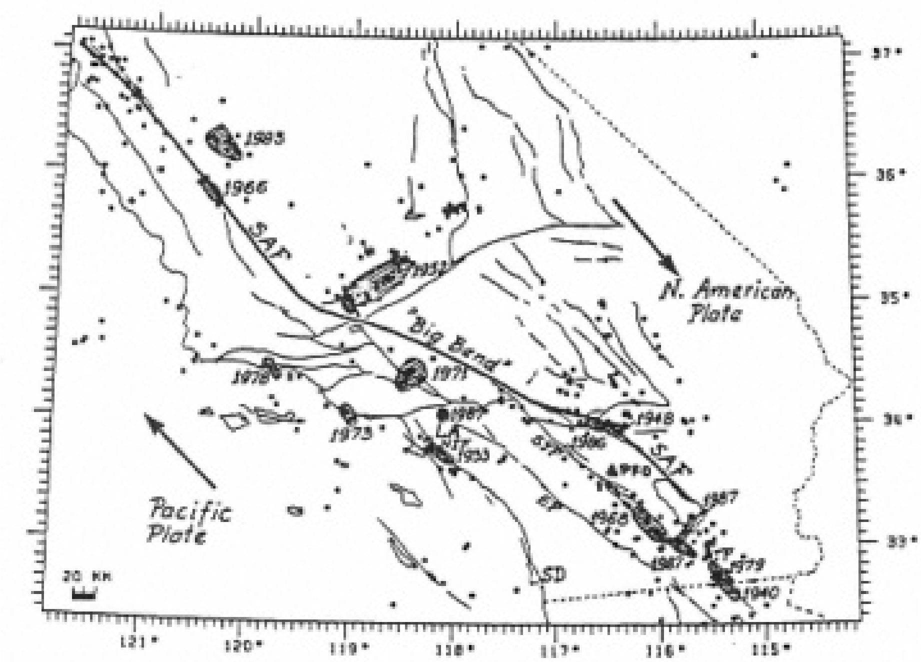


Figure 4. Major fault lines and earthquake epicenters in southern California. All events of magnitude >4.5 in the past 55 years are shown, with rupture zones of the largest events shaded. Fault abbreviations SAF, San Andreas; SJF, San Jacinto; EF, Elsinore; NIF, Newport-Inglewood; GF, Garlock; and IF, Imperial. PFO is the Pinon Flat Observatory.

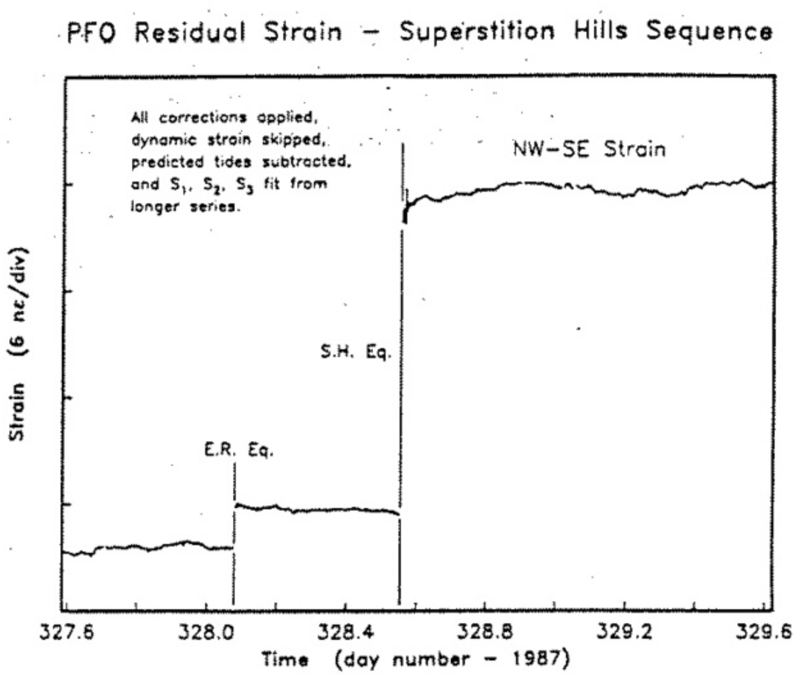


Figure 5. Superstition Hills events recorded at The Piñon Flat Observatory (See Figure 4).

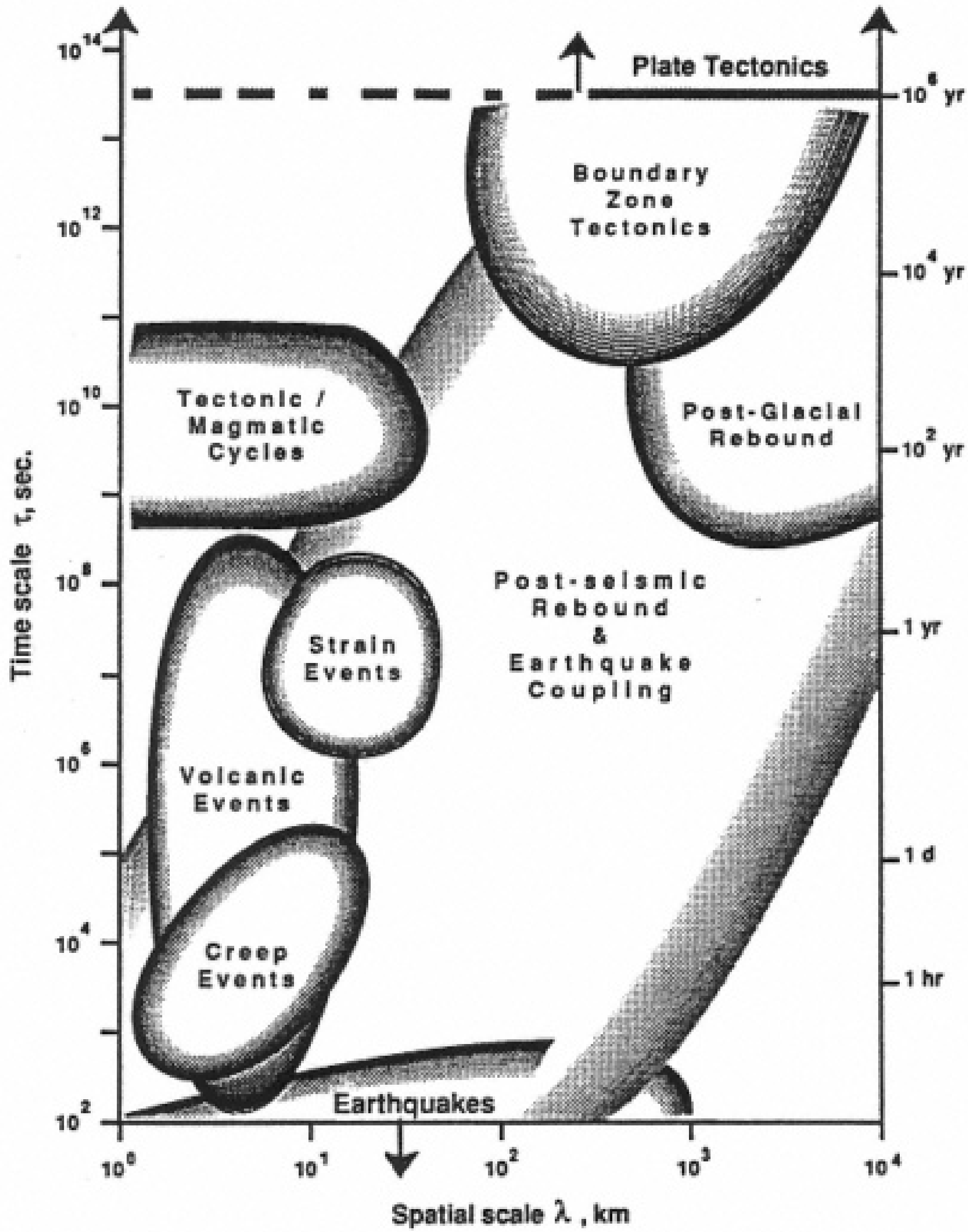


Figure 6. Map of geodetic signals in terms of spatial and temporal scales.

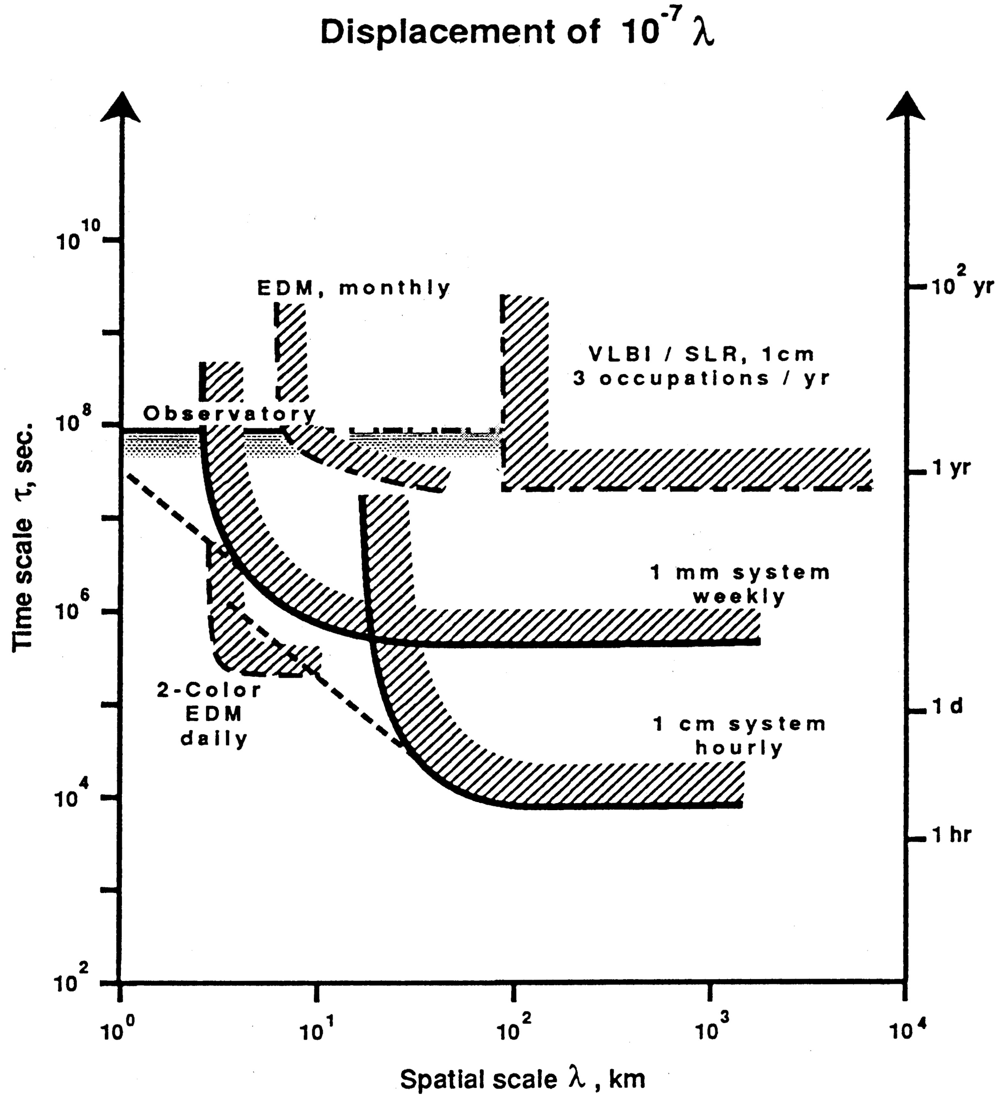


Figure 7. Detection capability of various geodetic techniques at the 10^{-7} strain level.

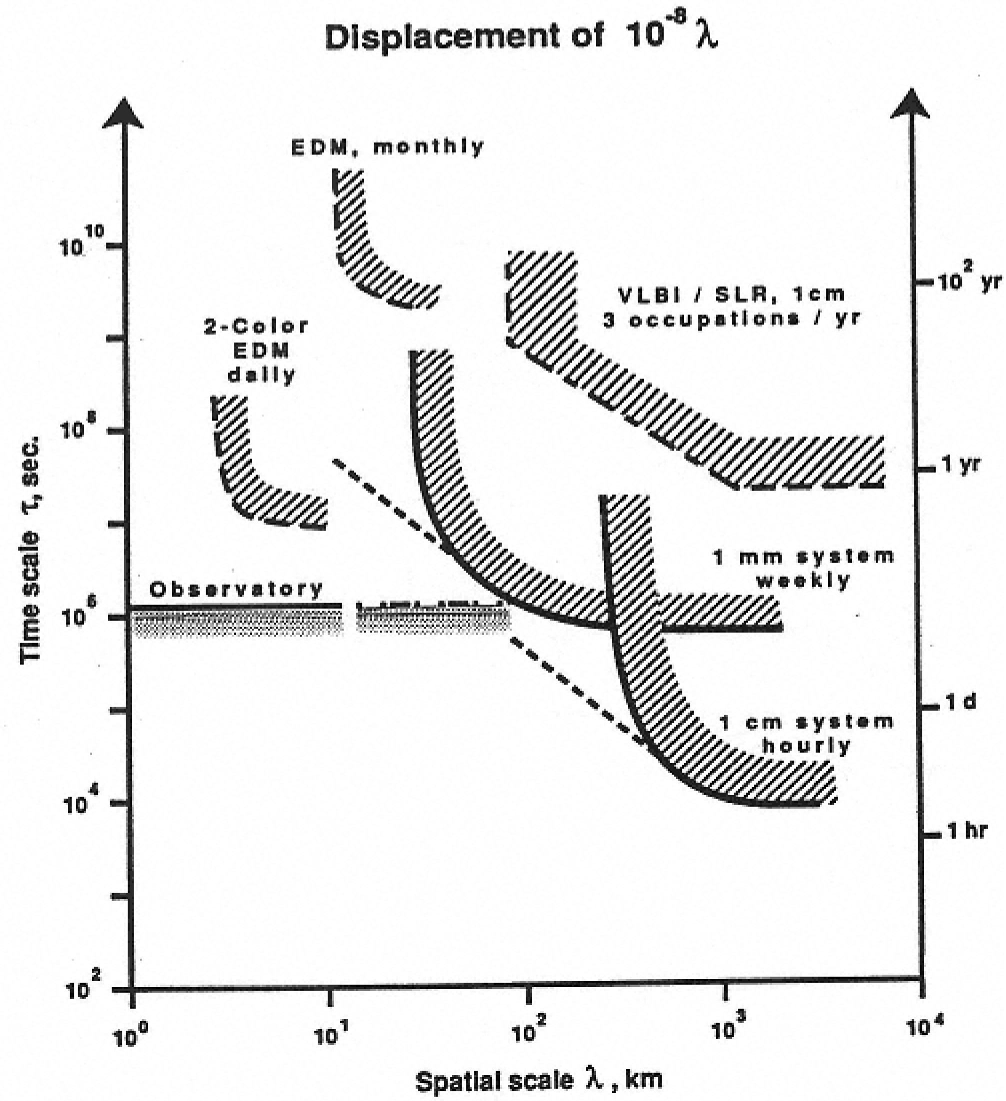


Figure 8. Detection capability of various geodetic techniques at the 10^{-8} strain level.

Table 3. EULER vectors for the rigid-plate model NUVEL-1 (from Gordon et al., 1988)

				Error Ellipse			
Plate Pair	Latitude °N	Longitude °E	ω (deg/My)	σ_{max}	σ_{min}	ξ_{max}	σ_{ω} (deg/My)
Pacific Region							
na-pa	48.7	-78.2	0.78	1.3	1.2	-61	0.01
co-pa	36.8	-108.6	2.09	1.0	0.6	-33	0.05
co-na	27.9	-120.7	1.42	1.8	0.7	-67	0.05
co-nz	4.8	-124.3	0.95	2.9	1.5	-88	0.05
nz-pa	55.6	-90.1	1.42	1.8	0.9	-1	0.02
nz-an	40.5	-95.9	0.54	4.5	1.9	-9	0.02
nz-sa	56.0	-94.0	0.76	3.6	1.5	-10	0.02
an-pa	64.3	-84.0	0.91	1.2	1.0	81	0.01
pa-au	-60.1	-178.3	1.12	1.0	0.9	-58	0.02
eu-pa	61.1	-85.8	0.90	1.3	1.1	90	0.02
co-ca	24.1	-119.4	1.37	2.5	1.2	-60	0.06
nz-ca	56.2	-104.6	0.58	6.5	2.2	-31	0.04
Atlantic Region							
eu-na	62.4	135.8	0.22	4.1	1.3	-11	0.01
af-na	78.8	38.3	0.25	3.7	1.0	77	0.01
af-eu	21.0	-20.6	0.13	6.0	0.7	-4	0.02
na-sa	16.3	-58.1	0.15	5.9	3.7	-9	0.01
af-sa	62.5	-39.4	0.32	2.6	0.8	-11	0.01
an-sa	86.4	-40.6	0.28	3.0	1.2	-24	0.01
na-ca	-74.3	-26.1	0.11	25.5	2.6	-52	0.03
ca-sa	50.0	-65.3	0.19	15.1	4.3	-2	0.03
Indian Ocean and African Regions							
au-an	13.2	38.2	0.68	1.3	1.0	-63	:0.00
af-an	5.6	-39.2	0.13	4.4	1.3	-42	0.01
au-af	12.4	49.8	0.66	1.2	0.9	-39	0.01
au-in	-5.5	77.1	0.31	7.4	3.1	-47	0.07
in-af	23.6	28.5	0.43	8.8	1.5	-74	0.06
ar-af	24.1	24.0	0.42	4.9	1.3	-65	0.05
in-eu	24.4	17.7	0.53	8.8	1.8	-79	0.06
ar-eu	24.6	13.7	0.52	5.2	1.7	-72	0.05
au-eu	15.1	40.5	0.72	2.1	1.1	-45	0.01
in-ar	3.0	91.5	0.03	26.1	2.4	-58	0.04

If Only We Had Better Gravity Data…

Marcia McNutt
Department of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology Cambridge, Massachusetts

INTRODUCTION

The Earth science community has entered an exciting new era in which, for the first time, the goal of understanding the full 3-dimensional structure of mass and energy transport within the Earth appears to be attainable. This revolution in Earth science, which will integrate the 2-dimensional surface kinematic pattern of plate tectonics with a 3-dimensional dynamic model, is largely the outgrowth of advances in global digital networks, supercomputing, and satellite geodesy. There is no question but that progress has been hindered by the lack of a high-resolution, extremely accurate, truly global gravity field. In compiling the following list of important problems that could be addressed with better gravity data, we did not confine ourselves to what might be achievable with any particular instrument or mission design. We recognize that a variety of approaches will be necessary, and that in many cases, improvements in the gravity field will be scientifically more significant if coupled with other geophysical and geological observations of similar quality.

OCEANIC LITHOSPHERE

Our knowledge of the gravity field over the oceans took a quantum leap forward with the advent of satellite altimetry in the 1970s. Over some areas of the oceans, we actually have better resolution in the gravity field than in the bathymetry. Information on the marine geoid from Geos-3 and SEASAT has led to a significant increase in our understanding of the thermo-mechanical structure and evolution of oceanic lithosphere from the midocean ridges where it is formed to the trenches where it is consumed. Nevertheless, a number of important problems remain to be solved because presently-available altimetric geoids lack sufficient accuracy, resolution, and/or continuity at shorelines. A sampling of some of these problems is given below.

Midocean Ridges. According to the theory of plate tectonics, midocean ridges (MORs) are a 2-dimensional volcanic line along which magma rises and accretes to the trailing edges of spreading plates in order to create new lithosphere. With the development of multi-beam swath mapping systems, we now view MORs as complex 3-dimensional structures (Figure 1) consisting of whole new classes of topographic features such as propagating rifts, overlapping spreading centers, 200 km undulations in axial relief, and more minor departures from axial linearity. It is thought that many of these features are the topographic expression of distinct, ephemeral magma chambers (Macdonald et al., 1987), which would produce gravity effects both by virtue of their density contrast as well as via their elastic deformation of the overlying plate (Madsen et al., 1984). Gravity data would provide key information on the origin and evolution of the features, but are, so far, unavailable at the requisite 1 to 2 mgals accuracy and 2 to 100 km resolution.

Fracture Zones. One of the more intriguing observations to emerge from analysis of the altimetric geoid is the failure of the simple plate model to describe the density structure of adjacent lithospheric plates of different ages across fracture zones. For example, Figure 2 shows the size of the geoid step across the Eltanin Fracture Zone as a function of average age of the contiguous lithosphere. Theoretical thermal models predict a constant geoid step at all ages for the half-space cooling model of the lithosphere, with the step gradually decreasing with increasing age for plate models such as that of Parsons and Sclater (1977). Contrary to these predictions, the observed geoid step rapidly decreases at young ages and suddenly reappears at older ages. A similar pattern has been observed on the Udintsev, Ascension, and Falkland-Agulhas fracture zones (Cazenave, 1984; Driscoll and Parsons, 1988; Freedman, 1987).

In addition to conductive cooling of thermal plates with differing ages, several other factors undoubtedly contribute to the geoid signature of fracture zones. Based on geoid, gravity, seismic, or topography studies, the effects of lithospheric flexure (Sandwell and Schubert, 1982), thermal stress (Parmentier and Haxby, 1986), crustal structure (Detrick and Purdy, 1980), peridotite intrusions (Fox et al., 1976), small-scale convection (Craig and McKenzie, 1986) and hot spot volcanism (McNutt et al., 1989), have all been suggested as significant. The best present geoid data do not have the accuracy or resolution to sort out the various contributions of these processes to the density structure at fracture zones. We need a gravity field accurate to 1 mgal at resolution of 50 km or less. To realize the full potential of such data, gravity field modeling must be also constrained by better topographic data from the oceans and seismic information on crustal structure. Until such data is forthcoming, we must question the adequacy of the thermal plate model in describing the density structure of fracture zones.

Subduction Zones. The largest gravity anomalies on Earth occur at trenches where oceanic lithosphere is subducted into the mantle. These zones are responsible for creating the greatest thermal, seismic, and geochemical anomalies found within the upper mantle. The underthrust plate is flexed and deformed by a number of loads, including stresses from motion relative to the convecting mantle, the weight of the overlying plate, the negative buoyancy of its own cold mass, thermal stress, and the density changes associated with phase changes in the mantle. With seasurface gravity we have observations (Watts and Talwani, 1974) and altimeter observations (McAdoo and Martin, 1984), we have been able to calibrate the rheology of the deformed lithosphere. Earthquake hypocenters (e.g., Isacks and Barazangi, 1977) and travel time anomalies (Creager and Jordan, 1984) provide maps of the geometry of the downgoing plate. Thermal plate models allow us to calculate the load associated with the cold slab (Toksoz et al., 1971). If we had a gravity or geoid map continuous from the undeformed seafloor, across the outer rise, trench, forearc, and island arc to the overriding plate, we would be able to calculate the stresses acting on the underthrust plate, and thereby learn much about lithosphere/asthenosphere interaction and aspects of mantle rheology, such as the degree to which the lower mantle resists slab penetration (Hager, 1984). The large amplitude of the anomalies leads to accuracy requirements of only 5 to 10 mgals at 100 to 200 km resolution for studying plate interactions, and 5 mgals at 1000 to 3000 km resolution for investigating mantle rheology. The necessity of having a field that spans the transition from ocean to continent leads to the requirement that at least some of the data be obtained using non-altimetric techniques.

Midplate Swells and Plateaus. The world's ocean basins contain more than 100 areas of elevated seafloor which extend at least 1000 km and stand more than 2 km above the adjacent oceanic crust. These features are generally classified as either oceanic swells, which are capped by active hot spot volcanoes, or plateaus, which display steeper margins and flatter tops. Gravity data collected by geodetic satellites are one of the primary means of studying the thermal and mechanical perturbations that occur when midplate swells or plateaus form.

For example, the amplitude of the geoid high over midplate swells has been the key observational constraint used to argue that the base of the lithosphere has been reheated by an upwelling mantle plume since the compensation depth is 60 to 70 km (Crough, 1978; McNutt, 1987). The large amplitude (10 m) and long wavelength (~1000 km) of the geoid signature from the thermal anomaly responsible for uplifting swells is more than adequately mapped by the altimetric geoid data presently available over the oceans. However, more precise gravity data with better resolution would contribute to studies of midplate swells in an indirect, but significant way. Gravity anomalies with an accuracy of a few mgals and with wavelength of 30–50 km have been the principal observational constraint used to measure the flexural rigidity or elastic plate thickness of the oceanic lithosphere (Figure 3) supporting the individual hot spot volcanoes capping these swells (Watts, 1978). Because the base of the elastic plate corresponds to an isotherm near 500°–600°C (McNutt and Menard, 1982), by measuring the elastic plate thickness as a function of distance along the subsiding thermal swell as it moves past the hot spot, we can chart the depth to the 500°–600°C isotherm as a function of time (McNutt, 1984). This view of the evolution of one isotherm provides a strong constraint on the details of the thermal structure imposed by the hot spot that cannot be resolved by the more general integral constraints on low density provided by the longer wavelength geoid anomaly over the long swell. A thorough understanding of the mechanism by which the hot spot reheats the lithosphere requires such knowledge of the vertical and lateral structure of reheating.

The more numerous oceanic plateaus have the following characteristics: Lack of focussed seismicity, non-lineated magnetic anomaly patterns, (generally) calcareous sediment caps, crustal thickness in excess of 15 km, and topographically-correlated geoid anomalies (Carlson et al., 1980; Nur and Ben-Avraham, 1982; Sandwell and Renkin, 1987). In total area, oceanic plateaus cover more than 3% of the seafloor. Therefore, they must play a significant role in both the evolution of the ocean basins and the formation of collision-type margins (Vogt, 1973; Ben-Avraham et al., 1981). Nevertheless, the origin and subsurface structure of these features remain enigmatic. Are they continental fragments or oceanic in origin, formed by excess volcanism at or near midocean ridges? The thick pelagic caps make direct sampling by dredging and even drilling difficult. Seismic refraction data are not always available, and in at least one case (Ontong-Java), the same set of travel times have been used to argue for both continental and oceanic origin. Since continental-type plateaus would have deeper isostatic roots than oceanic ones, gravity data provide important information as to the origin of plateaus. Satellite altimeter data have been used to estimate depths of compensation for a number of plateaus from the slope of geoid height versus topography (MacKenzie and Sandwell, 1986). For smaller plateaus, the accuracy of this procedure is limited by the accuracy and coverage of existing satellite altimeter data. A complete gravity/topography study requires a field accurate to 1 mgal at resolution of 50 km and greater.

Distribution of Volcanoes. Although the most prominent islands and seamounts occur as chains (on fast-moving plates) or clusters (on slow-moving plates) formed by hot spots, we suspect that the vast majority of oceanic volcanoes are smaller features erupted in a less organized, but still non-random, manner (Jordan et al., 1983); Smith and Jordan, 1987). Several factors control the distribution of oceanic volcanism. First, there must be a large supply of magma beneath the lithosphere. Second, the lithosphere's density and thermal structure must be such that the magma has enough hydraulic head (Vogt, 1974) and latent heat to penetrate it without freezing during ascent (Spence and Turcotte, 1985). Finally, the lithosphere must remain over the magma pool long enough for the volcano to develop (Gass et al., 1978). Since the rate at which such seamounts are erupted onto the seafloor is related to the thermal state of both the lithosphere and the deeper mantle, seamount distribution contains information on both thermal properties of the plate and the temporal evolution of the convecting mantle.

Almost all oceanic volcanoes lie beneath the ocean surface and thus most remain uncharted. At the present exploration rate, it will take several centuries to map significant portions of the seafloor using ships. It has already been demonstrated that gravity/geoid data can be used to locate uncharted seamounts (Dixon et al., 1983; Dixon and Parke, 1983; Sandwell, 1984), and for those of known dimensions, the age of the lithosphere at the time of emplacement of a seamount can be inferred from the gravitational signature of the flexural response of the lithosphere caused by loading of seamount (Cazenave et al., 1980; Watts and Ribe, 1984). Thus, high quality global coverage would allow a more accurate census of the age distribution of seamounts. Such a data set would place constraints on thermal evolution of the lithosphere and the time variability of mantle convection. The requirements for seamount studies are an accuracy of 1 to 5 mgals and resolution of 10 to 50 km.

CONTINENTAL LITHOSPHERE

The potential for accurate global gravity data to improve our understanding of lithospheric properties and processes is even more apparent for the continents, which are considerably more complex and less readily explained by plate tectonic concepts than the ocean basis. For example, evidence is mounting that the differences in the mechanical properties of continental and oceanic lithosphere are not simply explained by the presence of the thick, granitic continental crust, but rather require thermal and/or compositional differences extending to depths of 200 km or more. At present, gravity data accurate ±4 mgals at 100 km resolution are publicly available for only 22 percent of the Earth's land area (Figure 4), with political and geographical barriers preventing further acquisition by means of standard ground surveys. In order to comprehend the origin, evolution, and resource potential of that part of the planet which we inhabit, global gravity missions are a primary scientific priority.

Rifting and Continental Extension. The global distribution of continents today and the partitioning of their mineral and petroleum resources largely reflects the effects of continental rifting, and yet it is a process about which we understand very little. Does the location of rifting reflect the position of diverging currents in the mantle, a preexisting zone of weakness in the lithosphere, or both? Why do some rifting events fail after a short period of time while others succeed in leading to the formation of new ocean basins? How is the extension partitioned horizontally and vertically in the crust and lithosphere? Gravity data can bring important constraints on the problems concerning continental rifting in several ways.

Gravity anomalies over rifts are sensitive to the perturbed crustal structure from lithospheric stretching and any deep thermal anomalies responsible for doming and plate thinning. An outstanding problem in the study of extensional deformation is the disagreement among various measures of extension, such as heat flow, subsidence, and gravity anomalies, as to the total amount of lithospheric thinning in a single vertical column (Royden et al., 1983a, b; Wernicke, 1985). As compelling evidence for the discontinuous nature of extension in time as well as space (e.g., Wallace, 1984; Glazner and Bartley, 1984; Morgan et al., 1986) has continued to grow, the need for models that go beyond one layer extension has been recognized (e.g., Royden et al., 1983 a,b: Hellinger and Sclater, 1983). With particular reference to the Basin and Range, Wernicke (1985) proposes a simple shear model (Figure 5) for the continental lithosphere in which motion along a low angle detachment allows lithospheric thinning in regions far removed from the surface zone of normal faulting. In general, this geometrical model is capable of explaining thermal uplift on the flanks of rifted regions where no crustal thinning has occurred, although the effects of small-scale convection induced by large lateral thermal gradients have also been invoked to explain the same observations (Steckler, 1985; Moretti and Froidevaux, 1985; Moretti and Chenet, 1987). Gravity anomalies have the potential to distinguish between these two explanations by providing bounds on the vertical and lateral extent of the low-density material providing flank uplift and by mapping out variations in flexural strength of the lithosphere caused by thermal reheating. Broad constraints on thermal structure would be obtained with gravity data accurate to 1 to 2 mgals with a resolution of 100 km. Specific information on variations in flexural rigidity, given the low elastic plate thicknesses to be expected, would require similar accuracy but a resolution of 20 km or better.

Sedimentary Basins and Passive Margins. Sediment deposits in continental basins and on passive continental margins preserve a record of the Earth's geologic history. As repositories for fossil fuels, they represent the most economically significant geologic feature. The principal research topics include: Why do basins and margins subside? Why are the same basins periodically reactivated? Do sediment onlap/offlap patterns on passive margins reflect changes in eustatic sea level or temporal variations in lithospheric rheology? Gravity anomalies bear on these problems in several ways. For example, gravity maps of the Michigan Basin reveal a high-density body at the base of the sedimentary strata thought to correspond to a magmatic intrusion (Haxby et al., 1976). Cooling of this magma body may have supplied the driving force for basin subsidence. Additionally, gravity observations plus data on depths to distinct stratigraphic horizons yield estimates of the elastic thickness of the basin lithosphere as a function of time. The elastic thickness in turn constrains models of the long-term thermal evolution of the basin. Thus, gravity observations supply key information on both the driving forces for basin subsidence and the history of how those forces affect the mechanical behavior of the lithosphere. Global gravity data with ~50 km spatial resolution and accuracy of 1 to 2 mgals is required here. In order to study passive margins, it is particularly vital that we obtain a gravity data set continuous across the coastlines.

Mountain Belts. Gravity observations have already played a major role recently in completely overturning the accepted notion that mountain belts on the Earth's surface are compensated by simple crustal thickening through a form of Airy isostasy. Karner and Watts (1983) noted a consistent assymetry in the Bouguer gravity field across the Alps and Appalachians. The Bouguer gravity low, which results from the low-density material at depth compensating the excess mass of the mountains, is consistently offset towards the foreland basin to the west of the Appalachians and to the north of the Alps, while a prominent gravity high, unassociated with any topographic feature and not predicted by Airy isostasy, appears in the hinterland on the opposite side of the orogens (Figure 6). Karner and Watts (1983) demonstrated that this gravity pattern is consistent with a model in which the mountains are supported by a stiff elastic plate which has underthrust the mountains from the direction of the foreland in the process of continent-continent collision. The amplitude of the deflection of the elastic plate as revealed by the magnitude of the Bouguer gravity low requires loading by both the mountainous topography and by a buried high-density body in the hinterland, the mass presumably responsible for the Bouguer gravity high. This new model for the structure of mountain belts has thus established the validity of elastic flexure to describe the rheology of continental lithosphere and the existence of subsurface loads to maintain the deflection of foreland basins despite erosion of topographic loads.

Despite the importance we now place on buried loads in describing the conditions of mechanical equilibrium at mountain belts, the nature of these buried loads remains obscure. Subsurface loads from cold slabs (Sheffels and McNutt, 1986), dense obducted blocks (Karner and Watts, 1983), and normal stress applied from flow in the mantle (Lyon-Caen and Molnar, 1983) have all been used to supply forces and bending moments to the lithosphere beneath mountain belts. Do all these factors contribute to the compensation of orogenic belts at different times in their geologic evolution, or do the peculiarities of plate collision lead to fundamentally different loading conditions at different locations? We require additional studies of thrust belts at all stages of evolution with a variety of pre-collision tectonic settings (e.g., presence or absence of back-arc basins, different ages of colliding plates, etc.). A wide range exists on Earth; unfortunately, we lack observations of the gravity field over many, particularly the very youngest collision zones, due to difficult terrain and/or political problems with access.

Gravity coverage over continental orogens at wavelengths of 50–100 km (i.e., less than the flexural wavelength of the lithosphere) with an accuracy of 1 to 2 mgals would allow us to test models of lithospheric rheology, mechanisms of plate loading, causes of vertical tectonics in orogens and the details of continental suturing. For example, McNutt and Kogan (1987) used statistics of gravity anomalies in Eastern Europe and Central Asia to argue that steeply plunging continental plates beneath thrust belts are characterized by a low value of elastic plate thickness even for very old lithosphere. They explain their result as the effect of massive brittle and ductile failure of the plates at high strains, as might occur for a plate which behaves according to the rheological model shown in Figure 7. The unavailability of unclassified gravity profiles across the orogens used in their study prevents them from testing their hypothesis with forward modeling.

Deep Structure of the Continental Lithosphere. The thickness of oceanic plates has been determined based on the cooling half-space model. Generally speaking, it varies from almost zero thickness at the midocean ridges to about 100 km thick beneath old oceanic basins. However, the thickness of continental lithosphere has not yet been agreed upon. The results of seismic studies on the thickness of continental lithosphere are controversial, with maximum thicknesses ranging from no more than 200 km (Anderson, 1979) to over 400 km (Jordan, 1979a). The flexural observations from foreland basins adjacent to mountain ranges point to an asymptotic thermal plate thickness for continental lithosphere of the order of 250 km or greater, at least twice that for oceanic lithosphere. The question remains as to how such a cold continental keel can be maintained against convective destabilization. One viable hypothesis for the deep structure of continents proposes a chemically-induced density reduction in the lower continental lithosphere that offsets the density increase from cooling (Jordan, 1979b). Regardless whether the bottom of the lithosphere is defined as a thermal boundary or a chemical boundary, density anomalies will exist as a depth between 100–400 km across the boundary of a “continental root”. This horizontal density variation will give a surface gravity anomaly of about 1–5 mgal. The anticipated wavelength of the gravity anomaly will coincide with the length scale of the continent. Thus, an improved constraint on the thickness of the continental lithosphere can be derived based on improved surface gravity data and proper modeling of mantle thermal structure adjacent to the roots of continents. Earth scientists do not have, at present, a precise global gravity field to search for the gravity signal from deep continental thermal structure.

THE MANTLE

The problem of mantle convection is fundamental to understanding the evolution of the Earth. The outgassing of the oceans and atmosphere, the differentiation of the crust, volcanism, and all

tectonics--continental as well as oceanic--are ultimately dependent on energy sources within the mantle and core, and upon the transport of this energy and material by flow driven by thermal or compositional buoyancy. The oceanic crust and lithosphere are part of this convecting system: they make up its uppermost, cold thermal boundary layer. Their motion is associated with a flow pattern coming to the surface at the midocean ridges or other rifting areas and returning to the interior at subduction zones. Most oceanic crust, as well as its associated lithosphere, is recycled to the interior. The continental lithosphere, which consists of the continental crust and

sizeable pieces of subcontinental mantle, rides on top of the convective system. The velocities of the system are of the order of centimeters per year; the heat transport is an average of 0.08 W

m^{-2}. These values, together with the thermal and rheological properties of rocks, indicate that the system must be more complicated than the smoothest flow necessary for the observed plate motions. Phenomena such as changes in the plate tectonic pattern on time scales of tens of millions of years, long-term episodicity of volcanism in tectonically complex areas such as western North America, exceptionally high heat flow on the continental side of subduction zones, and higher than predicted heat flow and topography in some parts of ocean basins, all suggest that there are secondary scales of mantle flow not directly connected to the precisely measured plate tectonic pattern. Observations which see through the lithosphere and into the mantle are needed. The gravitational field provides one such observable.

Viscosity of the Mantle. The thermal state and mechanical behavior of the mantle are strongly dependent on its viscosity structure. There are two major issues to be resolved. From laboratory creep studies of minerals under mantle conditions (i.e., high temperature and high pressure), it is inferred that mantle rock should deform according to a power law non-Newtonian rheology (Kohlstedt and Hornack, 1981). However, a Newtonian model for the mantle has been utilized to delineate mantle viscosity structure: such a model can approximately fit both the isostatic glacial rebound data and the gravity data around Hudson Bay in Canada (Peltier and Wu, 1982). This result can perhaps be reconciled with the laboratory data if volatiles and inhomogeneities within the mantle alter its deformation mechanism such that it resembles a Newtonian fluid. The second issue concerns the magnitude of mantle viscosity. Some studies find that the viscosity of both the upper and lower mantle is about 10^{21} Pa-sec (Wu and Peltier, 1983), while other studies, such as the geoid plus seismic tomography study described in Figure 8, indicate that the viscosity of the upper mantle is 10^{19} Pa-sec with the viscosity of the lower mantle approximately one order of magnitude higher (Hager and Clayton, 1987). Studies utilizing non-Newtonian rheology indicate that the viscosity of the lower mantle is probably not constant but changes with a two- to three-orders-of-magnitude variation across the lower mantle (Karato, 1981). Geoid highs over subducted slabs can also be explained using a non-Newtonian model (McAdoo, 1982). Both issues can be addressed and plausible mantle deformation mechanisms can be discriminated using a set of gravity measurements that are dense and high resolution, e.g., 1 mgal accuracy over 500 km wavelengths, together with seismic tomographic studies and convection model computations (Hager and Clayton, 1987).

Vertical Scale of Mantle Convection. The vertical scale of mantle flow is a subject of current debate. Geochemical isotopic studies have been interpreted as suggesting the existence of a multilayer structure (Jacobsen and Wasserburg, 1981). However, geophysical arguments indicate that a single layer convective regime is more likely (Spohn and Schubert, 1982). If multilayer convection exists, it is hypothesized that the 670 km seismic discontinuity will be the boundary between separate flow systems in the upper and lower mantle. Due to the upwelling and down-going currents associated with mantle flows, undulations or vertical displacements at this boundary will occur with a wide range of wavelengths (Christensen and Yuen, 1984). Due to the attenuating effects of distance, the ones with most signal will be in the range of several thousand km (Busse, 1981). The gravitational characteristics of a chemically stratified mantle are quite different than those of a mantle with uniform composition. Consequently, high resolution gravity data can be used to better delineate the competing hypotheses.

The depth of slab penetration can be studied particularly effectively with better gravity data. Based on seismic investigations of travel time (Creager and Jordan, 1984, 1986), subducting slabs are thought to be able to penetrate into the lower mantle. Since seismic anomalies are directly related to density variations, the existence of deep penetrating slabs can be examined with the gravity data derived from the proposed global measurements. The gravitational signature of a deep subducted slab is particularly sensitive to the presence or absence of a chemical discontinuity at 670 km depth. If a discontinuity is present, dynamic compensation of the slab will occur at that depth, resulting in smaller gravity anomalies for a given density contrast. The current long wavelength gravity field (>4000 km) can be satisfied by either a model with normal slab density and mantle-wide flow or a model with high slab densities (caused by phase changes) and a chemical barrier to flow. The models can be discriminated using shorter wavelengths. Since subducting slabs lie beneath island arcs and typically span the ocean-continent transition, altimetric geoids are not sufficient to study this problem and gravity fields such as those obtained by a gravity-measuring satellite are required. The expected signal strength will be above 0.1 mgal with a length scale dictated by the angle of a subducting slab and the speed of slab subduction; a typical wavelength will be about 1000 km.

Small-scale Convection. The existence of small-scale convection (i.e., at horizontal scales much less than plate dimensions) beneath lithospheric plates has been predicted from the observed flattening of the depth-age and geoid-age relations for oceanic lithosphere, from laboratory experiments, and from theoretical convection calculations. For example, the fact that the slope of the geoid begins to flatten over older oceanic lithosphere (Parsons and Richter, 1981) has been one of they key observations in support of the thermal plate model (McKenzie, 1967) which does not allow conductive cooling to extend below about 125 km depth. Presumably, deeper cooling could be prevented by some sort of small-scale convection, such as the longitudinal rolls observed in a layer of constant viscosity undergoing horizontal shear in the laboratory by Richter and Parsons (1975). Theoretical calculations suggest that longitudinal rolls can exist only if the upper mantle viscosity is extremely low (Yuen et al., 1981) and that they may have a typical horizontal wavelength of about 150 km with an amplitude of 5 mgal (Buck, 1985).

One of the major discoveries of the SEASAT altimeter mission is gravity undulations with the predicted wavelength, amplitude, and orientation in the Central Pacific (Haxby and Weissel, 1986). However, in the Indian Ocean, crossgrain features with the same wavelength but even larger amplitudes (20–60 mgal) are thought to be due to buckling of the lithosphere in response to N-S compression of the Indian plate as it collides with Asia (Weissel et al., 1980; McAdoo and Sandwell, 1985). Even lithospheric boudinage, a pinch and swell instability resulting from platewide tensile stresses (Froidevaux, 1986; Zuber et al., 1986), might be capable of producing some of the crossgrain lineations. Therefore, before crossgrain lineations are used to constrain models of small-scale convection, we must map them with greater accuracy and in more detail to determine their origin. Do these features contain information concerning asthenosphere viscosity, or are they indicative of lithospheric stress and rheology? An improved gravity data set not only will be able to verify the existence or absence of such structures, it will also be able to delineate where such rolls begin and where they terminate as a function of plate age and spreading velocity. An accuracy of 1 mgal at 100 km resolution would allow these features to be traced to 20 percent of their amplitude.

Beneath the continents there is direct observational evidence from seismic tomography that small-scale convection also occurs. For example, below the Transverse Ranges in Southern California a curtain of high-velocity material extending down to a depth of 250 km is evidence of convective downwelling of the cold thermal boundary layer at the base of the lithosphere (Humphreys et al., 1984). This feature may explain the dynamics of the Big Bend of the San Andreas Fault. The gravity signature of this feature is calculated to be up to 15 mgal in amplitude with a wavelength of about 150 km. A model of the local gravity field (Sheffels and McNutt, 1986) indicates that this feature is, to a large extent, compensated from above by flexure of the overlying plates. In this particular instance, both gravity observations and velocity anomaly maps from seismic tomography were necessary in order to understand the interaction of the lithosphere with the asthenosphere. Once we have this system response well calibrated, it may be possible to identify regions of downwelling beneath continents from gravity and the surface geology alone.

Mantle Plumes. While the volcanoes associated with mantle plumes are lithospheric features, the source and ultimate cause of hot spot activity lies in the mantle below. Candidates for the formation and feeding of hot spots include chimney-like thermal plumes (Morgan, 1972), isolated hot blobs (Olson and Nam, 1986), and stripping away of the base of the lithosphere by convective instability (Richards et al., 1987). For each of these models, it is possible to predict a dynamic gravity model which could then be tested by observation. The depths of origin of hot spots can be addressed by looking at the long wavelength gravity variations. In order to separate the dynamic topography due to deep circulation from crustal and lithosphere effects, understanding the shorter wavelength variations is essential. To do this would require gravity data with an accuracy of 1 mgal at spatial resolution on the order of 100 km.

TEMPORAL VARIATIONS IN GRAVITY

A less obvious but equally important use of gravity data is the detection of time variations in the field. In most geophysical and geodetic studies, gravity is assumed to be static. Such an assumption is usually justifiable because expected changes in the field over typical observing lifetimes (~10^2 years) are relatively small. However, the precision of proposed satellite gravity measurements is such that heretofore undetected changes in gravity even over six months will be sensed. Furthermore, subsequent missions of similar design could provide additional time samplings of the field, thereby better resolving ongoing variations. Effects causing changes in gravity include ocean and solid Earth tides, postglacial rebound, secular melting of the ice caps, seasonal variations in ground water (e.g., snowpack) and great earthquakes. The magnitudes of these effects have been estimated in a recent study by Wagner and McAdoo (1986).

The characteristic periods for these effects range from minutes to thousands of years but to a large extent are global in nature, with the exception, perhaps, of the effects of great earthquakes. The cause of these changes in gravity are the orbital motion of the Sun and Moon and changes in the physical state of the solid Earth, oceans, and atmosphere. Some effects that should give rise to temporal variations in gravity detectable with an extremely sensitive gravity gradiometer, for example, are given below. It is difficult to anticipate all such effects, and thus it is particularly in the frontier area of temporal variations in gravity that completely unexpected discoveries could result.

Tides. The tides span a wide range of frequencies between about 12 hours and 19 years, and are related to the periods associated with the orbital motions of the Earth about the Sun, the Moon about the Earth and the rotational period of the Earth on its axis. A spacecraft in orbit about the Earth will sense all these changes in gravity (Figure 7) and also those induced by its own motion over the tidally distorted Earth primarily a once per orbital revolution effect. Because the frequencies of the tides are well known, they are separable from the signals of other sources of gravity anomalies, and provide new information of significant value to solid Earth and ocean dynamics.

Tidal variations in gravity can cause very large perturbations of a spacecraft orbit and the measurement of gravity or its gradient at altitude will provide considerable information about the tides. This technique has already been used (Williamson and Marsh, 1985) to estimate the long wavelength components of at least 12 tides. Generally, these tidal solutions provide new information on the ocean tides on the assumption that good models already exist for the solid Earth. In turn, these ocean tidal solutions then can be used to improve our knowledge of Earth-Moon interactions, including the change in the Moon's orbit and and the deceleration of the Earth on its axis due to the transfer of angular momentum from the Earth to the Moon (Williamson and Marsh, 1985).

Postglacial Rebound. The most recent episode of Pleistocene deglaciation began roughly 18,000 years ago. The massive continental sheets of ice were largely melted 5000 years ago (Wu and Peltier, 1983). Although the melting has nearly ceased, the solid Earth continues to rebound in response to this deglaciation. This rebound continues today because of the high average viscosity (order 10^{22} Pa s) of the Earth's mantle. Concomitant changes in the gravity field also continue today.

Figure 9 shows the geoid effect predicted from the present day rebound rate. When the rebound is decomposed in spherical harmonics, the largest components occur at very low degree and order (≤ 6). The maximum gravity change per year is 10^{-3} mgal. Such small variations in gravity associated with the second spherical harmonic have been measured as perturbations in the orbit of LAGEOS (Yoder et al., 1983) and been used to estimate the viscosity of the lower mantle. Higher harmonics of glacial rebound have not been observed from satellites although in several areas they are well constrained by shoreline emergence data (Wu and Peltier, 1983). Rebound of these higher harmonics contains information on radial variations in the viscosity of the upper mantle and asthenosphere, important for dynamic models of the Earth.

To understand the full extent and mechanism of the rebound, we require global models of the gravity field at wavelengths of 3000 km or longer and an accuracy of 10^{-3} mgals repeatedly measured at various epochs over a period of about a decade. Although changes in all the medium to long wavelength harmonics occur, those most easily detectableare restricted to the very low degree terms. There is reasonable expectation that variations up to degree and order four will be observable from low Earth orbit. Detailed knowledge of rebound will help solve the problem of the lateral and vertical variation of mantle viscosity.

Glacial Melting and Atmospheric Warming. One of the most important variations in gravity is due to the slow secular melting of glaciers (Meier, 1984). Climate studies show an accelerated warming of the atmosphere over the last century that is believed to be caused by increases in atmospheric carbon dioxide associated with burning of fossil fuels and deforestation (NASA Advisory Council, Earth Systems Science Committee, 1987). Global warming may induce glacial melting resulting in sea level increases and coastal flooding.

Accurate satellite gradiometer missions could monitor the small changes in the global gravity field caused by the redistribution of mass from the glaciers to the oceans. The current estimate of yearly sea level rise, caused by melting of small glaciers, is 5×10^{-4} m. However, this number is very uncertain because only a few glaciers have been monitored over the last century (Meier, 1984). The yearly change in gravity over the oceans due to glacial melting is only 2×10^{-5} mgal. In glaciated areas, the yearly gravity change may be somewhat higher because the area is smaller.

Although these predicted temporal variations in gravity are extremely small, they are global and could perhaps be measured by a satellite gravity gradiometer which can render the highest resolution at long wavelengths by virtue of its long integration time. One method of increasing the signal associated with secular glacial melting is to measure over a longer period of time. This could be done with two or more missions spaced at five-year intervals.

Seasonal Variations in Groundwater and Excitation of the Chandler Wobble. The Chandler wobble is an apparent 427-day precession of the Earth's instantaneous pole of rotation about its axis of greatest moment of inertia. The Earth's anelasticity damps the Chandler wobble, such that in the absence of some excitation, the rotation pole and axis of figure would eventually coincide. The rate at which the wobble is damped depends upon the largely unknown viscosity structure of the mantle and core. Its Q (i. e., π times damping time divided by wobble period) could lie anywhere between 70 and 600. If the wobble is reexcited only occasionally, such as, for example, by mass movements associated with very great earthquakes, then its Q must be very large. If the forcing is frequent, then Q must be small. A knowledge of the mechanism which excites the Chandler wobble would constrain its Q, which in turn would reveal the viscoelastic structure of the lower mantle and core.

Winter in the Northern Hemisphere is accompanied by a dramatic increase in continental water storage associated with the development of ice deposits and snowpacks (Hinnov and Wilson, 1987). Hinnov and Wilson (1987) and Wahr (1982) have suggested that this seasonal fluctuation in groundwater storage excites the Chandler wobble, implying a small Q. To test whether groundwater fluctuations excite the Chandler wobble and to determine its Q, better observations are needed. Due to a lack of hydrologic and meteorologic data, especially in Asia, the spatial and temporal variations in groundwater are not accurately determined. Typical annual variations in surface groundwater height over the continents are 0.1 m. During the seasons this water is transferred from the oceans to the continents and back to the oceans. Gravity variations associated with this mass transfer will be about 4×10^{-3} mgal. To determine the transfer function between the excitation and the observed wobble, a gravity mission should last several years. A shorter mission (about 6 months) could be used to confirm that groundwater plays a major role in exciting the Chandler wobble.

Volcanoes and Earthquakes. Gravity data have a key role in monitoring precursors to volcanic eruptions and in determining the response of the Earth to faulting. For example, Rundle (1982) has calculated the gravity changes associated with magma loading. Observatories at Hawaii (Dzurisin et al., 1980) and Mt. Saint Helens (Jachens et al., 1981) have detected these gravitational changes caused by magma inflation events, which signal impending eruptions. Walsh and Rice (1979) and Savage (1984) have modeled the change in surface gravity associated with dipslip faulting. For an infinitely long thrust fault they find the slipinduced changes in vertical gravity are proportional to changes in elevation. Upward continuation of the gravity signal from an earthquake to spacecraft altitude has been carried out by Wagner and McAdoo (1986). For the 1964 Alaska earthquake they show that a spacecraft passing over the region at 160 km altitude would have its velocity changed by approximately 15×10^{-6} m s^{-2} in a period of about 50 seconds as a result of the coseismic vertical motion, indicating a change of about 0.1 to 1 mgal over an area of about 400 km^2. Changes of this magnitude, if detectable from near Earth orbit, might lead to the study of pre-and post-seismic behavior on a global scale.

IMPLEMENTATION STRATEGY

The previous sections of this report have identified a large number of fundamental problems across a broad range of disciplines in Earth Science. The requirements for gravity information in order to address them, as summarized in Figure 10, are as diverse as the problems themselves, although we can identify a general need for shorter wavelength information of standard accuracies (a few mgals) and much more precise gravity estimates at long wavelengths. Clearly, no one gravity mission could possibly satisfy all of these requirements, especially if we consider the exciting possibility of monitoring changes in gravity over year to decade time periods.

Many of the problems in the marine environment, such as ridge crest processes, fracture zone structure, origin of midplate swells and plateaus, seamount distribution, and some aspects of small-scale convection, require accuracy of only a few mgals but resolution of 50 km or better. It is extremely difficult due to upward attenuation to attain such short wavelength information from any space mission that attempts to measure the geopotential directly. Another altimeter mission with more accuracy and better ground coverage than SEASAT is probably the best way to meet these objectives. The GEOSAT mission has already obtained denser track coverage than SEASAT, but most of the data remain classified. The ERS-1 altimeter mission, to be flown by ESA in the early 1990s, will carry a more accurate altimeter. Data from these missions will be a valuable resource for the Earth science community.

Despite the anticipated higher accuracy and better ground coverage from future altimeter missions, most of the disadvantages of using an altimeter to infer gravity cannot be overcome. The accuracy and resolution of the geoid are ultimately limited to a few 100s of mm and 30 km, respectively, by unmodelled oceanographic effects such as tides, waves, and currents. The data cannot be used to address any problems which cross onto the continents, such as those involving subduction zones. Finally, solution of many marine problems requires that the bathymetry as well as the gravity be known. Therefore, in many poorly surveyed areas oceanographic expeditions to acquire marine bathymetry and gravity simultaneously are the only alternative.

Almost all of the outstanding problems we identify in continental tectonics and structure and mantle convection can best be addressed with a low altitude, highly accurate space mission that measures the global geopotential directly. This strategy is the only feasible way to obtain a truly global gravity field without regard to political boundaries and shorelines. Such a mission would also provide complementary data of better accuracy but perhaps poorer resolution to what can be obtained by and altimeter over the oceans for addressing marine problems. The most cost-effective way to measure the geopotential from space with high accuracy is via gravity gradiometry. Several instruments have already been proposed for space deployment.

The acquisition of gravity data with a resolution better than 100 km probably cannot be achieved from space. Large areal surveys using portable land gravimeters are time-consuming and often politically impractical. Gravity measurements from fixed-wing aircraft and helicopters are possible, but acquisition costs are high. More recently, U.S. Navy and Air Force programs have led to the production of ship and airborne gravity gradiometers which allow substantially large gravity surveys within a reasonable time. A global potential map is outside the range of even airborne surveys, but for some of the problems described in the oceanic and continental lithosphere sections requiring high resolution data in a limited area, this strategy might be feasible.

ACKNOWLEDGEMENT

This report has been prepared from contributions to a NASA-sponsored workshop held during February 1987, in Colorado Springs. Participants were: John Apel, David McAdoo, Lindrith Cordell, Marcia McNutt, Charles Finley, Ho Jung Paik, Thomas Fischetti, Richard Rapp, Edward Flinn, David Sandwell, Bradford Hagar, David Smith, Albert Hsui, David Sonnabend, Andy Lazarewicz, and Louis Walter. Major contributions by Bradford Hagar, David McAdoo, and David Sandwell are acknowledged.

REFERENCES

Anderson, D. L., The Deep Structure of Continents, J. Geophys. Res. 84, 7555–7560,, 1979.

Ben-Avraham, Z., A. Nur, D. Jones, and A. Cox, Continental Accretion and Orogeny: From Oceanic Plateaus to Allochthonmous Terrain, Science, 213, 47–54, 1981.

Buck, W. R., When Does Small Scale Convection Begin Beneath Oceanic Lithosphere?, Nature, 313, 775–777, 1985.

Busse, F. H., On the Aspect Ratios of Two-Layer Mantle Convection, Phys. Earth Planet. Int., 24, 320–324, 1981.

Carlson, R., N. Christensen, and R. Moore, Anomalous Crustal Structures in Ocean Basins: Continental Fragments and Oceanic Plateaus, Earth Planet. Sci. Lett., 51, 171–180, 1980.

Cazenave, A., Two Distinct Trends for Cooling of the Oceanic Lithosphere, Nature, 310, 401–403, 1984.

Cazenave, A., B. Lago, K. Dominh, and K. Lambeck, On the Response of the Ocean Lithosphere to Seamount Loads from GEOS-3 Satellite Radar Altimetry Observations, Geophys. J. Roy. Astron. Soc., 63, 233–252, 1980.

Christensen, U. R., and D. A. Yuen, The Interaction of a Subducting Lithospheric Slab With a Chemical or Phase Boundary, J. Geophys. Res., 89, 4389–4402, 1984.

Craig, C. H. and D. P. McKenzie, The Existence of a Thin Low-Viscosity Layer Beneath the Lithosphere, Earth Planet. Sci. Lett., 78, 420–426, 1986.

Creager, K. C., and T. H. Jordan, Slab Penetration Into the Lower Mantle, J. Geophys. Res., 89, 3031–3049, 1984.

Creager, K. C., and T. H. Jordan, Slab Penetration Into the Lower Mantle Beneath the Mariana and Other Island Arcs of the Northwest Pacific, J. Geophys. Res., 91, 3573–3589, 1986.

Crough, T., Thermal Origin for Mid-Plate Hot-Spot Swells, Geophys. J. Roy. Astron. Soc., 55, 451–459, 1978.

Detrick, R. S. and G. M. Purdy, The Crustal Structure of the Kane Fracture Zone From Seismic Refraction Studies, J. Geophys. Res., 85, 3759–3777, 1980.

Dixon, T. H., M. Naraghi, M. K. McNutt, and S. M. Smith, Bathymetric Prediction From SEASAT Altimeter Data, J. Geophys. Res., 88, 1563–1571, 1983.

Dixon, T. H. and M. E. Parke, Bathymetry Estimates in the Southern Oceans From SEASAT Altimetry, Nature, 304, 406–411, 1983.

Driscoll, M. L. and B. Parsons, Cooling of the Oceanic Lithosphere: Evidence From Geoid Anomalies Across the Udintsev and Eltanin Fracture Zone System, Earth Planet. Sci. Lett., 88, 289–307, 1988.

Dzurisin, D., L. A. Anderson, G. P. Eaton, R. Y. Koyanagi, P. W. Lipman, J. P. Lockwood, R. T. Okamura, G. S. Puniwai, M. K. Sako, and K. E. Yamashita, Geophysical Observations of Kilauea Volcano, Hawaii, 2, Constraints on the Magma Supply During November 1975–September 1977, J. Volcan. Geotherm. Res. 7, 241–269, 1980.

Fox, P. J., E. Schreiber, H. Rowlett, and K. McCamy, The Geology of the Oceanographer Fracture Zone: A Model for Fracture Zones, J. Geophys. Res., 81, 4117–4128, 1976.

Freedman, A. P., Marine Geophysical Applications of SEASAT Altimetry and the Lithospheric Structure of the South Atlantic Ocean, Ph.D. Thesis, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 1987.

Froidevaux, C., Basin and Range Large Scale Tectonic: Constraints From Gravity and Reflection Seismology, J. Geophys. Res., 91, 3625–3632, 1986.

Gass, I. G., D. S. Chapman, H. N. Pollack, and R. S. Thorpe, Geological and Geophysical Parameters of Mid-Plate Volcanism, Phil. Trans. R. Soc. Lond. Ser. A., 288, 581–597, 1978.

Glazner, A. F., and J. M. Bartley, Timing and Tectonic Setting of Tertiary Low-Angle Normal Faulting and Associated Magmatism in the Southwestern United States, Tectonics, 3, 385–396, 1984.

Hager, B. H., Subducted Slabs and the Geoid: Constraints on Mantle Rheology and Flow, J. Geophys. Res., 89, 6003–6015, 1984.

Hager, B. H., and R. W. Clayton, Constraints on the Structure of Mantle Convection Using Seismic Observations, Flow Models, and the Geoid in Mantle Convection, W. R. Peltier, Ed., in press, 1987.

Haxby, W. F., D. L. Turcotte, and J. M. Bird, Thermal and Mechanical Evolution of the Michigan Basin, Tectonophysics, 36, 57–75, 1976.

Haxby, W. F., and J. K. Weissel, Evidence for Small-scale Mantle Convection From SEASAT Altimeter Data, J. Geophys. Res., 91, 3507–3520, 1986.

Hellinger, S., and J. G. Sclater, Some Comment on Two-layer Extensional Models for the Evolution of Sedimentary Basins, J. Geophys. Res., 88, 8251–8270, 1983.

Hinnov, L. A., and C. R. Wilson, An Estimation of the Water Storage Contribution to the Excitation of Polar Motion, Geophys. J. Roy. Astron. Soc, 88, 437–459, 1987.

Humphreys, E., R. W. Clayton, and B. H. Hager, A Tomographic Image of Mantle Structure Beneath Southern California, Geophys. Res. Lett., 11, 625–627, 1984.

Isacks, B., and M. Barazangi, Geometry of Benioff Zones: Lateral Segmentation and Downwards Bending of the Subducted Lithosphere, in Island Arcs, Deep Sea Trenches, and Back-Arc Basins, Maurice Ewing Ser., Vol. 1, edited by M. Talwani and W. C. Pitman III, pp. 99–114, AGU, Washington, D. C., 1977.

Jachens, R. C., D. R. Spydell, G. S. Pitts, D. Dzurisin, and C. W. Roberts, Temporal Gravity Variations at Mount St. Helens, March–May, 1980, in the 1980 Eruptions of Mount St. Helens, Washington, U.S. Geological Survey Professional Paper #1250, pp. 175–181, 1981.

Jacobsen, S. B., and G. J. Wasserburg, Transport Models for Crust and Mantle Evolution, Tectonophysics, 75, 163–179, 1981.

Jordan, T. H., The Deep Structure of Continents, Sci. Amer., 240, 92–107, 1979a.

Jordan. T. H., Mineralogies, Densities, and Seismic Velocities of Garnet Iherzolites and Their Geophysical Implications, in the Mantle Sample: Inclusion of Kimberlites and Other Volcanics, edited by F. R. Boyd and H. O. A. Meyer, AGU, Washington, D. C. 1979b.

Jordan, T. H., H. W. Menard, and D. K. Smith, Density and Size Distribution of Seamounts in the Eastern Pacific Inferred From Widebeam Sounding Data, J. Geophys. Res., 88, 10508–10518, 1983.

Karato, S., Rheology of the Lower Mantle, Phys. Earth. Planet. Int., 24, 1–14, 1981.

Karner, G. D., and A. B. Watts, Gravity Anomalies and Flexure of the Lithosphere at Mountain Ranges, J. Geophys. Res., 88, 10449–10477, 1983.

Kohlstedt. D. T., and P. Hornack, Effect of Oxygen Partial Pressure on the Creep of Olivine, in Anelasticity in the Earth, Ed., F. D. Stacey, M. S. Paterson, and A. Nicolas, Geodynamics Ser., 4, 101–107, 1981.

Lyon-Caen, H., and P. Molnar, Constraints on the Structure of the Himalaya From an Analysis of Gravity and Flexural Model of the Lithosphere, J. Geophys. Res., 88, 8171–8191, 1983.

Macdonald, K. C., J.-C. Sempere, P. J. Fox, and R. Tyce, TectonicEvolution of Ridge-Axis Discontinuities by the Meeting, Linking, or Self-Decapitation of Neighboring Ridge Segments , Geology, 15, 993–997, 1987.

MacKenzie, K., and D. Sandwell, Geoid Height Versus Topography for Oceanic Swells and Plateaus, EOS, Trans. Amer. Geophys. Union, 67, 1229, 1986.

Madsen, J. A., D. W. Forsyth, and R. S. Detrick, A New Isostatic Model for the East Pacific Rise Crest, J. Geophys. Res., 89, 9997–10016, 1984.

McAdoo, D. C., On the Compensation of Geoid Anomalies Due to Subducting Slabs, J. Geophys. Res., 87, 8684–8692, 1982.

McAdoo, D. C., and C. F. Martin, SEASAT Observations of Lithospheric Flexure Seaward of Trenches, J. Geophys. Res., 89, 3201–3210, 1984.

McAdoo, D. C., and D. T. Sandwell, Folding of the Oceanic Lithosphere, J. Geophys. Res., 90, 8563–8569, 1985.

McKenzie, D. P., Some Remarks on Heat Flow and Gravity Anomalies, J. Geophys. Res., 72, 6261–6273, 1967.

McNutt, M. K., Lithospheric Flexure and Thermal Anomalies, J. Geophys. Res., 89, 11180–11194, 1984.

McNutt, M. K., Temperature Beneath Midplate Swells: The Inverse Problem, in Seamounts, Islands, and Atolls, Geophys. Monograph #43, edited by B. Keating, R. Batiza, and G. W. Boehlert, AGU, Washington, D. C., 1987.

McNutt, M. K., K. M. Fischer, S. Kruse and J. Natland, The Origin of the Marquesas Fracture Zone Ridge and Its Implications for the Nature of Hot Spots, Earth Planet. Sci. Lett., 91, 381–393, 1989.

McNutt, M. K., and M. G. Kogan, Isostasy in the USSR, II. Interpretation of Admittance Data, in the Composition, Structure, and Dynamics of the Lithosphere-Asthenosphere System, Geodyn. Ser., #16, edited by K. Fuchs and C. Froidevaux, AGU, Washington, D. C., 1987.

McNutt, M. K., and H. W. Menard, Constraints on Yield Strength of the Oceanic Lithosphere Derived From Observations of Flexure, Geophys. J. Roy. Astron. Soc., 71, 363–394, 1982.

McNutt, M. K., and L. Shure, Estimating the Compensation Depth of the Hawaiian Swell With Linear Filter, J. Geophys. Res., 91, 13915–13923, 1986.

Meier, M. F., Contributions of Small Glaciers to Global Sea Level Change, Science, 226, 1418–1421, 1984.
Moretti, I., and P. Y. Chenet, The Evolution of the Suez Rift: A Combination of Stretching and Secondary Convection, Tectonophysics, 133, 229–234, 1987.
Moretti, I., and C. Froidevaux, Physical Models of Extensional Tectonics, in Modeling the Thermal Evolution of Sedimentary Basins, J. Burrus, ed., Ed Technip, Paris, 107–123, 1985.
Morgan, P., Deep Mantle Convection: Plumes and Plate Motions, Amer. Assoc. Petrol. Geol. Bull., 56, 203–213, 1972.
Morgan, P., W. R. Seager, and M. P. Golombek, Cenozoic Thermal, Mechanical, and Tectonic Evolution of the Rio Grande Rift, J. Geophys. Res., 91, 6263–6276, 1986.
National Aeronautics and Space Administration Advisory Council, Earth Systems Science Committee: Earth Systems Science: A Closer View, Washington, D. C., November 1987.
Nur, A., and Z. Ben-Avraham, Oceanic Plateaus, the Fragmentation of Continents, and Mountain Building, J. Geophys. Res., 87, 3644–3661, 1982.
Olson, P., and I. S. Nam, The Formation of Sea Floor Swells by Mantle Plumes, J. Geophys. Res., 91, 7181–7191, 1986.
Parmentier, E. M., and W. F. Haxby, Thermal Stresses in the Oceanic Lithosphere: Evidence From Geoid Anomalies at Fracture Zones, J. Geophys. Res., 91, 7193–7204, 1986.
Parsons. B., and F. Richter, Mantle Convection and the Oceanic Lithosphere, Sea, 7, 73–117, 1981.
Parsons, B., and J. G. Sclater, An Analysis of the Variation of Ocean Floor Bathymetry and Heat Flow With Age, J. Geophys. Res., 82, 803–827, 1977.
Peltier, W. R., and P. Wu, Mantle Phase Transition and the Free Air Gravity Anomalies Over Fennoscandia and Laurentia, Geophys. Res. Lett., 9, 731–734, 1982.
Rapp, R. H., and J. Y. Cruz, Spherical Harmonic Expansions of the Earth's Gravitational Potential to Degree 360 Using 30' Mean Anomalies, Report #376, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio, 1986.
Richards, M. A., B. H. Hager, and N. H. Sleep, Dynamically Supported Geoid Highs Over Hotspots: Observations and Theory, J. Geophys. Res., 7, 7690–7708, 1988.

Richter, F. M., and B. Parsons, On the Interaction of Two Scales of Convection in the Mantle, J. Geophys. Res., 80, 2529–2541, 1975.

Royden, L., F. Horvath, A. Nagymarosy, and L. Stegena, Evolution of the Pannonian Basin System, 2, Subsidence and Thermal History, Tectonics, 2, 91–137, 1983a.

Royden, L., F. Horvath, and J. Rumpler, Evolution of the Pannonian Basin System, 1, Tectonics, Tectonics 2, 63–90, 1983b.

Rundle, J. B., Deformation, Gravity, and Potential Changes Due to Volcanic Loading of the Crust, J. Geophys. Res., 87, 10, 729–10, 744, 1982.

Sandwell, D., A Detailed View of the South Pacific Geoid From Satellite Altimetry, J. Geophys. Res., 89, 1089–1104, 1984.

Sandwell, D. T., and M. L. Renkin, Compensation of Swells and Plateaus in the Northern Pacific: No Direct Evidence for Mantle Convection, J. Geophys. Res., 4, 2775–2783, 1988.

Sandwell, D., and G. Schubert, Lithospheric Flexure at Fracture Zones, J. Geophys. Res., 87, 4657–4667, 1982.

Savage, J. C., Local Gravity Anomalies Produced by Dislocation Sources, J. Geophys. Res., 89, 1945–1952, 1984.

Sheffels, B., and M. McNutt, Role of Subsurface Loads and Regional Compensation in the Isostatic Balance of the Transverse Ranges, California: Evidence for Intracontinental Subduction, J. Geophys. Res., 91, 6419–6431, 1986.

Smith, D. K., and T. H. Jordan, Seamount Statistics in the Pacific Ocean, J. Geophys. Res., 4, 2899–2918, 1988.

Spence, D. A., and D. L. Turcotte, Magma-Driven Propagating Cracks, J. Geophys. Res., 90, 575–580, 1985.

Spohn, T, and G. Schubert, Modes of Mantle Convection and the Removal of Heat From the Earth's Interior, J. Geophys. Res., 87, 4682–4696, 1982.

Steckler, M. S., Uplift and Extension at the Gulf of Suez-Indications of Induced Mantle Convection, Nature, 317, 135–139, 1985.

Toksoz, N. N., J. W. Minear, and B. R. Julian, Temperature Field and Geophysical Effects of a Downgoing Slab, J. Geophys. Res., 76, 1113–1138, 1971.

Vogt, P., Subduction and Aseismic Ridges, Nature, 241, 189–191, 1973.

Vogt, P., Volcano Height and Plate Thickness, Earth Planet. Sci. Lett., 23, 337–348, 1974.

Wagner, C. A., and D. C. McAdoo, Time Variations in the Earth's Gravity Field Detectable With Geopotential Research Mission Intersatellite Tracking, J. Geophys. Res., 91, 8373–8386, 1986.

Wahr, J. M., Effects of the Atmosphere and Oceans on the Earth's Wobble, Geophys. J. Roy. Astro. Soc., 70, 349–372, 1982.

Wallace, R. E., Patterns and Timing of Late Quaternary Faulting in the Great Basin Province and Relation to Some Regional Tectonic Features, J. Geophys. Res., 89, 5763–5770, 1984.

Walsh, J. R., and J. R. Rice, Local Changes in Gravity Resulting From Deformation, J. Geophys. Res., 84, 165–170, 1979.

Watts, A. B., An Analysis of Isostasy in the World's Oceans, 1, Hawaiian-Emperor Seamount Chain, J. Geophys. Res., 83, 5989–6004, 1978.

Watts, A. B., and N. M. Ribe, On Geoid Heights and Flexure of the Lithosphere at Seamounts, J. Geophys. Res., 89, 11152–11170, 1984.

Watts, A. B., and M. Talwani, Gravity Anomalies Seaward of Deep-Sea Trenches and Their Tectonic Implications, Geophys. J. R. Astron. Soc., 36, 57–90, 1974.

Weissel, J. K., R. N. Anderson, and C. A. Geller, Deformation of the Indo-Australian Plate, Nature, 287, 284–291, 1980.

Wernicke, B., Uniform-Sense Normal Simple Shear of the Continental Lithosphere, Canada, J. Earth Sci., 22, 108–125, 1985.

Williamson, R. G., and J. G. Marsh, Starlette Geodynamics: The Earth's Tidal Response, J. Geophys. Res., 90, 9346–9352, 1985.

Wu., P., and W. R. Peltier, Glacial Isostatic Adjustment and the Free Air Gravity Anomaly as a Constraint Upon Deep Mantle Viscosity, Geophys. J. Roy. Astro. Soc., 74, 377–449, 1983.

Yoder, C. F., J. G. Williams, J. O. Dickey, B. E. Schutz, R. J. Eanes, and B. D. Tapley, Secular Variation of Earth's Gravitational Harmonic J_2 From LAGEOS and the Nontidal Acceleration of Earth Rotation, Nature, 303, 757–762, 1983.

Yuen, D. A., W. R. Peltier, and G. Schubert, On the Existence of a Second Scale of Convection in the Upper Mantle, Geophys. J. Roy. Astro. Soc., 65, 171–190, 1981.

Zuber, M. T., E. M. Parmentier, and R. C. Fletcher, Extension of Continental Lithosphere: A Model for Two Scales of Basin and Range Deformation, J. Geophys. Res., 91, 4826–4838, 1986.

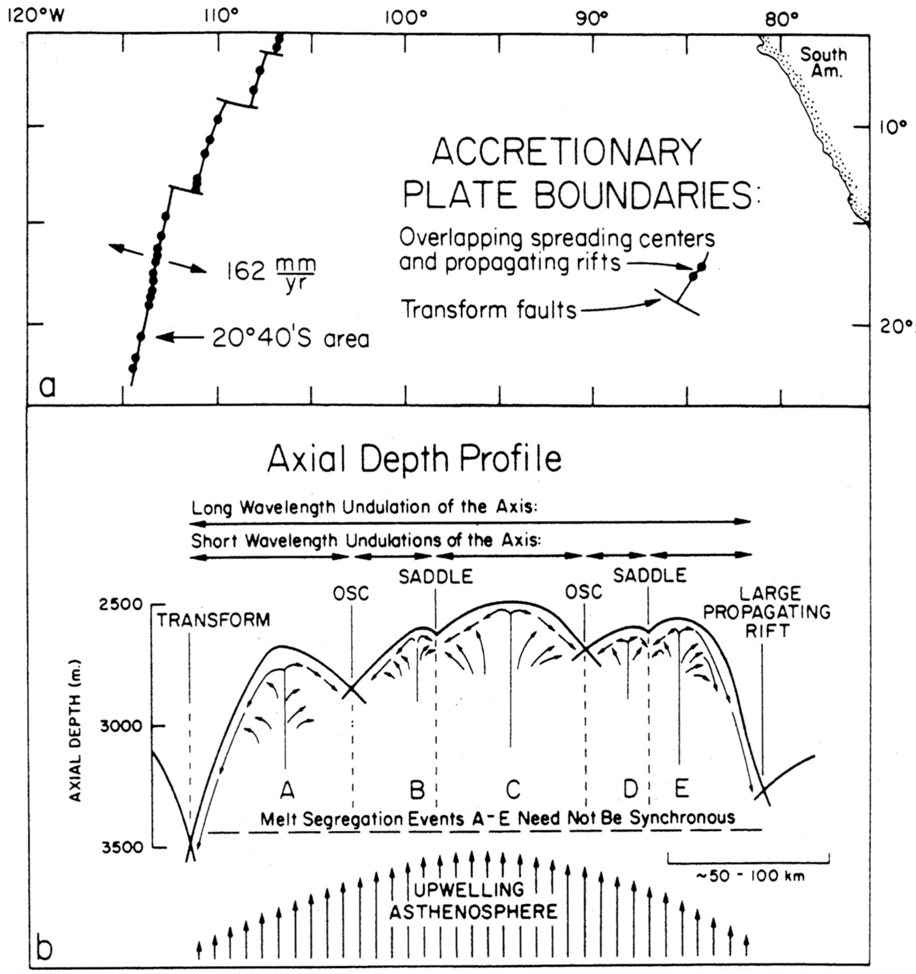


Figure 1.

(a) Overlapping spreading centers (OSCs)and propagating rifts on southern East Pacific Rise. Smallerridge-axis discontinuities such as saddle points, devals, and small, non-overlapping offsets are not shown here. (b) Model formidocean ridge segmentation. Major episodes of meltsegregation occur near regions A–E in upper mantle, leadingto replenishment and swelling of axial magma chamber. Magmamigrates downhill along-strike, away from sources of melt. Smallridge-axis discontinuities occur at distal ends of thesemagma pulses. OSCs occur where magma pulses have misalignedsuch that they produce overlapping en echelon offset. From Macdonald et al. (1987).

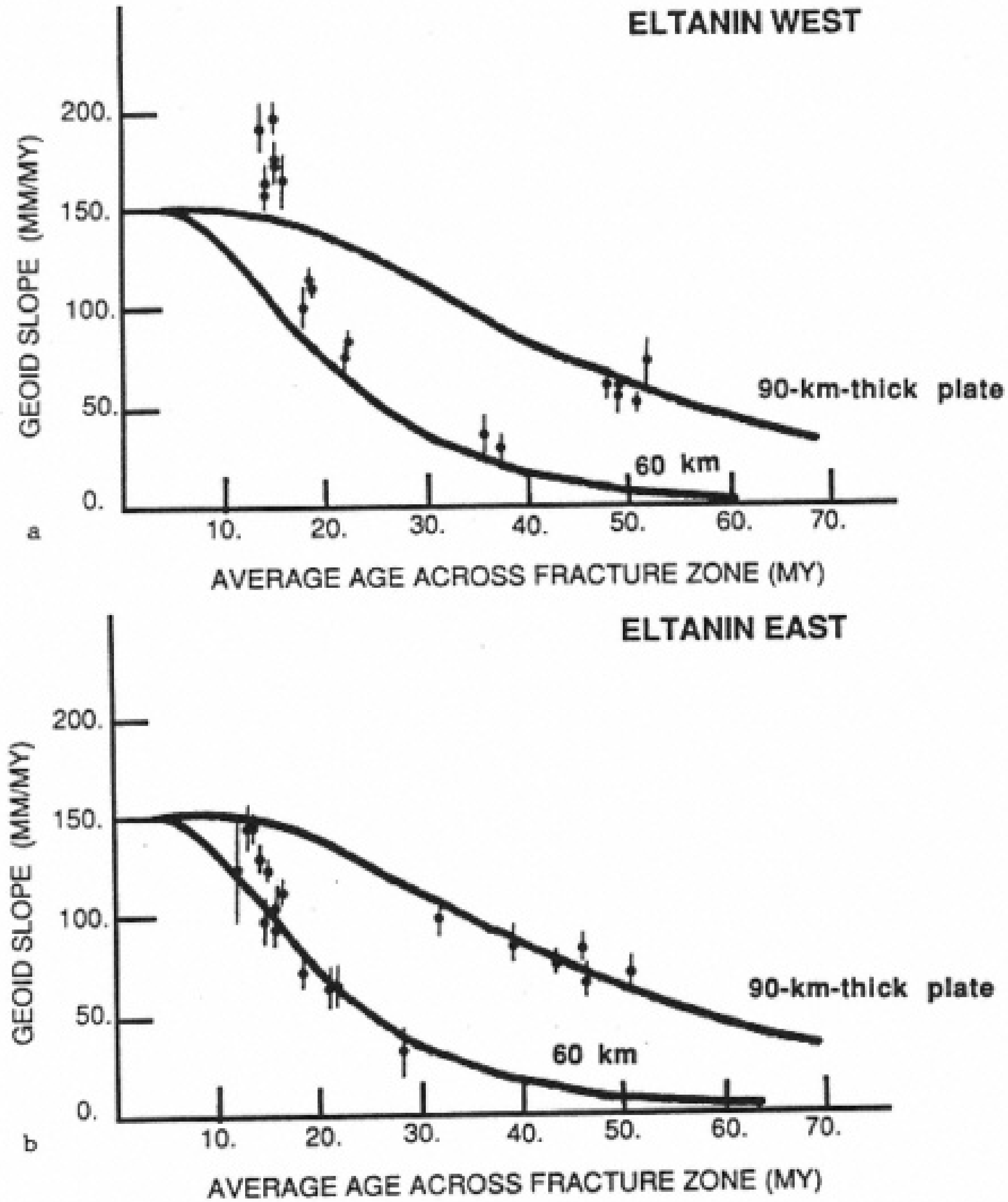


Figure 2.
Geoid slope estimates across the Eltanin Fracture Zone system. The dots represent the observed change in geoid across the fracture zone divided by the age contrast plotted as a function of the average of the ages on both sides of the fracture zone. The continuous curves give the expected relationship between geoid slope and average age for thermal plate models with lithospheric thicknesses of 90 km and 60 km. (a) Data for the western (Antarctic) limb of the fracture zone. (b) Data for the eastern (Pacific) limb. After Driscoll and Parsons (1987).

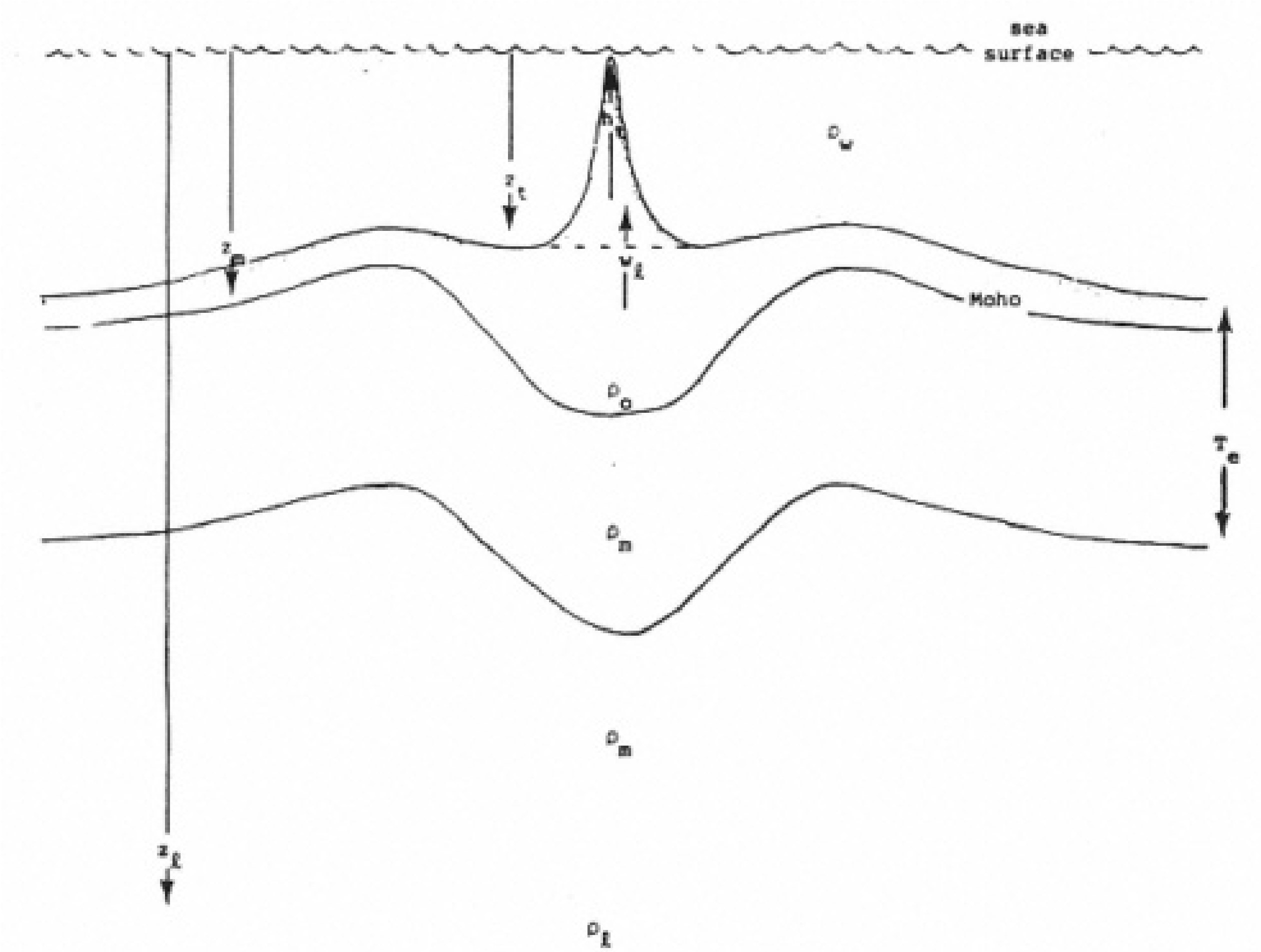


Figure 3.
Schematic cross section of a midplate swell. The volcano with height h_t and density P_o loads an elastic plate of thickness T_e, which corresponds to the depth to the thermal-controlled elastic ductile transition. The flexure of the elastic plate is superimposed on a broader midplate swell of height w buoyed up by some low-density anomaly P of thermal origin at depth z. The average depth of the bathymetry below the sea surface and the depth to the oceanic Moho are z_t and z_m, respectively. The density of seawater and mantle material are P_w and P_m, respectively. From McNutt and Shure (1986).

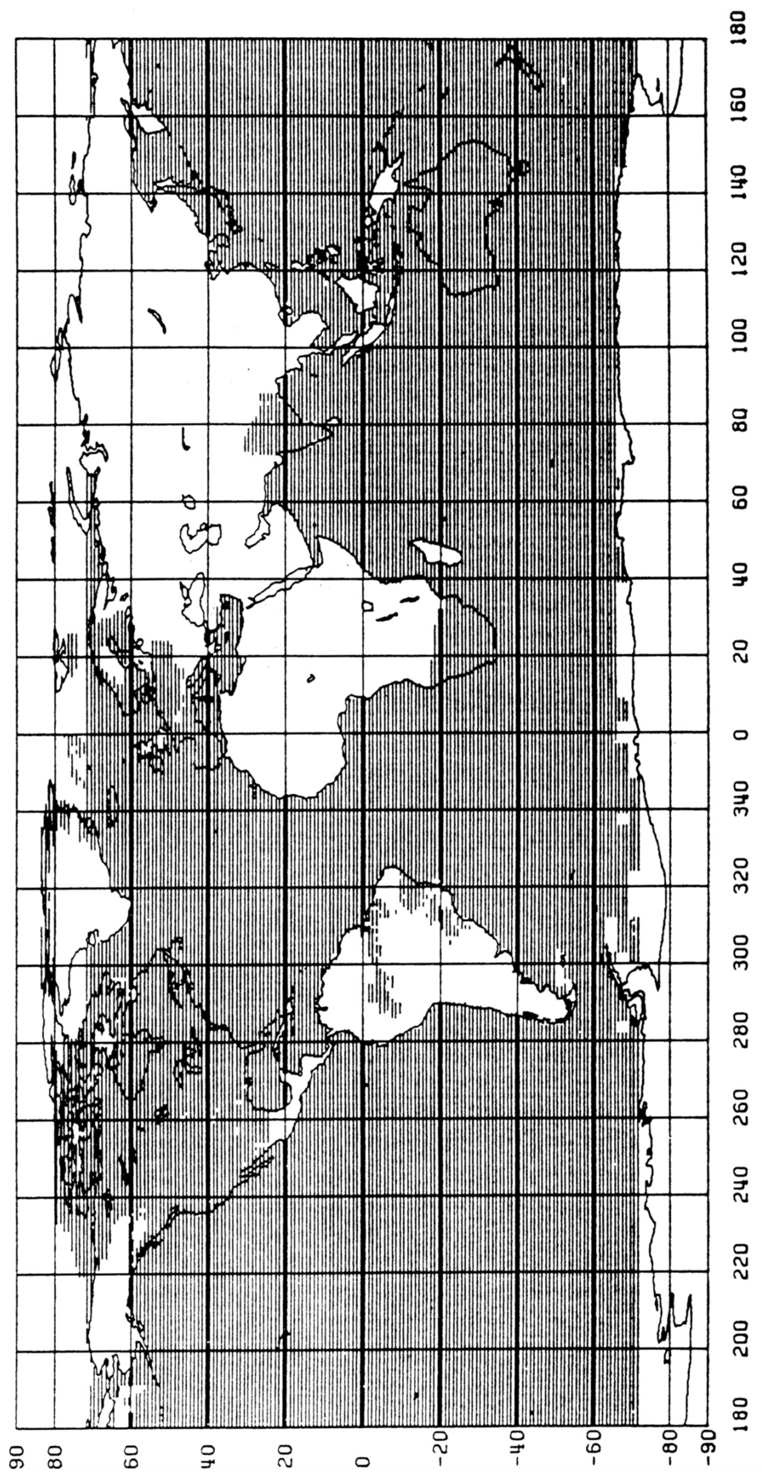


Figure 4.
Blank areas show locations of 1°x1° cells without gravity data from either terrestrial or altimetric measurements. From Rapp and Cruz (1986).

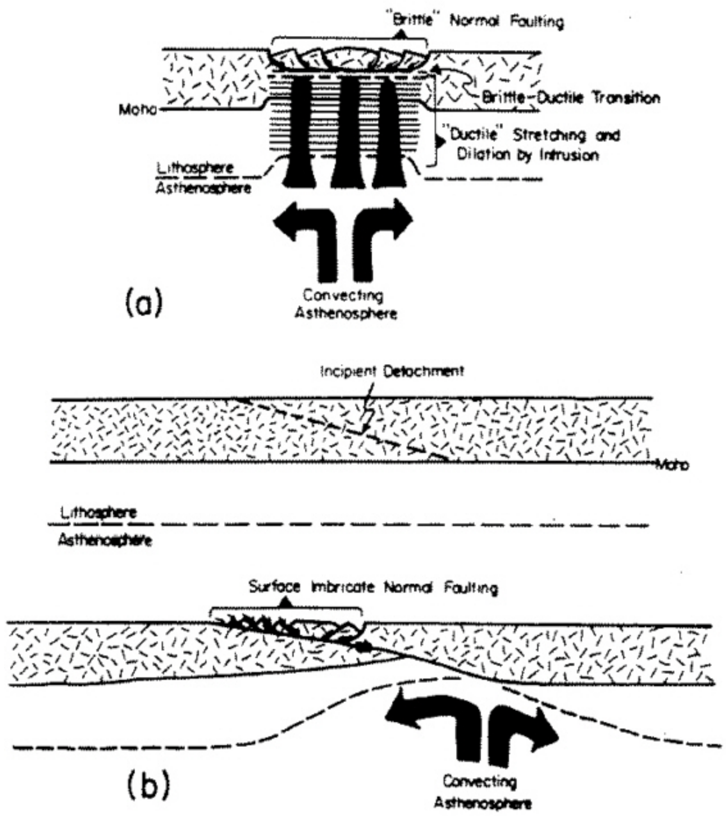


Figure 5.
End-member models of strain geometry in zones of lithospheric extension. (a) "Pure Shear" model, in which crust and mantle lithosphere are attenuated uniformly along any given vertical reference line. (b) "Simple-shear" model, in which relative extension of crust and mantle lithosphere along any given vertical line is non-uniform. From Wernicke (1985).

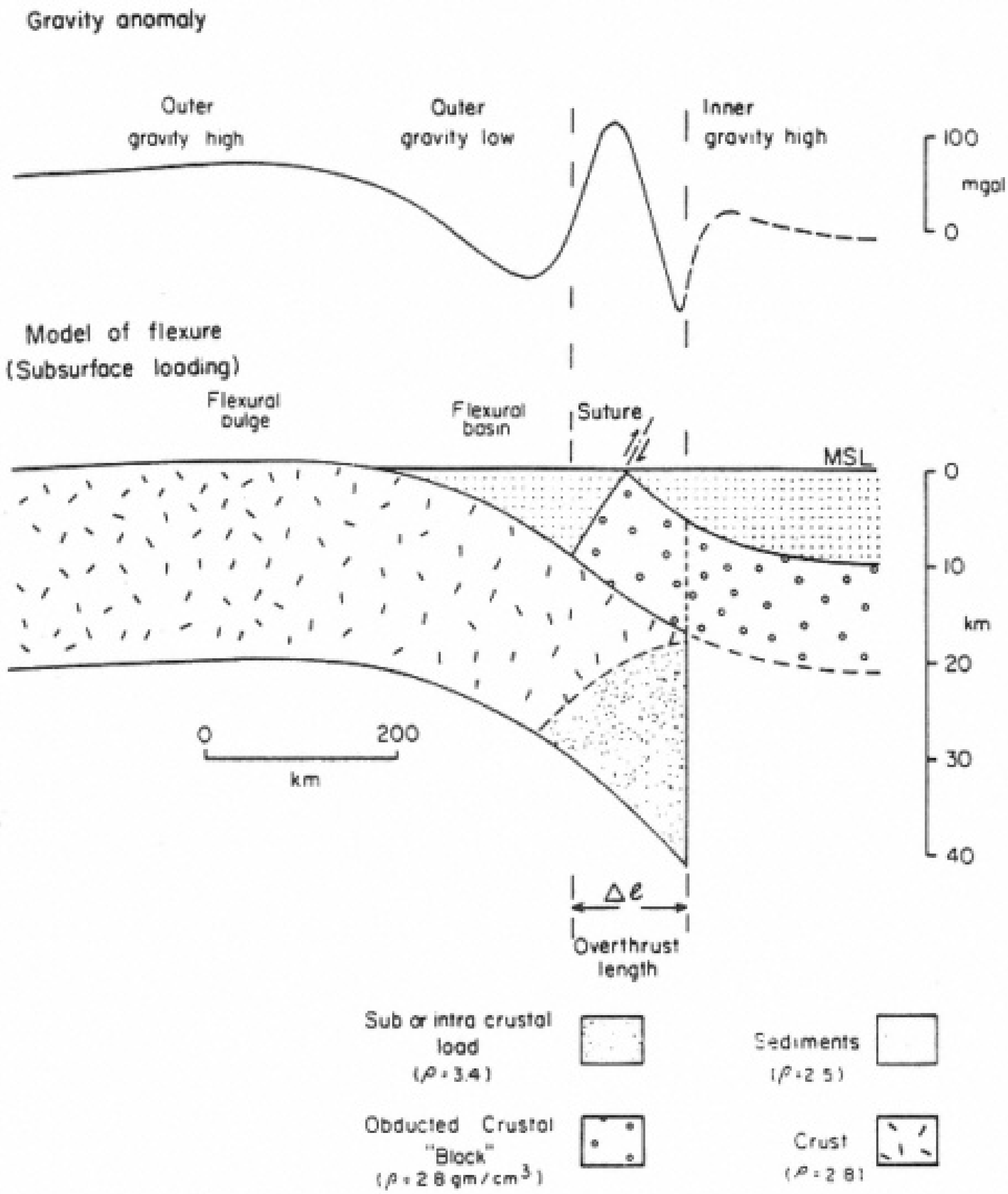


Figure 6.
Schematic model of the crustal structure and the predicted (Bouguer or free air) gravity anomaly over a completely eroded orogen. The flexure of the elastic plate on the left side is maintained by the weight of either an obducted block from the overriding plate or some intracrustal load. The gravity anomaly is characterized by a positive negative couple in which the low in the foreland is due to flexural depression of the basement and the high in the hinterland is caused by the excess mass of the buried load. From Karner and Watts (1983).

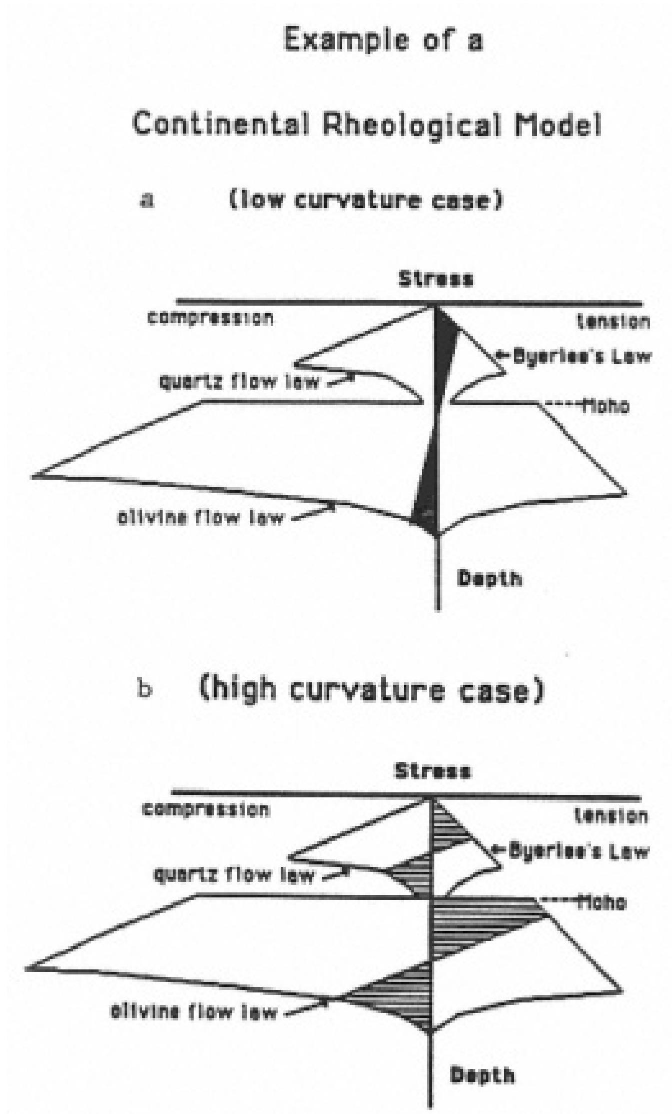


Figure 7.
Example of how a section of continental lithosphere, if bent to high curvature, can behave as a very thin elastic plate. Elastic stresses for the case of low plate curvature (a) and high plate curvature (b) are superimposed on a possible model for the strength of continental lithosphere as a function of depth. The strength in the upper crust and upper mantle is controlled by frictional sliding according to Byerlee's law. Strength in the lower crust and at the base of the elastic plate is limited by ductile flow laws for quartz and olivine, respectively. For the case of low plate curvature (a), the strength of rocks in the lower crust is not exceeded, and the plate flexes as one unit of relatively large thickness. For the case of high plate curvature (b), the strength in the lower crust is exceeded so that the plate decouples, flexing as two thin plates. Even though the depth to the elastic/ductile transition is the same for the two plates, the effective elastic thickness for case (b) is only half that for case (a).

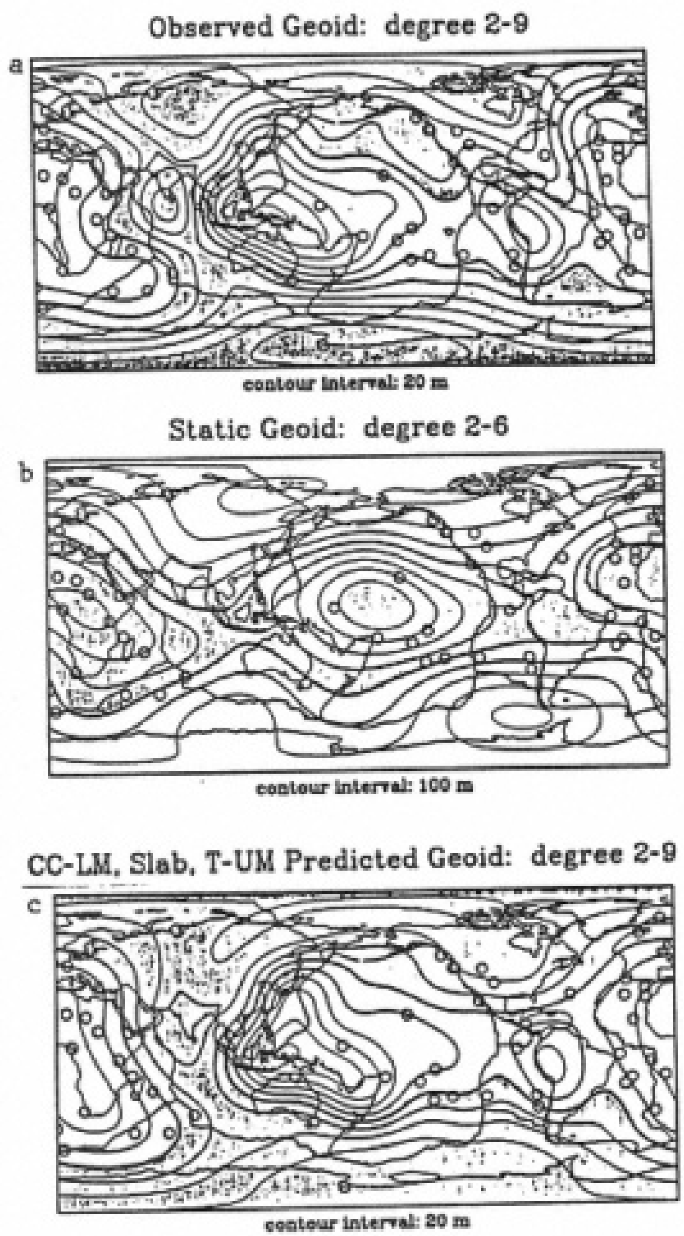


Figure 8.
Example of how geoid data can constrain the viscosity structure of the Earth. (a) The observed geoid at spherical harmonic degrees 2–9; (b) the predicted geoid from converting seismic velocity anomalies derived from mantle tomography to density anomalies, assuming that they are statically maintained within a rigid Earth. This geoid has the right pattern but the wrong sign. (c) The geoid predicted assuming that the density anomalies used in (b) drive convection in the mantle with an order-of-magnitude viscosity increase between the upper and lower mantle. The total geoid is now the sum of the effects of the density anomalies themselves plus the deformations they induce on the Earth's free surface and the core/mantle boundary. The model provides a remarkably good fit to the observed geoid. Open circles show locations of hot spots. From Hager and Clayton (1987).

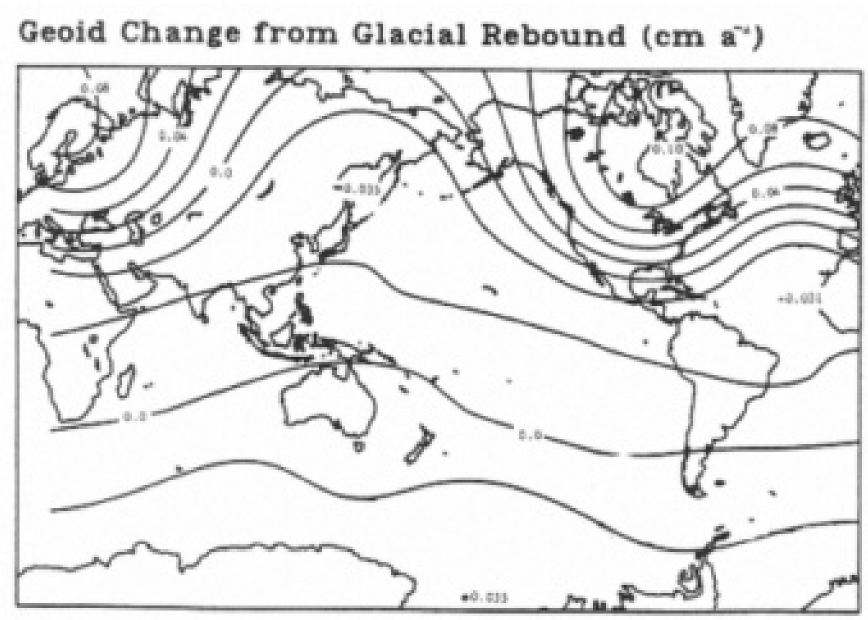


Figure 9.
Geoid change predicted from simplified model of postglacial rebound. Units are cm/yr. From Wagner and McAdoo (1986).

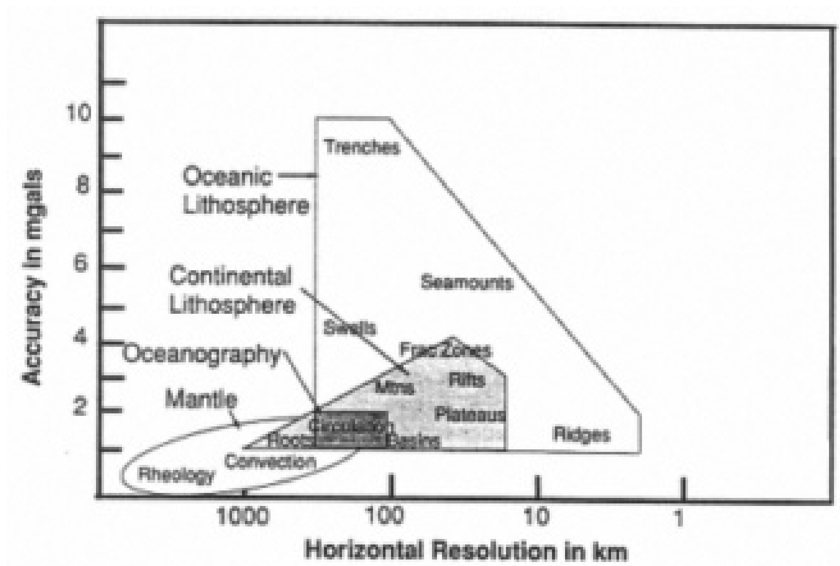


Figure 10.
Summary of requirements for gravity measurement accuracy as a function of spatial resolution for the problems discussed here midplate swell of

Common Interests in Geodetic and Oceanographic Research

Victor Zlotnicki
Jet Propulsion Laboratory
California Institute of Technology Pasadena, California

INTRODUCTION

The oceanographic community is planning comprehensive and ambitious studies of the circulation of the oceans on a global scale and in the key tropical region where such events as the El Nino/Southern Oscillation (ENSO) are more evident. The ocean is the key component of the climate system that temporarily stores heat and transports energy to different regions, hence these oceanographic studies are part of the World Climate Research Programme (WCRP, 1983). The World Ocean Circulation Experiment (WOCE) plans to provide during the 1990's "…the first comprehensive, global survey of the physical properties of the oceans…to establish the first global baseline for the long-term behaviour of the ocean, and to test computer models of the ocean circulation, vitally needed to understand global climate dynamics and to predict decadal climate change" (U.S. Science Steering Committee for WOCE, 1986). The Tropical Oceans-Global Atmosphere (TOGA) experiment is aimed at determining the extent to which the coupled tropical ocean-global atmosphere system is predictable on time scales from months to years, to learn whether the system can be modelled accurately enough to allow prediction, and to design an observational program that would allow operational prediction (WCRP, 1985).

Both WOCE and TOGA include as a key element measurements of sea level using in-situ and satellite systems and such measurements effectively require precision in position and gravity field knowledge that are current geodetic research. These areas of common interest are reviewed to understand the accuracy of the relevant geodetic quantities that would significantly help oceanographic research. The influence of ocean currents and changes in the volume of ice on Earth rotation provides other links between current oceanographic and geodetic research, but these will not be discussed here.

GEOID AT FIXED TIME VS TIME AVERAGED CIRCULATION

Oceanographers know the difference between the mean sea surface and the geoid to approximately 20 to 30 cm from 100 years of hydrographic data; however, due to data coverage, such a calculation is very accurate and has resolution of about 100 km in the North Atlantic and within 1000 km of the continents in the N. Pacific, is adequate and has resolution of a few hundred km in the central Pacific and the coastal areas of the southern ocean, and is fairly inaccurate and has resolution worse than 1000 km over most of the southern oceans (see Levitus, 1982, pp. 118–119). Oceanographers can measure the difference between sea surface and geoid to ±10 cm (Roemmich and Wunsch, 1982) using additional in-situ measurements. While such measurements are of limited coverage due to the slow speed of ships (relative to the quasi-synoptic coverage of an altimetric satellite), they provide essential information about the deep ocean layers, whose motion can be quite different, even opposite, to the surface motion defined by the difference between altimetric and geoidal height.

To improve on the oceanographic capabilities, the combined error of altimetric sea surface minus geoid heights should not exceed about 10 cm rms over all wavelengths larger than that implied by the Rossby radius of deformation. The radius of deformation (e.g., Gill, 1982) is the smallest length scale over which exact geostrophic balances can occur. The radius associated with the first baroclinic mode, about 10–30 km (Gill 1982, sec. 7.5) at mid-latitudes, implies a shortest wavelength between 63 and 189 km (the wavelengths increase towards the equator). Features whose size is comparable to the radius of deformation often are close to geostrophy but shift in position, such as the Gulf Stream axis, or drift, like eddies, so they can be measured as altimetric variability without using a geoid model. However, the extent to which such features remain in a fixed position in the oceans, perhaps tied to bathymetric features, and require a geoid model to be measured, is not clearly known. At wavelengths much larger than the minimum, let us say greater than 1000 km, geostrophy is approximated to a closer degree, and the error in the difference (altimetric sea surface-geoid) should be much lower than 10 cm, ideally of the order of 1 cm.

A more accurate gravity field model at wavelengths of 400 km and longer, as the Geopotential Research Mission's goal is, would help in quantifying permanent sea surface topography in two ways: (1) by reducing the error in the geoid that is subtracted from the corrected altimetric measurement, and (2) by reducing the error in the corrected altimetric measurement, whose largest error source is the uncertainty in the satellite height above the reference ellipsoid, dominated by imprecise modelling of the gravity field (Stewart et al., 1986, p. 7).

GEODETIC REFERENCE FRAMES AND THE POSITION OF TIDE STATIONS AND ALTIMETRIC SATELLITES

Several studies of global sea level trends (especially Barnett, 1983) show that sea level records since the beginning of the century contain a component coherent across various places on the earth: a trend of apparent rising global sea level at a rate of 15 cm/century. The interpretation of this result is complicated by the effect on this signal of inaccurately known continental uplift or subsidence, unknown geoid changes, and by the difficulties of precision levelling across whole continents or from a continental margin to a mid-ocean island. The first need, therefore, is to differentiate sea level motion from land motion and from geoid change at wavelengths greater than about 4000 km, over one decade, with accuracy of about 1 cm.

Since both tide gauges and altimeters measure sea level, the former providing excellent time coverage, the latter excellent space coverage, it is essential to have sea level stations and altimetric satellites in the same reference system at any time. If different reference frames must be used, it is essential to know their relative positions, at all times, to an accuracy better than that required for sea level measurements. This allows not only calibration of altimetric measurements from one satellite, but also comparison of the results of different satellite missions carried out at different times.

The U.S. Science Steering Committee for WOCE (1986) suggested taking advantage of GPS and VLBI developments that promise to allow the position of tide stations to be known at the 1 cm order of magnitude in the vertical relative to a geocentric reference frame. Diamante et al., (1987) have argued that the installation on the order of 100 well distributed and calibrated tide stations, accurately positioned with differential GPS measurements relative to a network of VLBI stations, plus another 150 less accurately positioned tide stations, would provide absolute sea level accurately enough for tropical studies (e.g., Wyrtki, 1984) and global sea level studies.

The reference frame implicit in the previous paragraph is different from that used to track altimetric satellites, a fact that could cause errors when tide gauges are used to calibrate satellite altimeters (e.g., Wunsch, 1986). As pointed out by Mather et al. (1979), observations to either extra-galactic sources (as in VLBI) or to earth-orbiting satellites (laser ranging or range-rate measurements) are sensitive to the instantaneous rotation vector of the earth relative to the observing station. In addition, the satellite observations are sensitive to the instantaneous position of the earth's centre of mass (‘geocentre') because it is the focus of any ellipse that osculates the orbit. At present, altimetric satellites are positioned with respect to a geocentric reference system known in a time averaged sense, because the coordinates of the observing stations are constrained by the observations to many satellites over many years. A global network of VLBI stations (to which the tide stations are proposed to be tied) defines a different reference frame, both because of its insensitivity to the earth's center of mass and because of different averaging times over which the reference frames are implicitly defined (see Mather et al., 1979, for estimates of the rates of change in coordinates). Hence, it is important that not only the position of tide stations, but also the relative positions of the two reference frames be measured with sufficient accuracy and frequency to allow decadal and secular studies of sea level change with both tide gauges and altimeters.

TIME VARYING MEAN SEA SURFACE VS TIME VARYING GEOID

Let us assume that accurately positioned tide gauges detect a tilt of 1 cm across an ocean basin over a period of a few years. Does it reflect a subtle change in ocean circulation or an equally subtle change in the geoid?

Wagner and McAdoo (1986) modelled a variety of possible changes in the geoid due to the mass redistributions in the "solid" earth associated with post-glacial rebound, ablation of continental glaciers, large earthquakes and steady-state plate tectonics. They find plausible rates of change in geoid height of the order of 0.1 cm/yr at wavelengths greater than 4000 km, a number of the same order of magnitude as the coherent component of tide records usually associated with sea level rise. Rubincam (1984) and Yoder et al. (1983) actually measured, albeit indirectly, the decrease in the J2 coefficient of the earth's gravity field between 1976 and 1981, based on the measured acceleration of the ascending node of the LAGEOS satellite. Rubincam's (1984) value of $(-8.2 \pm 1.8)\ 10^{-19}\ s^{-1}$ is equivalent to a geoid change of 0.016 cm/yr × $(3 \sin^2 ø-1)$, where ø is geographic latitude.

Even though any such tilts in mean sea surface would be interpreted in conjunction with other oceanographic data, it is highly desirable to avoid ambiguities by a long-term program of gravity monitoring. Any such geoid changes may also cause changes in the position of the geocentre, which affects the reference system from which altimetric satellites are observed and ultimately, the accuracy of altimetric observations of the ocean. Mather et al. (1979) point out that a network of 25 stations measuring absolute gravity to ± 10 microGal can detect geocentre motion of at least 1 cm/yr after 10 years. We also know that changes in the second harmonic of the earth's gravity field can be measured from changes in the orbit of the LAGEOS satellite. The elements seem to be available for a long-term program to monitor changes in the longest wavelengths of the gravity field that would distinguish these from changes in ocean circulation.

ACKNOWLEDGEMENTS

Discussions with Professors Carl Wunsch, Richard Rapp and Byron Tapley were very helpful. The research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

Barnett, T. P., Recent Changes in Sea Level and Their Possible Causes , Climatic Change, 5, 15–38, 1983.

Diamante, J. M., T. E. Pyle, W. E. Carter, and W. Scherer, Global Change and the Measurement of Absolute Sea Level, Prog. Oceanography, 18, 1–21, 1987.

Gill A. E., Atmosphere-Ocean Dynamics, Academic Press, 662 pp., 1982.

Levitus S., Climatological Atlas of the World Ocean, NOAA Prof. Paper 13, National Oceanic and Atmospheric Administration, U.S. Dept. of Commerce, (available from the author at Geophys. Fluid Dyn. Lab., P.O. Box 308, Princeton, NJ 08540), 1982.

Mather, R. S., C. Rizos, R. Coleman, and E. G. Masters, Geodetic Reference Systems for Crustal Motion Studies, Tectonophysics, 52, 15–37, 1979.

Roemmich D. and C. Wunsch, On Combining Satellite Altimetry with Hydrographic Data, J. Mar. Res. 40, 605–619, 1982.

Rubincam, D. P., Postglacial Rebound Observed by LAGEOS and the Effective Viscosity of the Lower Mantle, J. Geophys. Res., 89, 1077–1087, 1984.

Stewart, R., L.-L. Fu, and M. Lefebvre, Science Opportunities from the Topex/Poseidon Mission, National Aeronautics and Space Administration, JPL Publ. 86-18 (available from Topex Project, Jet Propulsion Lab., Pasadena, CA 91109), 1986.

U.S. Science Steering Committee for WOCE, Status Report on U.S. WOCE Planning, U.S. WOCE Planning Report Number 3, 229 pp., U.S. Planning Office for WOCE, College Station, TX, 1986.

Wagner, C. A. and D. C. McAdoo, Time Variations in the Earth's Gravity Field Detectable with Geopotential Research Mission Intersatellite Tracking, J. Geophys. Res. 91, 8373–8386, 1986.

WCRP-Joint Scientific Committee and Committee on Climatic Changes and the Ocean, Large-Scale Oceanographic Experiments in the World Climate Research Programme, WCRP Publication Series No. 1, 121 pp., Intergovernmental Oceanographic Commission (UNESCO), Paris, France, 1983.

WCRP-Toga Scientific Steering Group, Scientific Plan for the Tropical Ocean and Global Atmosphere Programme, WCRP Publication Series No. 3, 147 pp., Intergovernmental Oceanographic Commission (UNESCO), Paris, France, 1985.

Wyrtki, K., The Slope of Sea Level Along the Equator During the 1982/1983 El Nino, J. Geophys. Res., 89, 10419–10424, 1984.

Wunsch, C., Calibrating an Altimeter: How Many Tide Gauges is Enough? , J. Atmosph. and Oceanic Technology, 3, 746–754, 1986.

Yoder, C. F., J. G. Williams, J. O. Dickey, B. E. Schutz, R. J. Eanes and B. D. Tapley, Secular Variation of Earth's Gravitational Harmonic J2 Coefficient from LAGEOS and Non-Tidal Acceleration of Earth Rotation , Nature, 303, 757–762, 1983.

Lasers for Geodesy in the Year 2000

David E. Smith
NASA Goddard Space Flight Center

INTRODUCTION

Laser ranging to earth satellites and the Moon began in the sixties with range accuracies of several meters and has developed over the last two decades to the one centimeter level to become a major geodetic tool for addressing global and regional scale geophysical problems. Major contributions have been made to our knowledge and understanding of the Earth's gravity field, the shape and size of the Earth, the motions of the major tectonic plates, the earth and ocean tides, and to our understanding of the orbit of the moon. During the next decade there will be significant advances in the technologies used in laser ranging which will be used to help meet the ever-increasing demands in scientific problems for greater accuracy, greater frequency of measurements, less time required to make measurements and, of course, lower cost.

In 1983 NASA organized a meeting called the Airlie House Conference (Walter, 1984) to discuss the scientific requirements for future space geodetic systems, including laser ranging. The report of this conference spelled out the measurement requirement of millimeters on a local, regional, and global scale for many future scientific problems.

In the following sections I briefly describe the technological developments that are expected over the next few years that will influence the growth of laser ranging as a geodetic tool. I will also discuss ranging and altimetry systems that are being developed by NASA as part of its research and development program.

TECHNOLOGY DEVELOPMENTS

Two main areas of technology development during the next decade that will significantly influence the growth of laser ranging in all its forms are in the transmitter and in the receiver. present day transmitters consist of a flashlamp pumped, mode-locked, Q-switched Nd:YAG short pulse laser oscillator, a double-pass amplifier, and a frequency-doubling crystal (ref. Cohen and Degnan, 1987). Because of the need for increased accuracy and more efficient operation, future systems are expected to employ diode pumping and operate at two frequencies (532 and 355 nm). Diode pumping of the laser offers the potential for greater prime power efficiency, longer lifetimes, and a smaller, lighter weight system. Dual frequency operation permits the atmospheric delay to be derived directly from the measurement instead of relying on the modeling of the atmospheric delay based on a knowledge of pressure, temperature, and humidity at a point on the Earth's surface.

Diode pumping of the laser is an order of magnitude more efficient than flashlamp pumping, and therefore, results in greater laser lifetime. This is of less importance for ground-based laser ranging but is critical for space-based systems where maintenance is very restricted. Further, great efficiency means less heat is generated so that higher pulse rates can be used and systems can be smaller. In addition, unlike the flashlamp, the method of failure of a diode system is gradual.

At the present time flashlamps used in laser ranging systems have effective lifetimes on the order of 10^6 shots--months of actual operation or days of continuous operation, while the potential for diode pumping is on the order of 10^9 shots--decades of actual operation or years of continuous operation. This development of long lifetime lasers makes spaceborne systems viable.

A further development, which in part results from the greater efficiency of the diode pumping, is the shortening of the transmitted laser pulse length by two or three orders of magnitude. Present laser tracking systems have pulse lengths of 10 to 100 picoseconds (millimeters), but because of the increased accuracy requirements and the two-color capability, the lasers of the near future will need to have pulse lengths of the order of 10^{-15} seconds. Lasers with pulse lengths of 10×10^{-15} seconds are already operating in laboratories (Degnan, private communication, 1987).

Table 1 shows a comparison of some characteristics of a laser transmitter in 1987 with a system that can be expected before the year 2000.

The major development in the receiver will be the introduction of the streak camera which effectively transforms the return pulse from the time domain to the spatial domain. Present-day laser ranging systems employ a photomultiplier to measure the amplitude of the signal and the time of flight. In the two-color lasers of the future the streak camera will act as a sub-picosecond resolution timing system for the green (532 nm) and near-ultraviolet (355 nm) pulses, thus enabling the dispersion caused by the atmosphere to be measured on a pulse by pulse basis (Cohen et al., 1987). If the separation of the green and near-ultraviolet return pulses can be determined to 200 to 300×10^{-15} seconds, the atmospheric correction can be determined to 1 mm. This is approximately the capability of present streak tubes, and improvements over the next decade are expected to lead to resolutions of 50 to 100×10^{-15} seconds. Atmospheric corrections will then be obtainable at the sub-millimeter level and the overall range accuracy of the tracking system to the one millimeter level.

Finally, in order to measure the range to an object, we must consider the contribution of the cornercube array to the total error. The most precisely configured array in orbit is that of the LAGEOS spacecraft which is believed to be accurate to the few millimeter level (Fitzmaurice et al., 1977), but the majority of arrays are not this precise. In order to fully utilize the accuracy of the millimeter systems of the future, improved accuracy retro-reflector systems need to be designed to sub-millimeter level. The arrays will need to be smaller, so as to minimize pulse spreading, but large enough to provide sufficient return signal. This may cause pulse rates to increase from their present 5 to 10 pulses per second to the order of 1000 pulses per second, and for return pulses to be averaged to produce an accurate measurement. The design of the laser array and precise knowledge of the array's position with respect to the center of mass of the spacecraft could become a limiting factor in the development of increasingly accurate laser ranging systems.

GROUND-BASED LASER RANGING

Table 2 shows a comparison of some of the principal components of a ground-based laser ranging system of 1987 with a projected system of the future. The future system reflects the new technology developments previously described and includes the expectation that at least part of the future system will be automated. The degree of automation will depend on the operational philosophy rather than the technology. The development of fully automated ground-based systems will certainly be possible in the near future because to a large degree they will employ the same technology as the spaceborne systems. Fully automated systems could have many advantages, particularly in the cost of operation, but real-time communications systems will be needed for data relay and for routine operation. (For a full description of satellite laser ranging, the reader is referred to Degnan, 1985).

SPACE-BASED LASER RANGING

During the next decade we can reasonably expect that significant progress will be made toward putting a laser ranging system into space. In the mid-seventies the preliminary design of such a system was developed for operation from the Space Shuttle and/or a free flying spacecraft (Vonbun et al., 1977; Smith, 1978). A spaceborne laser is currently scheduled to fly on one of the EOS (Earth Observing System) platforms for launch in the 1997 timeframe (Degnan and Cohen, 1988).

The concept of the spaceborne laser is that precise range measurements can be made from an orbiting space platform to a network of laser retro-reflector arrays on the earth's surface and that from these measurements the relative locations of the retro-reflector arrays can be derived. Figure 1 shows the concept of the spaceborne laser. The general characteristics of the EOS system are shown in Table 3. As the spacecraft passes over the region, the laser makes range measurements to the retro-reflectors according to a pre-determined pattern, remaining on an array for just a few seconds before moving to the next. The laser has been designed to operate at two frequencies, 532 nm and 354 nm, in order to be able to remove the large effects of atmospheric refraction, and therefore, the need for meteorological sensors at the ground arrays. The basic measurement will be accurate to one centimeter or better, including any remaining atmospheric error.

LASER ALTIMETRY

A recent development important for geodesy has been the introduction of the laser altimeter. Already operating experimentally from aircraft, the laser altimeter sends a short pulse of radiation to the Earth's surface which is “reflected” back to the receiver in the aircraft. Conceptually, the laser altimeter is similar to the laser ranging system except that it operates without retro-reflectors, using only the radiation scattered back from the surface below. The system can provide topographic information over almost any surface, including land, snow, ice, oceans, and cloud tops. A spaceborne altimetric capability is planned as part of the EOS spaceborne laser facility. The fundamental frequency of the laser (1064 nm) is used for altimetry, and is not needed for ranging. The altimetric component will provide decimeter precision topography of the Earth's surface with horizontal resolutions of 80 and 160 meters (the latter with contiguous spots). This level of horizontal resolution is extremely difficult to obtain from a conventional radar altimeter.

A laser altimeter is also being developed for operation around the Moon (Garvin et al, 1987; and 1988). This developmental system, the Lunar Observer Laser Altimeter (LOLA), could fly on the proposed Lunar Observer orbiting 100 km above the lunar surface and provide decimeter vertical precision topographic profiles with horizontal resolutions of tens of meters. LOLA is designed to operate in two horizontal resolution modes. In the mapping mode, the system would obtain 300 meter spots that would be interpolated to provide a 1–2 kilometer global lunar topographic grid appropriate for geodetic and geophysical studies and regional characterization of all major terrain types. In the high resolution mode, LOLA would obtain 30 meter spots to be used for localscale geological profiling. The absolute limitation on the accuracy of this system will be determined by the knowledge of the spacecraft orbit and the lunar gravity field in particular.

Table 4 shows some of the performance characteristics of the altimetric functions of the EOS spaceborne laser ranger and the developmental lunar altimeter. A point of special importance in Table 4 is the factor of sixty greater power that is planned for the EOS system as compared to the lunar system. This simplifies the lunar system considerably by increasing the life expectancy of the system and reducing its size and weight. The Moon is an ideal object for a laser altimeter because a low altitude orbit is possible and atmospheric absorption is not a problem. In fact, a LOLA-class altimeter could be modified for orbital or fly by missions for any solid planetary body without an atmosphere (Garvin et al, 1987a).

CONCLUSIONS (PREDICTIONS)

The next decade will see the introduction of several new technologies into laser ranging, in particular: diode pumped lasers, two frequency ranging, streak cameras, and shorter pulse lengths. These technologies and developments will lead to (1) ground-based laser ranging systems accurate at the 1 mm level, including the effect of the atmosphere; (2) spaceborne laser ranging at the 1 cm level, or better, from the EOS platform or a similar system; and (3) laser altimetry providing topographic information of the Earth and Moon, and possibly the planets, at the 30 cm vertical precision level with decameter horizontal resolution. These capabilities, in conjunction with those occurring in the other areas of geodetic measurement science, will have a profound influence on geodesy and many associated disciplines.

ACKNOWLEDGEMENTS

I should like to acknowledge the thoughts and ideas of many of my colleagues at GSFC in this forage into the near future. In particular, Dr. Steven Cohen, Dr. John Degnan, Dr. James Garvin, and Dr. Maria Zuber have provided me the benefit of their own crystal balls in this gaze into the next decade plus. However, I take full responsibility for the predictions contained above. In addition, I am pleased to acknowledge the financial support provided for the instrument development by Dr. E. A. Flinn, NASA Geodynamics Program, for the ground-based and space-based ranging; and by Dr. L. Evans, NASA PIDDP Program, for the lunar laser altimeter.

REFERENCES

Cohen, S. C., and J. J. Degnan, Spaceborne Laser Ranging from EOS , Proceedings of IGARSS '87 Symposium, Ann Arbor, 18–21 May 1987.

Cohen, S. C., J. J. Degnan, J. L. Bufton, J. B. Garvin, J. B. Abshire, The Geoscience Laser Altimetry/Ranging System, IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-25, No. 5, September 1987.

Degnan, John J., Satellite Laser Ranging: Current Status and Future Prospects, IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-23, No. 4, July 1985.

Degnan, J. J., and S. C. Cohen, Applications of a Spaceborne Laser Ranger on EOS, SPIE LASER '88 Conference, 1988.

Fitzmaurice, M. W., P. O. Minot, J. B. Abshire, H. E. Rowe, Prelaunch Testing of the Laser Geodynamics Satellite (LAGEOS), NASA Goddard Space Flight Center Tech. Paper 1062, October 1977.

Garvin, J. B., J. L. Bufton, J. B. Abshire, M. T. Zuber, Laser Altimetry in Planetary Geology, Lunar Planet. Sci. Conf. XVII, 381–391, 1987.

Garvin, J. B., M. T. Zuber, J. L. Bufton, Planetary Laser Altimetry , IUUG XIX Assembly, 99, 1987a.

Garvin, J. B., M. T: Zuber, J. L. Bufton, Lunar Observer Laser Altimeter: Geoscience Applications, Lunar Planet. Sci. Conf. XIX, 379–380, 1988.

Smith, D. E., Spaceborne Ranging System, Proc. 9th GEOP Conf., Dept. of Geodetic Science Rept. No. 280, Ohio State University, October 1978.

Vonbun, F. O., W. D. Kahn, P. D. Argentiero, D. W. Koch, Spaceborne Earth Applications Ranging System, J. Spacecraft & Rockets, Vol. 14, pp. 492–495, 1977.

Walter, L. S., Geodynamics, Proc. of Airlie House Conf., NASA Conf. Publ., #2325, 1984.

Table 1: The Laser Transmitter

1987		2000
1 color	---------->	2 colors
flash lamp pumping	---------->	diode pumping
flash lamp efficiency:		
1% – 2%	---------->	20% – 30%
laser lifetimes:		
10^6 shots	---------->	10^9 shots
pulse length:		
100×10^{-12} secs	---------->	10^{-15} secs

Table 2: Ground-Based Laser System

	1987	2000
Accuracy (best systems)	1 cm (excl. atm)	1 mm (incl. atm)
Pulse rate	5–10 pps	100–1000 pps
Laser	Nd:Yag 532 nm	Nd:Yag 532, 355 nm
Energy	100 mj	10 mj
Pulse length	100×10^{-12} secs	10×10^{-15} secs
Data communication	tapes, etc. via mail	real-time via satellite
Automation	none	some--> complete

Table 3: Specifications for the EOS Spaceborne Laser

Laser	frequency doubled & tripled, mode locked Nd:Yag
Pulsewidth	100 picoseconds, FWHM
Energy/pulse	120 millijoules (1064 nm)
	60 millijoules (532 nm)
	20 millijoules (354 nm)
Beam divergence	0.1 milliradians
Maximum pulse rate	40 pps
Receiver telescope diameter (ranging)	18 cm

Table 4: Characteristics of the EOS and lunar laser altimeters

	EOS	Lunar
Orbital altitude	800 km	100 km
Frequency	1064 nm	1064 nm
Pulsewidth	100 ps	3 ns
Energy/pulse	120 mj	2 mj
Receiver telescope diameter	50 cm	25 cm
Spot size	80, 160 m	30, 300 m
Vertical precision	10–50 cm (excl. atmosph.)	30 cm
Pulse rate	40 Hz	10, 50 Hz

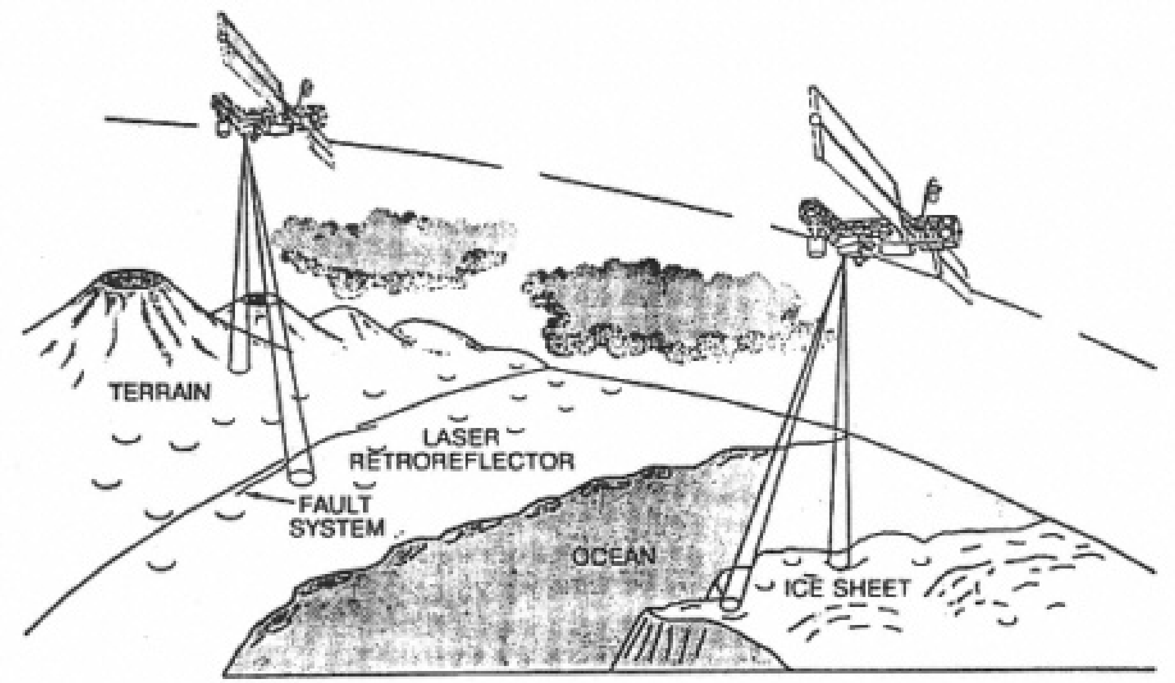


Figure 1.
Spaceborne laser ranging and altimetry concept.

Seafloor Geodesy by the Year 2000

F. N. Spiess
Scripps Institution of Oceanography University of California, San Diego LaJolla, California 92093

INTRODUCTION

Other papers in this group make it clear that high technology geodesy, with particular relevance to geodynamic problems, is a well-advanced maturing field of endeavor as applied to terrestrial situations. Lasers, satellites, high precision clocks, advanced signal processing and data reduction capabilities are moving us into second and third generation versions of systems brought to operation initially in the 1970s. Unlike the terrestrial situation, however, seafloor geodesy in the geodynamic context still only consists of paper analyses, workshop reports, and the beginnings of testing of a few potential system elements at sea. Nevertheless, the principal message of this paper is that by the year 2000 we will be discussing at least a few real multi-year data sets and using them to constrain our models of the structure and dynamics of the crust beneath the sea-its genesis, its evolution as it moves away from the mid-ocean ridges, its destruction in the trenches, and the effects of its interaction with continents and islands.

If this prediction is to be realized, we need to achieve three things:

1. Development of some new elements of undersea technology.
2. Application of a variety of systems at a few interesting initial sites.
3. Programmatic support frameworks for subsequent long periods of observation.

This paper will discuss each of these three topics briefly in the order listed.

TECHNOLOGY

A discussion of technology must start with some measurement goals. Seafloor geodynamic problems are such that first generation systems would be useful if they could produce accuracies of a few centimeters for measurements of horizontal or vertical position change, strain measurements of a few parts in a million, and tilt determinations in the range of a few microradians. Beyond that, whatever means are used to make these measurements must be compatible with the seafloor environment and capable to being used to monitor changes over periods of many years.

An encouraging aspect of the technological problem is that a large number of options are potentially available. Some of these involve the possibility of fairly direct transfer of well-established land techniques, while others are derived from undersea developments related to other goals.

At least five differences between subaerial and subsea circumstances drive the design and development of seafloor geodetic systems (NAS, 1983):

1. Electromagnetic radiation at frequencies high enough to be useful is highly absorbed in the sea because of the electrical conductivity of seawater. The only useful range is the optical one, but even there, one's capabilities are limited to tens, or perhaps a few hundred meters.
2. Acoustic energy of appropriate wavelength can be transmitted effectively over distances of many kilometers.
3. Ambient pressure at the seafloor is not only large, it is relatively free of large amplitude short wavelength or short time scale irregularities.
4. The seafloor environment is quite stable in terms of temperature and sediment water content.
5. Access to the seafloor is much more difficult than for most terrestrial sites.

These considerations lead to some immediate conclusions about the most fruitful directions that system developments are likely to take. First, acoustic systems will dominate for measurement of horizontal or slanting distance measurements involving distances of more than about 100 m. At the same time, shorter range systems now in use on land, but that suffer from environmentally induced monument motion problems should behave better in the deep sea. Finally, one must strive for simplicity in system installation and maintenance.

The techniques that should be most readily transferred from the land are those that operate effectively over rather short distances. Mechanical or laser strain measuring systems, as well as both short and long baseline tiltmeters (Agnew, 1987) also fall in this category. They will benefit from having substantially reduced “noise” due to monument instabilities. Some of them, however, introduce collateral requirements for substantial power and for continuity of operation. They also introduce challenging installation problems, as will be discussed below.

Laser ranging through seawater (as opposed to systems in which the path is contained in a pipe or optical transmission fiber) would be limited to ranges of at most a few hundred meters by the attenuation of light in seawater-about 200 decibels per kilometer in clear water. Speed of propagation of light in seawater is only known to about a part in 10^4. The change of speed with temperature and salinity (Stanley, 1971) is small, however; thus, as long as the laser power levels are modest enough that they do not create appreciable localized temperature increases, they can be useful in detecting path length changes with time. The localized heating problem prevents one from overpowering the high attentuation by brute force introduction of massive transmitted power.

Borehole oriented approaches using strain or tilt measuring devices installed in deep-sea drilling program holes should have no problem. Similarly, well-logging methods used to detect changes in the cross-sectional shapes of holes or to document orientation of breakouts in order to infer the nature of stresses in the uppermost parts of the crust should operate as well at sea as on land once there are established systems for doing wireline re-entry into deep-sea drill holes (Langseth & Spiess, 1987).

One new class of systems would involve the use of pressure measurements to determine the distance of the seafloor below the sea surface. With present state of the art in measurement of temperature and salinity as a function of depth, and knowledge of how to convert these measurements to water density (Saunders, 1981), one should be able to convert pressure measurements to depths with accuracy of a part in 10^5 (Reid, 1984). Quartz crystal pressure gauges offer the best possibility as far as accuracy is concerned (Irish and Snodgrass, 1971), although these come in a variety of configurations, some of which actually utilize mechanical elements (e.g., Bourdon tubes and bellows) as intermediate links and thus suffer from drift problems that degrade their performance in this context (Watts and Kontoyiannis, 1986; Wearn & Larson, 1984; Busse, 1987).

The major new class of systems entering geodesy are those utilizing underwater sound transmission. They are based on over 50 years of ocean acoustics research and development (primarily oriented toward submarine detection) (Eckart, 1968; Urick, 1975) but with a long history of application in marine geology and geophysics (Spiess, 1987). Early discussion of their usefulness in geodesy occurred at a symposium in 1966 (Speiss, 1966), although at that time the goals set were not as stringent as those being discussed here.

The principal limitation on use of sound propagation through seawater for distance measurement is our ability to know the velocity with which to convert travel time to distance. In this, one has a choice between using a direct measuring sound velocity meter or measuring pressure, temperature and salinity and converting these, via empirical relationships based on laboratory data into sound velocity.

At the present time the latter approach is only good to a few parts in 10^5 (Lovett, 1978). The direct measurement method relies on the fact that, even for geodetic purposes, sound propagation in seawater is essentially non-dispersive (Urick, 1975). One can thus use a small, dimensionally stable device, operating in the megahertz frequency range, to measure travel times, calibrate it in pure water (for which the relationship between pressure, temperature, and sound speed are known to a part in 10^6 (Greenspan, 1972)), and use the resulting velocity determination for systems operating as low as 10 kHz. The best such meter devised to date had its motivation in geodetic application and is capable of a little better than a part in 10^5 (McIntyre and Boegeman, 1986). Clearly, there is room for technological improvement in this area.

The limit on sound velocity measurement capability translates directly into a distance measuring accuracy limit of the same amount, 1 in 10^5. This breaks down at the short range end at distances of only a few meters, at which there is difficulty in determining the locations of the acoustic centers of the transducers. At the long range end it simply does not meet our goal of a maximum of a few centimeters uncertainty beyond about 10 km. There is, however, a fundamental environmental constraint. Since the sound velocity field must be determined experimentally by moving one, or perhaps a few, point measuring instruments through it, there can be dynamic situations in which the spatial variations of water temperature and salinity can vary rapidly enough with time that the sound velocity averaged over long travel paths may not, in a practical sense, be knowable to the same accuracy as the individual measurements. Particularly in shallow coastal waters the effective accuracy may not be better than a part in 10^3, while in the open ocean, paths including near surface water may only be good to a part in 10^4. Good understanding of the local physical oceanography, leading to proper distribution of measurements and averaging can probably make some improvement on these numbers. Near-bottom deep water may often show microstructure at the level of a part in 10^5 (Spiess, 1980). In any event, every site at which acoustic measurements are made must include, at the same time, a determination of the sound velocity structure, including some evaluation of the magnitude and time scales of its variability.

One other aspect of the environment arises because water motion velocities are not usually negligible compared with the velocity of sound (1500 m/sec). For accuracies of a part in 10^5 one must know the along track component of water velocity to 1.5 cm/sec (0.03 knot). In this case, it is thus preferable to measure round trip travel time by using reflectors or echo repeaters, since, under those conditions, the ratio of water velocity to propagation velocity only enters as its square, and all realistic deep ocean currents become negligible (NAS, 1983).

Acoustic systems take on three forms, depending on the path lengths over which one works. For distances of a few hundred meters one should be able to achieve a part in 10^5 (few mm) accuracy using direct two-way transmission between fixed bottom units by operating in the 100 kHz (15 mm wavelength) regime. Over such short ranges it should be possible to have very good knowledge of the sound velocity structure along the path as well.

As path lengths approach a kilometer or more one is forced to somewhat lower frequencies by the fact that attenuation is less (Fisher and Simmons, 1977). One is also forced to abandon the direct path approach because, in nearly isothermal water (typical of deeper situations) the sound velocity increases with depth because of increasing pressure. Under these circumstances sound rays curve upward (Spiess, 1966) and, unless one places the acoustic elements on towers (with the added complexity of having to account for tower tilt), no direct path between transponders will exist. Under these circumstances an intermediate, near-bottom towed vehicle can be used to range simultaneously on three or more transponders from many different locations in the area being surveyed. A large number of observations are then used to determine the internal geometry of the transponder array, with the advantage of averaging over acoustic paths traversing differing portions of the area and thus averaging over the time-varying aspects of the sound velocity field. Computer simulations of this type of system, using sets of 300 observations distributed through a four transponder array having a 2 km radius resulted in baseline length errors of 1 to 2 cm when errors having a Gaussian distribution with 10 cm standard deviation were inserted into the range data (Spiess, 1985a).

The third type of acoustic system under development is designed to relate points on the seafloor to vehicles at the sea surface, with the goal of tying from the surface vehicle to points on land using GPS technology. Composite systems of this kind will support determination of baselines of lengths of hundreds of km, and first generation versions are expected to have uncertainties of the order of a few cm (Spiess, 1985b).

The difficulty of making the surface-to-bottom tie is that it must encompass the more rapidly changing complex uppermost layers of the ocean. As pointed out initially by Bender (1982), this effect can be mitigated if one uses a set of three seafloor transponders and operates close to the point for which all three travel times are equal. Under these conditions the uncertainty due to lack of knowledge of the sound velocity is proportional to the distance one is away from the central point (Spiess, 1985b). For example, if one can operate within 100 m of the center, one only needs the sound velocity to within a part in 10^4 to achieve centimeter accuracy. This simple picture is degraded by the existence of horizontal gradients of sound velocity across the region traversed by the sound paths. Fortunately, the only oceanographic effects that can produce gradients large enough to create problems in the open sea (away from major fronts) are those associated with the higher frequency internal waves (periods of tens of minutes to a few hours), and these can be reduced by averaging over periods of a day or so.

Conventional commercially available transponders capable of operating over km ranges operate by having a circuit that recognizes an incoming pulse and then transmits a pulse at some other frequency. This process, for systems operating in the 5 to 20 kHz regime, introduces timing uncertainties equivalent to a meter or more uncertainty in range (NAS, 1983). A method for overcoming this problem is to use some version of a signal re-transmitting system such that the phase relationships are maintained between the incoming and outgoing signals. One method that has been developed to implement such a system has been to build a transponder that contains a digital shift register delay line having a number of microsecond steps capable of holding several milliseconds of signal. A spread spectrum coded timing signal is then transmitted with the transponder interrogation. The transponder continuously digitizes the incoming acoustic energy and puts the samples into the delay line. Upon recognition of the interrogation, the transponder shifts from listening to transmitting and sends out the contents of the delay line. The outgoing acoustic waveform thus is delayed by an amount known to within a microsecond and the returning signal retains its phase relationship with the original signal (Spiess et al., 1980). Transponders of this type have been built and tested at sea in March of 1988, showing timing uncertainties having an individual pulse standard deviation of 10 μsec, corresponding to ranging uncertainties of less than 1 cm.

All of the various systems described or implied above have a common need with respect to capabilities of installing, tending, and removing complex equipment in the deep ocean. Such capabilities start with requirements on ships that must support any on-site activity, as well as transporting the necessary people and equipment to the area. These functions have been recognized and included among the ship performance requirements developed by the ocean science community through studies and committees, particularly those sponsored through UNOLS (University National Ocean Laboratory System) (UNOLS, 1986). Characteristics of particular importance in this context include good seakeeping, transfer of cumbersome heavy loads from the ship to the seafloor, adequate onboard storage space, and dynamic positioning. No individual ships in our present academic ocean research fleet meet all of these requirements, although two of them (Knorr and Melville) are scheduled for major upgrading during the coming year that will put them close to being adequate for these purposes. Beyond that, one new ship is programmed for construction starting this year with Navy funding and a second is in the preliminary design phase with NSF support.

Four types of systems are in various stages of development relative to performing the kinds of tasks involved in setting up and maintaining seafloor geodetic devices on the deep seafloor: manned submersibles, tethered neutrally buoyant swimming vehicles (conventional Remote Operated Vehicles-ROV's), cable supported dynamically positioned devices, and seafloor supported manipulative systems (tractors, etc.).

The only one of the four for which there is substantial deep-sea operating experience is the manned submersible category. Alvin, operated for the research community by Woods Hole Oceanographic Institution (WHOI) has been used for many years to carry out manipulative tasks, primarily in the contexts of marine geology and benthic biology (particularly hydrothermal vent-related experiments). The French submersible Nautile has recently been used to position a nearly neutrally buoyant winch assembly in a deep-sea drilling program borehole re-entry cone (Langseth and Spiess, 1987). These craft have the advantage of a man actually on the site, although that often is a disadvantage in that safety consideration places substantial limits on the kinds of objects one is willing to handle. Since these craft operate at close to neutral buoyancy, they have limited load handling capabilities. Overall, they can carry out some of the necessary tasks, but their deployment and safety considerations limit their usefulness.

ROV's have been in use in shallow water for many years, particularly as complements to diving operations in offshore oil and gas field development. These systems rely primarily on small, nearly neutrally buoyant vehicles positioned in the water by vertical and horizontal thrust propeller systems (Wernli, 1984). They have proven quite useful for both observation and manipulation. Since the actual work vehicle is coupled by a neutrally buoyant wire to the ship or to an intermediate, cable supported “garage”, these units are fairly effectively decoupled from the heaving motions of their surface support ship. Thus, they trade stability for load handling capability. Two such systems are emerging for deep water use by the ocean science community. One of these is the Argo-Jason system, (Ballard, 1982; Yoerger and Harris, 1986) under development at WHOI, with the other built by ISE (Langseth and Spiess, 1987) to support Canadian research efforts. Both of these are pushing a major technological advance-the use of fiber optic information transmission links built into the long main strain cables that support the “garage” unit from which the neutrally buoyant work vehicle operates. The high data rates expected from these telemetry systems will allow operation of full bandwidth television systems for viewing the vehicle's surroundings and the tasks they are carrying out.

Cable supported dynamically positioned systems have the capability of working with heavier loads albeit without as effective decoupling from the motion of the surface suspension point as the conventional ROV's. One system of this type has moved into operating condition for ocean research in the past year, developed with NSF support at Scripps

Institution of Oceanography (SIO). Its primary function is to place instruments accurately on the seafloor (including into drill holes) and either monitor their outputs (providing a telemetry link to the supporting surface ship) or release then for later recall or recovery. This device, being directly cable supported, can carry loads in excess of 1000 kg negative buoyancy (Spiess et al., 1987).

The fourth category-bottom-supported manipulative devices-seem best adapted to many of the tasks envisioned in the implementation of seafloor geodetic observations. Resting on the seafloor, such devices, although cable connected to a surface ship for power and control, can be almost completely decoupled from ship motion. At the same time they are able to manipulate heavy objects or carry out fine scale assembly operations using their reaction against the (relatively) solid seafloor. One device of this type is emerging from the development stage at SIO-RUM III (Anderson and Horn, 1984) is scheduled to carry out its first deep-sea tasks this summer, installing precision transponder mounts and seafloor hydrophone assemblies in basins off southern California.

Summarizing the technological situation, it appears that both a wide range of instruments, and the capabilities for installing and monitoring them, could be available for conduct of deep seafloor geodetic observations within the next year or two.

INITIAL OPERATIONS

Initial use of systems of these kinds present two requirements. They should address interesting geodynamic problems, and they should be at sites conveniently located relative to operating bases so that it is easy to make frequent visits to carry out developmental, as well as routine observational functions. As new terrestrial geodetic systems evolved, the tectonically active southern and central California areas provided an appropriate laboratory region meeting these requirements. Our present understanding (Minster and Jordan, 1984) implies that the offshore regions adjacent to the southern California coast are equally interesting and appropriate. It is clear that there are localized areas of tectonic activity in the Continental Borderland, and that, in order to understand fully the present day dynamics of the North American/Pacific Plate boundary, we must have deep ocean reference points beyond the Channel Islands.

Localized networks utilizing seafloor strain measurements and short range direct path acoustic systems could have their initial trials within the borderland itself. Some sites should be shallow enough that decreasing temperature with depth would counter increasing pressure to produce sound velocity gradients leading to downward, rather than upward, refraction. This would in turn allow fairly useful ranges for direct linkages between near-bottom points. In these regions, however, there would have to be intensive measurement of sound velocity to cope with changing oceanographic conditions.

Composite GPS/acoustic systems would play a most significant role by establishing reference points in the deep ocean, not only off the southern California borderland, but off the central California coast as well. Ties between these points and the well-established VLBI sites in northeastern California (e.g., Owens Valley) and in Arizona (Yuma) would provide the necessary constraints on the contemporary total motion between the Pacific and North American plates, as well as satisfying the requirements of being available for re-occupation with minimal ship operating cost.

For many of the approaches mentioned in the previous section, the most exciting and fruitful zones for initial application would be at the crests of the mid-ocean ridges and rises (NRC, 1988). Geodetic measurements of all kinds would provide very useful constraints on models of lithosphere formation, volcanic activity and hydrothermal circulation. The most logistically convenient, well-studied site for initial implementation of such studies would be on the East Pacific Rise at 21° north latitude (Normark, 1980; RISE Group, 1980). That location is, however, within waters controlled by Mexico and there have occasionally been delays in obtaining clearances to re-occupy the site. it may be possible to generate a continuous program there, similar to those set up on land by the seismologists, by helping develop Mexican scientific community interest in seafloor geodesy. The other two immediate choices are the Juan de Fuca spreading axis off the Canadian and northwestern U.S. coast and the East Pacific Rise at 13° N. If it were not for a restrictive weather window, the Juan de Fuca site would be a very desirable one; however, operations there with our normal research ships in support of development of new approaches would be limited to the summer months with resulting constraints on coordinating ship schedules with other research activities. The 13° N site's disadvantage is its distance (about 1500 miles) from the nearest major oceanographic research ship operating base. Other than that, it is an attractive location since it has a stretch of well-developed simple structure including hydrothermal activity, but with examples of overlapping rift zone features (Sempere, 1986) nearby. The pros and cons of these locations will be debated over the coming year and, hopefully, the beginnings of geodetic site occupations will take place early in the 1990s.

PROGRAMMATIC CONSIDERATIONS

Successful geodetic programs have one essential requirement-a long-term commitment by the participants-both individual leaders and, even more important, sponsoring agencies. In the U.S. the two groups that have maintained the necessary commitment over the years on land have been the U.S. Geological Survey and the National Geodetic Survey (surviving the multi-step transformation into a NOAA element from the Coast and Geodetic Survey). Neither of these groups has been particularly aggressive in attempting to develop an oceanic capability.

This lack of push probably arises because applications of marine geodesy do not fit logically into their missions as developed in the terrestrial environment. The one exception is in the USGS mission of understanding earthquake phenomena. Here, there would be clear gains in having reference points on the oceanic side of the San Andreas complex, but even more exciting is the prospect of being able to measure convergence across the Aleutian Trench or uplift along the Vancouver Island/Washington continental slopes.

The only agency that has shown consistent support for development of oceanic capabilities has been the NASA Geodynamics program. It recognized a number of years ago (Walter, 1983) that space-based techniques (particularly GPS) could have a more nearly global capability if it were possible to make ties from continental sites into ocean areas. This extension has largely taken place by occupation of island locations. While these are particularly convenient for VLBI and satellite laser ranging operations, the program also recognized the desirability of being able to establish reference points in important places where no islands were available, most obviously on the seaward flanks of the major trenches on the ocean margins. It is hoped that this program will continue, at least to a point at which USGS and/or NOAA will pick up the challenge.

The other logical agency that might support oceanic geodetic activity is the National Science Foundation. Geodesy in general has had scant support from that quarter, in part because geodetic impacts on basic earth sciences have not been well established until rather recently. The gradual development of our awareness of the many facets of plate tectonics, and the fact that nearly all plate boundaries lie in the ocean, give considerable emphasis to the desirability of including research in ocean floor geodesy within the NSF Ocean Sciences purview. A problem in this is that, while de facto NSF provides continuous support for various aspects of ocean science, the general pattern for administering its ocean floor research grants (except deep-sea drilling) has been based on short-term commitments. This pattern is, however, being somewhat altered and some portion of its new initiatives are targeted for coordinated research activities focused on longer term goals. Within this context, the RIDGE initiative (NRC, 1988) does provide a logical context for ocean geodesy as a part of ocean science research.

CONCLUSION AND ACKNOWLEDGEMENTS

Although there are many steps yet to be taken along the way, it appears that, by the year 2000, there may very well be the technology and the programmatic commitment for significant seafloor geodetic activities. The research and application opportunities are there both in relation to our basic understanding of the crust of the earth beneath the sea, and the context of related impact on man's activities.

The technology discussed in Section II is quite diverse and its development has been carried out by a larger number of contributors than even the list of references implies. The author's geodesy-oriented activities have been supported primarily by NASA, but with small inputs from the Office of Naval Research, NOAA, and NSF.

REFERENCES

Agnew, D. C., Continuous Measurement of Crustal Deformation, in Methods of Experimental Physics, edited by C. G. Sammis and T. L. Henyey, Vol. 24, Part B, Geophysics: Field Measurements, pp. 409–435, Academic Press, 1987.
Anderson, V. C., and R. C. Horn, Remote Underwater Manipulator: RUM III, Trans. Soc. Automotive Engineers, San Diego, 1984.
Ballard, R. D., Argo and Jason, Oceanus, 25, 30–35, 1982.
Bender, P. L., National Bureau of Standards, University of Colorado, Boulder, personal communication, 1982.
Busse, D. W., Quartz Transducers for Precision Under Pressure, Mechanical Engineering, 109, No. 5, 1987.
Eckart, C., Principles and Applications of Underwater Sound, Reprinted by Dept. of the Navy, Headquarters Naval Materiel Command, Washington, D. C., 1968.
Fisher, F. H., and V. F. Simmons, Sound Absorption in Sea Water, Acoust. Soc. Am. 62(3), 558–564, 1977.
Greenspan, M., Acoustic Properties of Liquids, in American Institute of Physics Handbook, McGraw-Hill, New York, 1972.
Irish, J. D., and F. E. Snodgrass, Instruments and Methods-Quartz Crystals as Multipurpose Oceanographic Sensors-I Pressure, Deep-Sea Research, 165–169, 1971.
Langseth, M. G., and F. N. Spiess, Science Opportunities Created by Wireline Re-Entry of Deep-Sea Boreholes, Workshop held at Scripps Institution of Oceanography, 23–24 February 1987, Joint Oceanographic Institutions as part of the U.S. Science Support Program for the Ocean Drilling Program, 1987.
Lovett, J. R., Merged Seawater Sound-Speed Equations, J. Acoust. Soc. Amer., 63, 1713, 1978.
McIntyre, M., and D. E. Boegeman, A New Sound Velocity Measurement System, Proceedings of Intl. Symposium on Marine Positioning, Reston, Virginia, 1986.
Minster, J. B., and T. H. Jordan, Vector Constraints on Quaternary Deformation of the Western United States East and West of the San Andreas Fault, in Tectonics and Sedimentation Along the California Margin edited by J. K. Crouch and S. B. Bachman, p. 187, Pacific Section of Soc. of Economic Paleontology and Mineralogy, 1984.

NAS Committee on Geodesy, Seafloor Referenced Positioning: Needs and Opportunities, Panel on Ocean Bottom Positioning of the National Research Council's Committee on Geodesy, National Academy Press, Washington, D. C., 1983.

NRC, The Mid-Ocean Ridge: A Dynamic Global System, p. 352, National Academic Press, Washington, D. C., 1988.

Normark, W. R., Definition of the Plate Boundary Along the East Pacific Rise Off Mexico, Marine Geodesy, 4, 29–43, 1980.

Reid, J. L., personal communication, 1984.

RISE Group, East Pacific Rise: Hot Springs and Geophysical Experiments , Science, 207, 1421–1433, 1980.

Saunders, P. M., Practical Conversion of Pressure to Depth, Amer. Meteorological Soc., 11, 573–574, 1981.

Sempere, J., Occurrence and Evolution of Overlapping Spreading Centers , Ph.D. Thesis, University of California, Santa Barbara, CA, p. 227, 1986.

Spiess, F. N. Underwater Acoustic Positioning: Applications, Proceedings of the 1st Marine Geodesy Symposium, 93–101, 1966.

Spiess, F. N., Acoustic Techniques for Marine Geodesy, Marine Geodesy, 3(1), 13–27, Crane, Russak & Co., Inc., New York, 1980.

Spiess, F. N., C. D. Lowenstein, D. E. Boegeman, and F. V. Pavlicek, Precision Transponder and Method for Communication Therewith, U.S. Patent 4, 214, 314, 1980.

Spiess, F. N., Analysis of a Possible Sea Floor Strain Measurement System, Marine Geodesy, 9(4), 385–398, 1985a.

Spiess, F. N., Suboceanic Geodetic Measurements, IEEE Trans. on Geoscience and Remote Sensing, GE-23(4), 502–510, 1985b.

Spiess, F. N., J. A. Hildebrand, and D. E. Boegeman, New Systems for Seafloor Studies, EOS, 1335, 1987.

Stanley, E. M., The Refraction Index of Seawater as a Function of Temperature, Pressure and Two Wavelengths, Deep-Sea Research, 18(8), 833–840, 1971.

UNOLS Fleet Replacement Committee, Science Mission Requirements for New Oceanographic Research Ships, 1986.

Urick, R. J., Principles of Underwater Sound, McGraw-Hill, New York, N. Y., 1975.

Walter, L. S., Geodynamics, NASA Conference Publication 2325, 1983.
Watts, D. R., and H. Kontoyiannis, Deep-Ocean Bottom Pressure and Temperature Sensors Report: Methods and Data, University of Rhode Island, Graduate School of Oceanography, Narragansett, R. I., GSO Technical Report Number 86-8, 1986.
Wearn, Jr., R. B., and N. G. Larson, Measurements of the Sensitivities and Drift of Digiquartz Pressure Sensors, Deep-Sea Research, 29, No. 1, 111–134, 1984.
Wernli, R. L., Operational Guidelines for Remotely Operated Vehicles , p. 200, Marine Tech. Soc., 1984.
Yoerger, D., and S. E. Harris, Argo Jason: Integrated Capabilities for Exploration of the Sea Floor, AUVS, 1986.

The Accuracy of Position Determinations by VLBI Expected by the Year 2000

Alan E. E. Rogers
MIT Haystack Observatory Westford, Massachusetts 01886

INTRODUCTION

Very-Long-Baseline Interferometry (VLBI) was first used for geodetic measurements in 1972 following the development of VLBI in the late 1960s for the astronomical study of radio sources whose angular size is a few millarcseconds or less and whose structure cannot be resolved with conventional connected element interferometers. In geodetic VLBI the relative arrival time (see Figure 1) of signals from distant quasars is measured by recording digitized samples on magnetic tape for cross-correlation at a central processor. A hydrogen maser frequency standard provides a precise frequency reference for the receiver's local oscillators as well as a time standard for the digitization and data formatting as illustrated in Figure 2. A delay precision which greatly exceeds the inverse recorded bandwidth is achieved by recording 14 2 MHz bandwidth channels from S- and X-band in a fashion illustrated in Figure 3. The method, which is known as “bandwidth-synthesis”, (Rogers, 1970) allows a very wide bandwidth to be spanned with acceptable delay sidelobes and ambiguity spacing provided enough channels are available with minimum redundancy in spacings. The S- and X-band channels are used to provide separate group delays whose appropriate linear combination provides a delay almost free of the ionospheric path delay. A complete description of the MKIII VLBI system has been given by Clark et al., (1985).

Present Level of Accuracy. For the past eight years geodetic measurements made using the MKIII have achieved centimeter level repeatability in the measurement of baseline length. A statistical analysis of the repeatability of 100 baselines measured by the NASA Crustal Dynamics Project has been made by Clark, (1987). The length repeatability from this analysis is approximately given by 5 mm plus 2×10^{-9} times the length and Table 1 lists the repeatability of some of the more frequently measured baselines. The fractional term is mostly the result of an r.m.s. error in the vertical of about 2 cm which maps into length in an amount which is proportional to length. Table 2 gives a list of the various error sources which will be discussed in more detail later. The root sum square (RSS) is approximately in agreement with present experience.

System Noise. The signal to noise ratio (SNR) is given by where A is the effective antenna area (or geometric mean for antennas of different sizes), F is the radio source flux, 2BT is number of data bits recorded in integration time T, T_3 is the system temperature, K is Boltzman's constant, and L is a digital processing loss factor of about 0.6. The standard deviation in the estimate of group delay is equal to the inverse of the product of SNR and r.m.s. spanned bandwidth (measured in radians/sec) and is equal to 80 ps (2.4 cm) at X-band for 10-meter antennas of 50 percent efficiency and 60-K system temperatures observing a 1 Jansky (10^{-26} Wm^{-2} Hz^{-1}) radio source with 200 seconds of integration at the normal 56 Mbit/s data recording rate of the MKIII. Improved receivers using High Electron Mobility Transistors (HEMT) are expected to reduce system temperature to 30 deg K and improvements in recorder technology already support a 512 Mbit/s rate. While the system noise is not a major error source, its contribution is important in experiments with weak geometry such as very long baselines for which the mutual visibility is poor or experiments with small mobile antennas.

Instrumental Calibration. Drifts in the receiver and cable delays are calibrated by injecting pulses (about 30 ps duration) into the receiver at a 1 MHz rate. These pulses generate a low-intensity rail of signals spaced 1 MHz apart which, when mixed to video, are made to appear as 10-KHz tones in the upper sideband passbands by offsetting the local oscillators. Extraction of the phases of these tones in the data processing serves to calibrate the the phase delay of each 2-MHz-wide passband. At present some systems are limited by the presence of spurious signals from the digital electronics which corrupts the phases of the calibration tones. This limit will be reduced to a negligible level by better isolation of the I.F. electronics. The electrical length of the cable which feeds the calibrator is electronically measured to allow correction for changes due to temperature or flexure.

Other Instrumental Errors. Antenna flexure and changes in the axis intersection with tower expansion are significant error sources in some of the larger antennas used as base stations although these errors can be accurately modelled. A more subtle instrumental error source exists in present systems which is due to the variation of group delay in the antenna feed with position angle. Since the calibration signals are injected after the feed, these errors go uncorrected. These errors are the result of frequency dependent polarization impurities in the present feed and can be isolated by examining the "closure" of delays measured on a triangle of baselines.

Atmospheric Limits. As can be seen from the table, the atmosphere is thought to be the largest error source. There are two distinctly different methods currently employed to correct for the atmospheric path delays. The first method is largely one of "self-atmospheric calibration" in which the unmodelled "wet" atmospheric zenith delay after correction of the data for the delay through the "dry" troposphere is derived by least square analysis every few hours or for each observation using a Kalman filter. In this method it is important to include observations at low elevations to separate the atmospheric delay signature from the signature of a station height as illustrated in Figure 4. In the second method the atmospheric delay for both dry and wet components is calibrated by barometric pressure and water vapor radiometry (WVR), respectively. The first or self-calibration method is limited by errors in the dry atmospheric "mapping function" (Davis et al., 1985), or elevation dependence which depends on the atmospheric lapse rate and height of the troposphere as well as the pressure and temperature at ground level. The second or direct calibration method is also limited by a knowledge of the mapping function although not so severely since low angle observations are not required and can be avoided. In addition, the direct calibration is limited by the accuracy of the ground pressure measurement needed to derive the dry atmosphere zenith path and any breakdown in the assumption of hydrostatic equilibrium used to derive the path length from ground pressure. Further, the derivation of the wet delay from the WVR measurements (Resch, 1984) is subject to errors due to the departure of the atmosphere from the model assumed for the estimate of the "retrieval" coefficients. Better site dependent algorithms for the conversion of WVR data to wet path delay are now being developed. However, because there may be a bias in the WVR determined wet delay or a bias in the dry delay which results from barometer calibration errors or other unknown cause, the direct method is usually combined with the estimation of a bias for each experiment. At the present state of the art, the two methods are producing almost the same level of accuracy or marginally better results with the second method and the latest generation of WVRs capable of measuring the wet delay to better than 1 cm as inferred from the repeatability of results. Both methods assume azimuthal symmetry for the dry delay and hence, are equally limited by the presence of horizontal gradients. Perhaps the real promise for future improvements will come with an understanding of the biases between WVR determined path delays and the wet path delay solutions using VLBI data which may well be errors in the model for the dry delay. In addition, the incorporation of a horizontal gradient and mapping function parameters in the self-calibration may improve results if there is enough "geometric strength" in the data to allow for the estimation of additional parameters which are highly correlated with existing parameters.

The Very Long Baseline Array (VLBA). The National Radio Astronomy Observatory (NRAO) under contract to the National Science Foundation (NSF) is building an array of ten new 25-meter antennas to form a network of stations dedicated to VLBI. The antennas which span the United States, with antenna in Hawaii and another in the Virgin Islands (see Table 3), will be equipped with the new generation of VLBI electronics and recorders (which support a data rate up to 512 Mbit/s). The VLBA will be used for both astronomy and geophysics. Geodetic VLBI methods will be used to determine the station locations, earth rotation, polar motion, earth tides, tectonic plate motion, ocean loading, other geophysical parameters needed to provide the highest accuracy astrometry which might, for example, be used to search for perturbations which might indicate the presence of planets in orbit around nearby stars. The ten VLBA antennas combined with an effective aperture of 79 meters or a single VLBA antenna using 256 Mbit/s recording could operate with an ultracompact transportable VLBI system using an antenna as small as one meter and reducing the cost of making high accuracy geodetic ties to remote locations.

SUMMARY

At the present time atmospheric calibration is the dominant error source and will probably remain as the ultimate limit. Instrumental errors and errors due to source structure are likely to be reduced to a negligible level in the next decade. It is likely that the fractional accuracy of baseline length measurements will improve to 1×10^{-9} or perhaps 5×10^{-10}.

ACKNOWLEDGEMENT

Geodetic VLBI research at the Haystack Observatory is supported by the National Aeronautics and Space Administration under Contract NAS-5-29120. Haystack Observatory has designed and prototyped data acquisition and recording systems for the NRAO VLBA under Contract AUI-216.

REFERENCES

Clark, T. A., et al., Precision Geodesy the MKIII Very-Long-Baseline Interferometer System, IEEE Trans. Geosci. Remote Sens., GE-23, 438–449, 1985.

Clark, T. A., Presented at NASA's Crustal Dynamics Project Users Meeting, October 1987.

Davis, J. L., et al., Geodesy by Radio Interferometry: Effects of Atmospheric Modeling Errors on Estimates of Baseline Length, Radio Sci. Vol. 20, pp. 1593–1607, 1985.

Resch, G. M., Water Vapor Radiometry in Geodetic Application in Geodetic Refraction, edited by F. K. Brunner, Springer-Verlag, New York, 1984.

Rogers, A. E. E., Very-Long-Baseline Interferometry with Large Bandwidth for Phase-Delay Measurements, Radio Sci., Vol. 5, pp. 1239–1248, 1970.

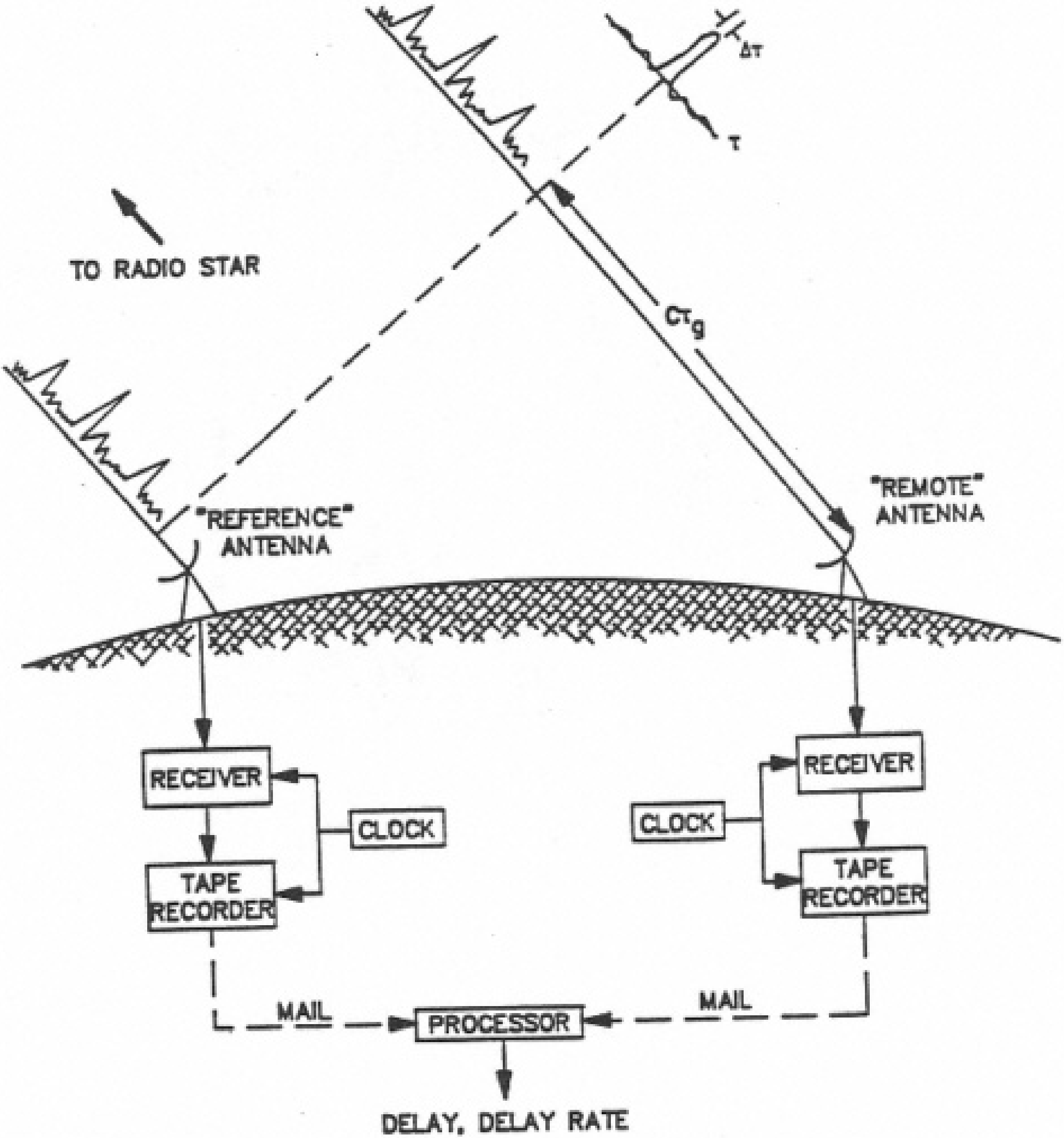


Figure 1.
Geodetic VLBI

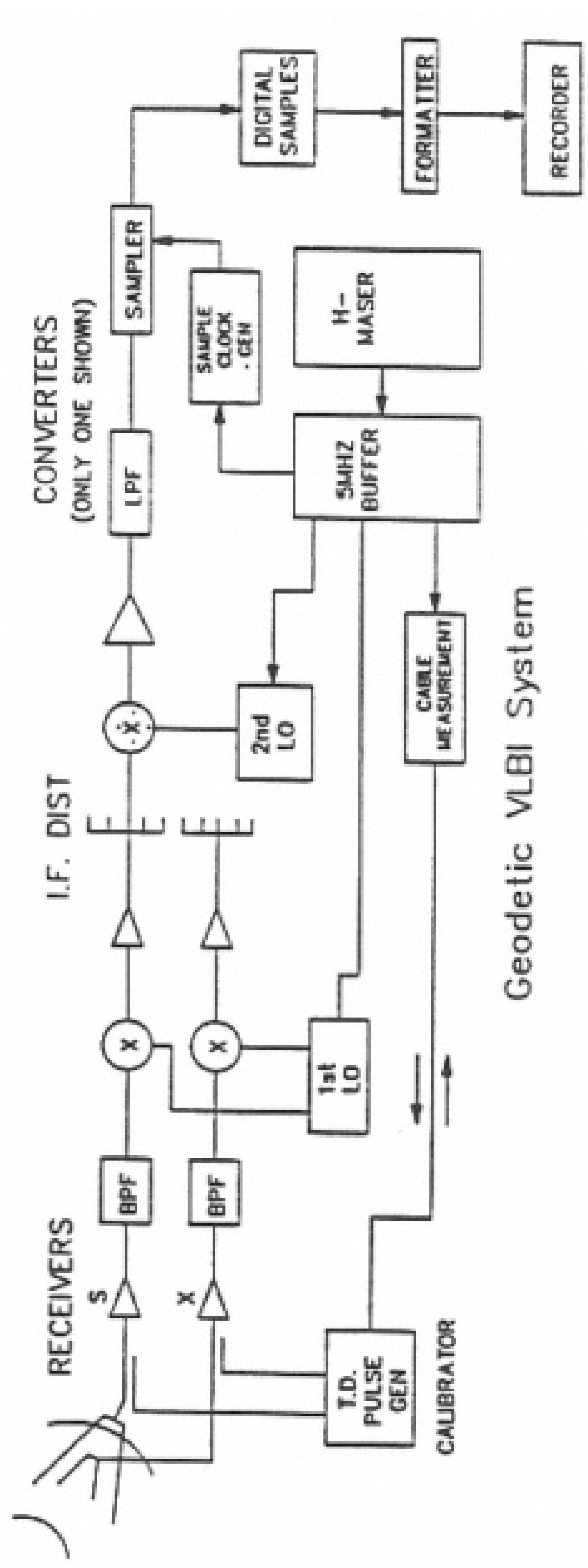


Figure 2.
Block Diagram of Geodetic VLBI System

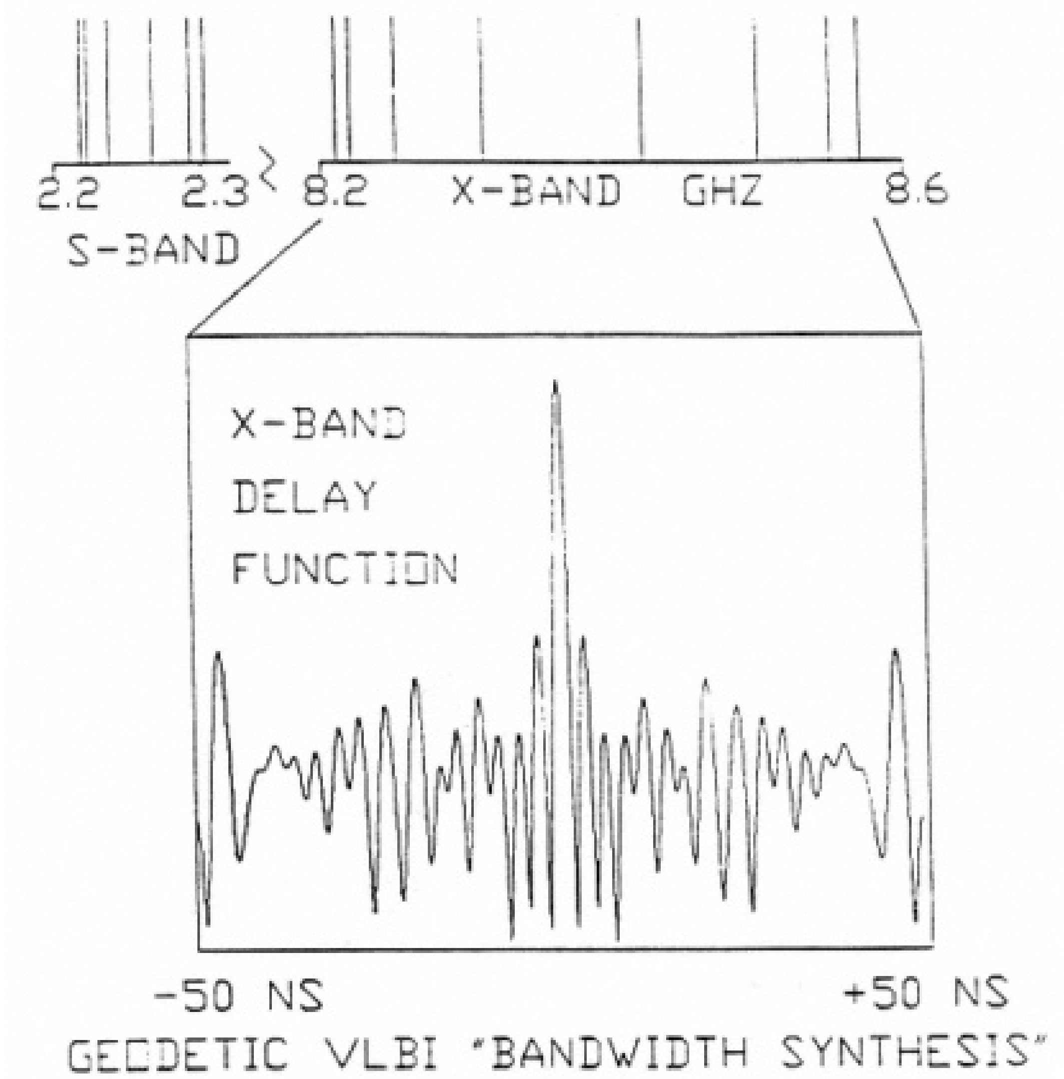


Figure 3.
S/X Frequencies Recorded for "Bandwidth Synthesis"

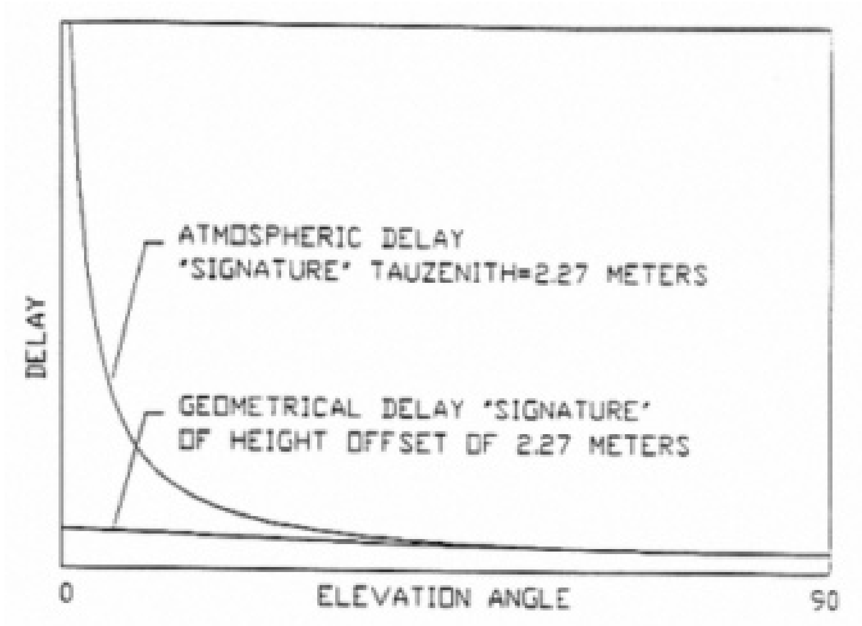


Figure 4.
Atmospheric Delay Signature

Demonstrated VLBI Baseline Repeatability

Baseline	Length km	RMS Repeatability Length mm	RMS Repeatability Transverse mm
Mojave - OVRO	245	7	6
Mojave - Monument Peak	271	9	8
Mojave - Vandenberg	351	11	12
Onsala - Wettzell	920	5	-
OVRO - Ft. Davis	1508	9	8
Mojave - Westford	3904	8	-
Haystack - Onsala	5600	15	-

Present Level of Accuracy
≈ 5mm plus 2×10^{-9} times the length

Table 1. Demonstrated VLBI Baseline Repeatability

ERROR SOURCES	ERROR EST.	COMMENTS
Instrumentation: Instr. noise(SNR≈ 70)	≈ 5 mm	Lower system temperatures and larger recording rates will improve
Antenna flexure	<5 mm in vertical	important for large antennas - can be modelled
Antenna tower expansion	≈ 5 mm	can be modelled
Spurious signals/cal errors	≈ 5 mm	can be improved
Polarization impurity	≈ 2 mm	new feeds will improve
Clock instability	≈ 2 mm	(1×10^{-14} over 10 min.)
Ionosphere	≈ 1 mm	S/X corrects almost perfectly
Source structure	≈ 2 mm	can be corrected - VLBA will make better maps
Atmosphere: dry troposphere	≈ 10mm in vertical	needs better pressure calibration (hydrostatic equib. errors?)
mapping function (or wet troposphere)	≈ 10mm in vertical	estimate additional parameters (improved WVR algorimthms)
horizontal gradients	≈ 2 mm	estimate with Kalman filter
RSS	≈ 20mm in vertical ≈ 8mm transverse	may be degraded by weak geometry

Table 2. Error Sources for Geodetic VLBI

LOCATION	N LATITUDE [deg,min,sec]	W LONGITUDE [deg,min,sec]	ELEVATION [m;MSL]	COMMENTS
Pie Town, NM	34 18 03.61	108 07 07.24	2371	First VLBI in 87
Kitt Peak, AZ	31 57 22.39	111 36 42.26	1916	antenna complete
Los Alamos, NM	35 46 30.33	106 14 42.01	1967	complete in 88
N. Liberty, IA	41 46 17.03	91 34 26.35	241	complete in 88
Brewster, WA	48 07 52.80	119 40 55.34	255	complete in 89
Fort Davis, TX	30.63	103.94	1615	complete in 89
St. Croix, VI	17.75	64.60	15	being acquired
Owens Valley, CA	37.23	118.28	1235	being acquired
Mauna Loa, HI	19.54	155.61	3200	negotiating site
"Northeast"				negotiating site

Table 3. VLBA Site Locations and Status Nov 87

GPS-Based Geodesy Tracking: Technology Prospects for the Year 2000

W. G. Melbourne
Jet Propulsion Laboratory California Institute of Technology Pasadena, California 91109

INTRODUCTION

The twenty-first century is a scant dozen years away, too near for many landmark advances in technology, yet long enough for a few surprises to emerge. To put it into focus one need only look backwards 12 years to the way things were in 1976, the year of our bicentennial. The Navstar/Global Positioning System (GPS) program had already been underway for two years. Meanwhile, we are still awaiting the launch of the first Block II satellite, which will mark the onset of the GPS operational phase. Of course, history is laden with inaccurate predictions of future technology trends based on current knowledge. My message is really ambivalent. Don't expect too much progress in the decade of the 1990s; on the other hand, do expect some surprises because technology often advances in quantum jumps.

In the case of GPS, the evidence today unmistakably suggests future trends for high accuracy geodesy and Earth satellite tracking. Given the current performance of GPS, particularly on regional baselines, one can say that GPS has arrived. Its attributes of high accuracy and relatively low cost make it the geodetic technique of choice in the 1990s for most of the regional deformation studies that require high temporal and spatial resolutions. It has even greater potential for the future. I would like to discuss this potential in terms of:

a. future accuracies of GPS geodetic systems,
b. future costs for data acquisition equipment, operations and data analysis, and
c. future applications in ground and satellite geodesy.

PROSPECTS FOR MILLIMETER ACCURACIES IN GPS-BASED GEODESY

What are the limiting error sources today for GPS systems and what are the prospects for their improvement? I think that we should take our cues on these questions from experience with VLBI. The horizontal accuracy from Mark VLBI III systems ranges from around 1/2 – 1 cm on regional baselines to 1 – 1-1/2 cm on transcontinental baselines (Clark et al., 1987; see chapter by Rogers) the vertical accuracy is about a factor of 2–3 worse because of tropospheric water vapor errors and limitations of observational geometry, i.e., one cannot observe below one's horizon. Roughly the same ratio between horizontal and vertical appears to hold in GPS accuracies. However, there is some suggestion that the more robust observation scenario with GPS systems, in which GPS satellite and ground receiver clock instabilities are isolated to prevent possible corruption of baseline estimates, yields better accuracies in the vertical component than are obtained with VLBI.

Today's evidence (Blewitt et al., 1988) supports the position that GPS horizontal accuracies are comparable to VLBI accuracies on regional baselines. For example, Figure 1 shows the agreement between GPS and VLBI-determined baselines over a range of lengths up to 2000 km for several measurement campaigns spanning 1985–1988. Centimeter-level agreement is the norm for regional baselines. On transcontinental baselines, GPS systems are not quite as accurate as VLBI because of current limitations in the control of GPS reference system errors, which should be remedied in the near future with stronger GPS fiducial and/or global tracking networks.

The principal errors limiting GPS performance today arise from:

a. mis-modeling and/or mis-calibrating the propagation media,
b. antenna multipath,
c. antenna phase center variability, and
d. reference system errors including the terrestrial reference frame, limitations in alignment with VLBI and SLR frames, and the GPS satellite ephemerides.

I will now summarize the present status of these error sources and try to estimate their future course in the 1990s. To reach the millimeter level will also require dealing with error sources that are presently masked by these “tall poles” cited above. For example, seasonal environmental effects such as ground water variability can certainly affect the height of monuments at the millimeter level in certain soil conditions. Ocean loading is another. GPS satellite multipath will have to be dealt with on longer baselines. I will mainly focus on the current limitations and the prospects for improving them.

PROPAGATION MEDIA

Tropospheric water vapor is the major error source for VLBI and GPS systems. Small errors can arise from mis-modeling or mis-calibrating the troposphere and the ionosphere.

Dry Troposphere. The total zenith delay of the dry troposphere is about 2 m and is readily determined barometrically to a precision of about 1 mm using standard atmosphere models. Departures from hydrostatic equilibrium of the dry component of the troposphere leads to zenith delay mis-modeling that can amount to a few millimeters in certain dynamical conditions. For baselines longer than the correlation length of these departures from hydrostatic equilibrium, differential delay errors will translate into comparable errors in the horizontal components and into larger errors in the vertical component. Bender (1987) has discussed the potential error arising from the dry component of the troposphere. To reach 1 mm accuracy it may be necessary to utilize synoptic meteorological data under some conditions; improved mapping functions may be required to transfer from the zenith to line-of-sight delays.

Ionosphere. There are small departures from the standard dual band plasma frequency correction for group delay and phase advance through the ionosphere. Also, for equatorial and high latitude observations, particularly those made during very high solar activity, ionospheric scintillations can be troublesome. The rapid phase variations and concomitant signal amplitude in the L-band carriers induces tracking errors in the tracking loops of some receivers and even loss of lock; on the other hand, some receivers of more recent design are capable of tracking signal frequency accelerations of up to several g's, even in adverse signal conditions, without significant error at the millimeter level. Code-tracking receivers definitely have a big advantage over codeless receivers in weakened and highly variable signal conditions. Consequently, this aspect of the ionospheric problem is likely to improve greatly, even with the high solar activities expected in the early 1990s, with the advanced receiver designs expected by then. The other aspect is the departure from the dual frequency correction resulting from ionospheric density gradients and from the small corrections in ionospheric refractivity due to the presence of the geomagnetic field. These effects, which depend on the third and higher degrees in carrier frequency, can amount to several millimeters at low elevations during high ionospheric activity. Another error contribution arises from the residual LCP component of the GPS signal which is about a factor of four below the RCP component. The propagation velocities of the RCP and LCP waves differ slightly in a plasma containing a magnetic field. The rejection of the LCP component by GPS antennas is not complete, typically 10–20 db lower (in power). As a result errors relative to the dual band correction up to a few millimeters can result. Additional modeling based on the dual frequency correction may be required to reach the millimeter level (Clynch and Renfro, 1982).

Water Vapor. Water vapor zenith delays range from 5 cm in dry and cold conditions, upwards to 50 cm in tropical conditions. Water vapor is not in hydrostatic equilibrium. As a result, in situ meteorological measurements to infer total water vapor delay can be in error by as much as 100% of the actual delay; they have fallen into disrepute. Currently, stochastic modeling and water vapor radiometry-based calibration are the only viable techniques for dealing with water vapor. Using the best technology currently available, mis-modeling and/or mis-calibrating the differential water vapor delay at the geodetic sites results in expected horizontal baseline errors that range from a few millimeters to perhaps a centimeter depending on the strategies used, on the local meteorological conditions, and on the length of the baseline (Elgered et al., 1988). At these error levels the correlation length for water vapor is well below 100 km for many regions (Treuhaft and Lanyi, 1987). For the vertical component these error levels should be increased by a factor of three. State-of-the-art water vapor radiometers have a stability over several hours of about 1K in measuring sky temperature (Jannsen, 1985). This translates into an accuracy in zenith delay of about 0.5 cm. Short period stability can be somewhat better but this is less important for baseline accuracy than long period stability. Assuming the WVR error can be characterized as a random walk process with an expected error accumulation of 0.5 cm after four hours of tracking, one obtains about 3 mm as the expected error in the horizontal baseline components.

A meteorologically benign region, such as the relatively dry western U.S. or even a tropical area with a uniform marine layer, would be characterized by low temporal and spatial variability in water vapor delay. For these regions, an equally successful approach is to stochastically model the tropospheric zenith delay as a piecewise constant random walk or as a first order Gauss-Markov process. The values of the constants in this time series are adjusted in a batch-sequential least squares process in which both the transition from one value to the next and the length of time over which the vertical delay is held fixed are a priori constrained by the stochastic model. By judicious choice of the values of the parameters in the stochastic model, one can obtain baseline repeatabilities that are competitive with those obtained with WVR calibrations (Herring et al., 1989; Tralli et al., 1988). Whether this approach will prevail in less benign tropospheric conditions with high spatial and temporal gradients such as those sometimes prevailing at Kokee Park, Kauai is undetermined.

A general consensus remains elusive on strategies for dealing with the water vapor problem toward achieving 1 mm horizontal baseline accuracy. A major reason for this is that the variability of water vapor delay, both spatially and temporally, is not well understood as a function of locale, season and current meteorological conditions. Of particular concern is the spectral power of the variability at periods of roughly a half hour and longer; these are the components that can partially mimic the signatures of the geodetic parameters in the tracking data and thus corrupt their estimates. Continuously operating arrays of GPS receivers, to be discussed later, offer some promise in developing a data base that should enable one to quantify the spectral properties of water vapor variability and to develop strategies for exploiting this knowledge.

Pushing the WVR-based methodology to yield 1 mm horizontal baseline accuracies will be a major technological challenge. To achieve this, the current dual channel WVR will have to be expanded to several channels and the sky temperature measurement accuracies improved to 0.1K, a tough requirement. Current WVR system temperatures are several hundred Kelvins. Future WVR's using HEMT technology will probably achieve around 100K, thus, the precision requirement will be around 0.1%. However, the accuracy requirement of 0.1K is more stringent requiring corresponding instrumental and calibration stability over several hours. Further, the retrieval algorithms for inversion of the measured sky temperatures to the water vapor-induced path delay, which involve similar but different integral equations along the line-of-sight, will have to be greatly improved over current algorithms; they are likely to become highly dependent on the region and the local meteorology. Also, mapping functions to map WVR line-of-sight delays to the local vertical or to the GPS directions will need further development.

To achieve millimeter-level performance by WVR calibration is going to be very expensive in the short term. Current WVR's cost around $125K and that price is likely to go up unless major strides are made in solid state digital RF systems; that is an ambitious undertaking today at these relatively high frequencies (20–60 GHz). However, if present development trends persist in microwave monolithic integrated circuits (MMIC technology), the mid-1990s may see virtually all-digital WVR's including digital sampling at these RF frequencies enclosed in a highly stable oven that occupies a volume of less than 10^6 cm; in addition to costing an order of magnitude less than current designs, it would also promise an order of magnitude improvement in instrumental stability. Using a phased array antenna for sky coverage would also improve performance and reduce cost. Nevertheless, the cost trend for the WVR is likely to move opposite to that of the GPS receiver technology in the early 1990s; the current factor of two in their relative unit cost is likely to grow to ten. Unless their costs can be reduced by an order of magnitude, WVR systems will not be affordable in large numbers and would be concentrated mainly at VLBI fiducial and other sites where the highest accuracy is sought.

While it may be theoretically possible for GPS-based geodetic systems to achieve millimeter horizontal baseline accuracies with major capital outlays, as a practical matter we may be mostly limited to accuracies attainable with the stochastic modeling approach. Stochastic modeling will reach an impasse at some accuracy level. The short period ($t<1/2$ hour) spectral components will effectively increase the carrier phase measurement noise; the longer period components can correlate with baseline offsets. These limiting accuracies are really unknown at this time because other error sources have masked the inherent accuracy of the stochastic approach and there are additional refinements to it that are yet to be made. I am optimistic, however, that water vapor-induced baseline errors can be driven well below 5 mm for many applications and that 1 mm can be approached, given sufficient averaging time. The question still unanswered is: what is the minimum averaging time to reach 1 mm, or alternatively, what is the expected baseline accuracy versus averaging time due to mis-modeled water vapor errors? This may be the limit on resolving short period geophysical signatures in the tracking data.

ANTENNA MULTIPATH

Antenna multipath refers to the perturbation of the received signal due to the presence of reflective objects in the vicinity of the antenna. These spurious signals are characterized by additional delay and usually smaller amplitude relative to the direct signal, and they contain phase shifts. They combine interferometrically with the main signal to cause phase and amplitude variations; these are processed by the receiver and manifest themselves in the measurements as carrier phase and group delay errors. Thus, both the carrier phase and the pseudorange measurements are corrupted by multipath, although in fundamentally different ways. Their manifestation also depends on the signal processing architecture of the receiver.

Carrier Phase Multipath. Carrier phase multipath depends on the signal processing scheme used to recover the carrier. For receivers using a half chip early/late gate correlation approach multipath can arise from objects up to distances of 450 m for C/A-based recovery and 45 m for P-code-based recovery. Figure 2 gives an example of multipath in carrier phase measurements. Here, the plots are the post-fit smoothed residuals of double-differenced carrier phase measurements of two GPS satellites from two ground receivers. Ionospheric effects have been eliminated with dual band measurements. The residuals are shown for two separate diurnal passes (an arbitrary vertical offset and the siderial period adjustment in the abscissa are included for illustrative convenience). The ground receivers were TI-4100s with the standard TI-4100 antenna. The RMS variation about zero mean is around 1-1/2 cm. Notice the high degree of diurnal repeatability; the cross correlation function peaks strongly for a zero time lag.

Multipath in the carrier phase measurements will certainly have to be dealt with if 1 mm baseline accuracies are sought. It is the long period spectral components of multipath variability, hence, reflections from nearby objects, i.e., roughly 1/2 hour and longer, for which one needs to be concerned. The shorter period components will tend to average out over several hours of observations and will have little corrupting influence on the baseline estimates. The high degree of diurnal repeatability in multipath suggests that a major fraction of it could be calibrated out by building up an empirical template from a number of successive passes.[1] But one would be vulnerable to

1. The Navstar/GPS Joint Program Office has recently announced that it is considering raising the GPS satellite altitudes about 50 km, which would increase their period by two (2) minutes. This would cause the ground tracks to drift westward 1 degree per day; thus, exact daily repeatability would no longer hold and would be replaced with an annual repetition. Near exact repeatability might occur at intermediate intervals but it would involve different satellites. Although this adjustment breaks an undesirable resonance and results in improved station keeping; however, it would be unfortunate for those wishing to calibrate diurnal multipath effects.

environmental changes around the antenna site during the intervening periods between revisitations with a GPS receiver; these could lead to different multipath signatures and hence, to multipath-induced baseline errors that are different on the two visitations. Thus, short-term baseline repeatability over a few days might be an optimistic measure of long-term repeatability spanning several years of site revisitations. These considerations also lead one to require standardization and configuration control of antenna/backplane assemblies over a span of years because of internal multipath in these systems. Counting on common mode error cancellations is a dangerous game when the outcome depends on the stability of two big numbers that are differenced!

Pseudorange Multipath Errors. Multipath error in the pseudorange depends on the dispersive character of the multipath effects across the power spectrum of the pseudorange PRN code. Pseudorange measurements are made by matching local models of the ranging code generated at the receiver with codes being received from the GPS satellite. In general, the received signal is a combination of direct and reflected signals with different delays, phases and amplitudes. The effect of this combined input is to distort the shape of the correlation function between the model and received signals. This shape distortion maps into error in determining the point of maximum correlation and this mapping function will depend on the particular correlation scheme used by the receiver. As an example, for a half chip early/late gate scheme, a peak error of ~5 ns occurs when a multipath signal is present with an amplitude of 0.1 and an additional delay of ~0.5 chip (1 P-code chip = 97 ns) relative to the direct signal. Because the effective "wavelength" of P-code pseudorange is two orders of magnitude greater than the carrier wavelengths, the pseudorange multipath error is typically a factor of 100 greater than carrier phase multipath error. An important exception is the case where the reflecting objects are within 1 carrier wavelength of the receiving antenna. In this case, the carrier phase multipath is reduced by less than order of magnitude relative to pseudorange multipath.

Figure 3 illustrates P-code pseudorange multipath effects for a typical environment using a TI-4100 antenna. The integration time for the points in this figure is two minutes; thus, thermal noise is an insignificant contributor to these residuals. Successive daily plots will attest to these residuals being primarily multipath-induced. RMS levels for P-code pseudorange typically are in the 1–2 m range for TI-4100 measurements; in some cases they are worse, even causing loss of lock; in some cases they are better but rarely, if ever, are they below 50 cm.

Multipath errors need not be anywhere near as high as those in Figure 3. With proper care in preparing a site with low multipath environment and with use of advanced antennas and, most importantly, backplane designs, multipath can be reduced by over an order of magnitude. Figure 4 shows the multipath level in a recent experiment using a Rogue receiver which has a Dorne & Margolin antenna coupled with a choke ring backplane designed at JPL (see insert on Figure 4). These results are from a single receiver tracking a single GPS satellite whose elevation with time is also shown. Here, the residuals are plotted in the form of two-minute points, the nominal output rate of the Rogue receiver, and also as half-hour running average points. The ordinate on this plot is the linear combination of P_1, L_1 and L_2 that eliminates both the ionospheric and range delays, where P_1 is the L1 pseudorange and L_1 and L_2 are the carrier phase measurements, all expressed in range units. This linear combination should be constant with time; variations are measures of multipath, which for this combination is strongly dominated by that on P_1. Also shown are two bands pertaining to performance specifications that we have placed on our antenna/backplane development effort. These specifications can in part be described by the requirement: multipath spectral power shall not cause P-code pseudorange residuals to exceed 30 cm for all periods of five minutes or less, nor exceed 5 cm for all periods of 30 minutes or longer. The latter condition is the most stringent one and the most relevant to baseline recovery accuracy. Except for very low elevations this test amply satisfied our long period multipath requirement. Future innovations on this design are likely to further improve these results. The particular test shown in Figure 4 demonstrates two points: (1) dramatic improvements in multipath can be achieved, so that smoothed P-code pseudorange accuracies below 5 cm appear feasible, and (2) high frequency multipath, which tends to originate from more distant objects, can be averaged down to relatively low levels. The very high repeatability of these residuals on a daily basis (with peak cross-correlation values well over 90% for this particular experiment) also suggests an approach for calibrating multipath effects in a stable environment.

The Utility of Pseudorange. A brief digression on the value of high accuracy pseudorange would seem to be in order. Most GPS practitioners make little use of the P-code pseudorange for high accuracy applications. The reasons for this have been, (a) its relatively coarse (meter-level) accuracy, primarily because of multipath and partly because the architecture of the TI-4100 receiver, the standard of the 1980s, through its multiplexing and baseband processing design, resulted in effectively SNR-limited P-code performance, and (b) the uncertain future availability of the P-code on the Block II satellites through possible actuation of the anti-spoofing (AS) function. Receiver manufacturers have exploited that uncertainty by developing systems for the civilian market that avoid use of the P-code, using instead the C/A code to recover L_1 and various nonlinear detection schemes (e.g., Costas loops, signal squaring, delay and multiply, code enhanced squaring) to recover the second harmonic of L2 and in some cases P_1, P_2, L_1–L_2 and P_1–P_2. Codeless recovery pays a price in the form of higher SNR-driven measurement errors, which is particularly significant for the P-code, and higher thresholds for signal detection and tracking the L2 carrier. However, receivers of the late 1980s and the 1990s are and will achieve nearly an order of magnitude improvement in SNR performance through elimination of multiplexing, use of the P-code, use of double sideband in-phase and quadrature processing, lower system temperatures, and elimination of interchannel biases through all-digital processing. Moreover, the average received signal strength (C/N_o) of the P-code modulation from the Block II satellites may be about 3 db higher.

The DoD has progressively relaxed and/or clarified its position on the operation of AS over the last several years; its current announced position is that, except in times of national emergency, AS will be off except for periods of test and training which would be scheduled relatively infrequently and announced in advance. The DoD policy on Selective Availability (SA) is another matter to be discussed later; but SA does not normally render inoperative the P-code tracking functions of receivers without decryption capability. However, receivers without decryption capability (i.e., without the so-called W-chip and the crypto-keyed information), regardless of whether they are P-code trackers or not, would not be able to recover the corrections to SA in real-time.

The discussion earlier suggests that 5 cm differential P-code pseudorange accuracies or better on 30-minute averages will be attainable for receiver and antenna/backplane systems that are well-designed and well-sited, and that it will be largely available to civilian users. Of what use is this capability given the already very high accuracy of carrier phase data which would easily be accurate to sub-millimeter levels if it were not limited by multipath and the troposphere? A number of applications come to mind.

1. Tracking GPS satellites with dual band carrier phase and 5 cm pseudorange would yield highly accurate global ephemerides obtained with far less tracking per pass than is required today. As a rough estimate, 5 cm is about 1 part in 10^8 on a baseline between two tracking stations 5000 km apart. If the effect of the satellite multipath on baselines of this length can be controlled or calibrated to 5 cm or better, one would expect to obtain ephemerides from a global network of GPS tracking stations that surpass 20 cm in accuracy, a factor of 5–10 better than today.

2. Carrier phase data that are phase connected, i.e., that have no cycle dropouts, provide range change information to nearly perfect accuracy but cannot measure absolute range unless the cycle ambiguities are resolved. Pseudorange measures absolute range but cannot compete with carrier phase in measuring range change. This suggests a synergistic approach (Hatch, 1982). By combining these data types one can map a whole time series of pseudorange observations to a common epoch with nearly perfect accuracy using the continuous and phase connected carrier phase observations. Thus, all of the pseudorange observations made over a span of hours can be brought to bear on a single time point and their errors thereby averaged down. For that epoch, geometric positioning is accomplished with an effective ranging accuracy that should be far better than the accuracy of the individual

pseudorange points. This quasi-geometric point positioning approach will have major applicability to tracking from dynamical platforms such as ocean platforms and buoys, aircraft and spacecraft, and in kinematic ground tracking. It will also be used for high accuracy orbit determination of future Earth satellite missions such as TOPEX/Poseidon which will carry a GPS flight-rated receiver (Yunck et al., 1985). This approach obviates the need for extensive dynamical models to generate the motion of the user platform.

3. It is well known that accurate pseudorange improves the accuracy of baseline determinations, particularly the east-west component (for most regions except very high latitudes) (Melbourne, 1985). Coarse pseudorange accuracy, e.g., 100 m, provides sufficient accuracy in clock synchronization so that negligible error is introduced in the ephemeris look-up epoch. From that level down to 1 m little improvement is realized. However, for pseudorange accuracy below a carrier wavelength, baseline accuracy typically improves by factors of 1 to 5, depending on the component and the situation. This is because with carrier phase data alone, the baseline estimation process must include the carrier cycle ambiguities or biases which can typically be estimated to an accuracy of about 1/2 carrier wavelength or better. In those situations where it is not possible to exploit the integer character of the cycle ambiguities and hence to fix their values, introducing pseudorange with an accuracy that is comparable to or better than the a posteriori cycle ambiguity accuracy adds significant information bearing on their values; it aids in breaking the high correlations that tend to prevail between the biases and the eastern component of the baseline. Figure 5 gives an example of the effect of introducing pseudorange into the estimation process with increasing accuracy ranging from no pseudorange to carrier range, the latter being the case where the cycle ambiguities are exactly known.

4. Pseudorange can also greatly assist in resolving the carrier cycle ambiguities. Cycle ambiguity resolution strategies without highly accurate pseudorange have had moderate success for well designed networks that contain a range of different baseline lengths. These techniques have been applied to networks up to about 1000 km in extent (Blewitt, 1989; Dong and Bock, 1989). But these schemes usually have not been able to correctly fix 100% of the cycle ambiguity integers in any given measurement session mainly because of ionospheric delay and ephemeris errors. Aiding the ambiguity resolution process with accurate pseudorange would virtually assure a 100% success rate at all baseline lengths. It can be shown (Melbourne, 1982) that the cycle ambiguity integers for L1 and L2 are given by the expressions:

$$(1) \quad n_1\lambda_1 = 4.091*P_1 - 3.091*P_2 - L_1$$

$$(2) \quad n_2\lambda_2 = 5.091*P_1 - 4.091*P_2 - L_2$$

where L and P are the carrier phase (modulo 2π) and P-code pseudorange measurements, respectively, all expressed in range units; and λ is the carrier wavelength. Here, n, P and L may be considered as residuals, and/or as singly or doubly differenced.

In the absence of measurement errors the left hand side of these equations are constants with time for a series of phase-connected carrier phase observations and P-code measurements. This suggests reducing the effect of measurement errors by averaging this system over an entire pass to obtain an ensemble mean for (n_1,n_2); short period measurement errors (including multipath) will average out, but long period effects will likely remain. If the long period effects are sufficiently low, the ensemble mean should lie close to an integer doublet and hence can be fixed. However, it can be shown that for a greater than 95% confidence level of picking the correct integer values for (n_1,n_2), the standard deviation of the ensemble mean must be less than 1/4 in each component; inspection of the above system of equations reveals that this is rather ambitious, requiring that the residual systematic pseudorange errors and averaged down multipath errors must in aggregate be no larger than about 1 cm. Nevertheless, Figure 4 suggests that the Rogue receiver/antenna system came close to achieving this goal in that particular experiment. The ordinate of Figure 4 is merely the linear combination of Equations (1) and (2) that eliminates P_2. There exists a similar plot for the combination eliminating P_1 and it shows comparable performance. One can infer from these two plots that the uncertainties in the ensemble means for the double differenced versions of n_1 and n_2 (for white Gaussian distributed errors) would be around 1/3 for this experiment, close but not quite close enough.

Short of an all-out direct assault on the cycle ambiguity problem as described above, to what other use could accurate pseudorange be made? Even if the pseudorange measurement accuracy were not quite sufficient to reliably resolve ambiguities, it certainly would provide a narrow bound on their possible integer values. This information should improve the efficiency of current carrier cycle ambiguity resolution techniques. Also, most techniques first try to fix the value of the difference, n_1–n_2, the so-called wide lane cycle integer. This corresponds to a wavelength of 86 cm and is therefore relatively easier to resolve. Without P_1 and P_2 the uncertainty in the differential ionospheric delay limits the reliability of this fixing to baselines or a maximum length that depends on solar activity, time of day, elevation angles of the GPS satellites, and geographical location. About 200–300 km appears to be a typical upper limit for campaigns that have been conducted in the continental U.S., Mexico and Central America. For receivers that do not use the P-code to recover L2, but rather, alternative nonlinear techniques that recover the second harmonic of L2, a wide lane ambiguity length of only 43 cm results, which is twice as vulnerable to ionospheric mis-modeling. With the P-code pseudorange the wide lane ambiguity can be fixed with very high reliability quite independent of baseline length (Blewitt, 1989). The above cycle ambiguity equations can be combined to yield

$$(3) \qquad n_1 - n_2 = 0.00652*P_1 - 0.00509*P_2 - L_1 + L_2$$

where the units of P and L are now in centimeters. These are very small coefficients on P_1 and P_2; the 95% confidence level requirement for correctly fixing $n_1–n_2$ is easily met by the measurements in the experiment shown in Figure 4. Even a pass of pseudorange from the TI-4100 usually can be averaged to satisfy the requirement. Having fixed $n_1–n_2$, one can use Equation (3) as an a priori constraint in estimating the individual integer ambiguities from the carrier phase observations (Blewitt, 1989).

Finally, it should be noted that the variations in Figure 4 are almost entirely due to multipath. Hence, if the antenna/backplane physical and electrical configuration is held invariant and the environment remains unchanged, one should obtain nearly identical residuals the next day but four minutes earlier because of the sidereal period of the daily exact repeating ground tracks of the GPS satellites. This means that whatever the values of the cycle integers that were chosen today, they are going to be the same tomorrow. Thus, the near repeatability of the multipath signatures allows one to connect all of the cycle ambiguities together over successive daily passes, thereby greatly reducing the number of degrees of freedom in a multi-day estimation solution. This technique would allow one to phase connect two carrier tracking segments interrupted by an extended outage using the information obtained from neighboring days.

If cycle ambiguities can be resolved on continental-sized baselines through a combination pseudorange and carrier phase-only techniques, it would result in a marked improvement in tracking accuracy for GPS ephemeris production. One would have carrier ranging effectively accurate to the limits of tropospheric mis-modeling, that is, sub-centimeter accuracy. The limiting error source in this case probably will be the mis-modeling of the phase center variation of the GPS satellite antenna array. This varies with boresight angle at the satellite and the latter ranges from zero up to a maximum of 14.3 degrees. The differential effect is baseline length dependent and negligible for regional baselines. By tracking the GPS satellite from rise to set using a high-gain steerable antenna (thereby minimizing ground multipath), one can obtain from the pseudorange and carrier phase measurements an estimate of multipath effects originating at the satellite. Recent tests have placed an upper limit of 5 cm variation over a several-hour pass (Young et al., 1985).

5. Quite apart from accuracy considerations there are a number of operational and data processing benefits to be accrued from receivers with pseudorange that should be evident from our previous discussion. The temporal invariance of the left hand sides of Equations (1) and (2), including invariance to clock instabilities and also to SA-induced dithering, suggests a very efficient technique for data editing, e.g., detecting and removing cycle dropouts (Lindqwister et al., 1988), and for real-time recovery from loss of lock by the receiver without loss of carrier phase connection. The latter will be quite useful in kinematic

surveying applications for dealing with signal loss resulting from obstructions, high accelerations and jerk. Real-time data compression by the receiver is another application afforded by accurate pseudorange. This has been practiced for some time by the USAF GPS Monitor Stations under the name, "Carrier-aided Pseudorange". Greatly shortened tracking sessions will achieve the same system accuracy as the current much longer tracking scenarios.

ANTENNA PHASE CENTER VARIABILITY

Antenna phase center variability arises from internal reflections and diffractions within the antenna/backplane assembly that shift the received phase as a function of signal direction. Errors arising from this effect should really be a non-problem at present system accuracies; unfortunately, it currently appears to be one. Unless rigorous standards and configuration controls are agreed to by the community, it is almost certain to preclude achievement of 1 mm baseline accuracies in the future. The phase shift from an omni antenna typically varies at the 1–2 cm level with wavefront direction (and is therefore indistinguishable from carrier multipath). This variability differs depending on the antenna design, the antenna and backplane physical and electrical configuration, and the local electrical and physical environment (Sims, 1985; Tranquilla, 1986). The mean phase center obtained by averaging over the solid angle above some elevation cutoff, e.g., 10 deg, and chosen to minimize the RMS variations, is one accepted definition of the phase center. The actual mean phase center, which results from averaging over a daily tracking session and which therefore depends somewhat on the tracking scenario and on the configuration of the GPS constellation, may differ from the conventional definition. Mis-modeling phase center variability can lead to errors in baseline estimates. The high accuracy attained to date with GPS geodetic systems is largely due to common mode error cancellation: antennas of the same manufacture and configuration have very similar phase center properties and thus their differences are slight. Using receivers with different antennas is an invitation to trouble in the form of significantly degraded repeatabilities (at the one centimeter level or higher) in baseline solution comparisons made over successive experiments, even though intra-experiment repeatability might be substantially better. Moreover, even with antennas with identical phase center variability with azimuth and elevation, one is likely to incur a baseline length dependent error owing to the differing direction of a GPS satellite as viewed at the ends of the baseline. There are obvious calibration experiments to tie different systems together and to convert from one scenario to another that one could perform in addition to standardization wherever practicable.

The strategies for defining and maintaining a terrestrial reference frame follow two rather different approaches. Simply put, one approach uses the fiducial concept (Davidson et al., 1985; Bock et al., 1986; Beutler et al, 1986; Lichten and Border, 1987) and the other uses instead the dynamical framework of the GPS constellation.

The Fiducial Network. The fiducial concept uses strong a priori information on the location of certain fiducial sites spanning a network of GPS stations. The a priori data are baseline vectors between GPS sites that are collocated with VLBI sites, augmented with geocentric coordinates obtained from collocations of VLBI and SLR. By processing concurrent tracking data from the fiducial and the geodetic sites and incorporating the a priori site coordinates, one effectively aligns the GPS orbits with a reference frame defined by these fiducial points and thereby transfers the fiducial information via the GPS ephemerides to the geodetic sites within the network. There are variations on this approach, mainly concerning how the ephemerides are generated and controlled, and how their information is transferred to the geodetic sites; but strong a priori fiducial information from VLBI and SLR is the hallmark of the fiducial approach.

Systematic errors in the fiducial approach arise from:

(a) errors in the alignment of the GPS receiver phase center with the intersection of axes of the VLBI telescope,
(b) errors in the VLBI/SLR-derived fiducial coordinates, and
(c) errors in the observations and modeling used in the GPS ephemeris generation process.

Alignment errors arise from the tie of the GPS phase center to a nearby benchmark and from the benchmark to the intersection of axes of the telescope. With a common environmental configuration and with identical equipment design and manufacture at each end of the baseline, the error in tying the antenna phase center to a benchmark largely cancels in baseline results (probably completely at the centimeter level but probably not at the millimeter level); but those arising from surveys between neighboring benchmarks and the intersection of axes, particularly of large antennas, have had a troublesome history over the past four years. Even today, residual errors of 1 cm are likely to be lurking in these ties; reducing this to the millimeter level will be a major undertaking.

The repeatability of the lengths of about 150 baselines determined by VLBI over the past four years can be expressed as

$$\text{Repeatability} = (5 + 3 * B * 10^{-3}) \text{ mm}$$

where B is baseline length in kilometers (Clark, 1988). This suggests that the accuracies of relative fiducial coordinates obtained from VLBI observations on regional scales are probably sub-centimeter in the horizontal components and probably 1–2 cm on transcontinental baselines with robust observation sets and with accompanying Earth orientation observations good to the milliarcsecond level. Improving these system accuracies to the millimeter level on regional baselines will require significant advances in dealing with the water vapor problem and in the development and deployment of VLBI data acquisition systems that span a wider bandwidth for group delay measurements than the Mark III system. Doubling the current ~400MHz spanned bandwidth should reliably yield carrier phase delays at S and X-band. That would give VLBI about an order of magnitude improvement in precision over the current Mark III system in measuring delay (although not that much improvement in system accuracy). Thus, GPS reference frame definition and maintenance using the fiducial approach on regional applications is likely to be paced by system improvements in VLBI.

In addition to VLBI-derived site coordinates, there are errors arising from aligning this fiducial frame with a geocentric frame derived from SLR. At current accuracy levels this is a minor error source for regional applications but troublesome when comparing results from different groups. Standardization will be welcome in the area of reference frames. Other error sources arise from mis-modeling Earth platform processes such as ocean loading, Earth tides and Earth orientation. These currently cause errors in horizontal baseline estimates that are probably at the level of a few millimeters; they will need to be refined to achieve millimeter accuracy.

GPS ephemeris errors for well-modeled data analysis systems appear to be somewhat data noise-limited in regional fiducial applications, particularly for GPS satellites with poor viewing geometry. It was just noted that the VLBI-derived horizontal components of the baselines used in fiducial control (typical lengths of 1000–3000 km) are probably accurate to better than 1 part in 10^8. That should translate into decimeter-level ephemeris accuracies. However, the best ephemerides produced to date, based on multi-day tracking arcs, stochastic modeling of non-gravitational accelerations and the tropospheric variability, and other sophisticated refinements, appear to have accuracies, when averaged over the entire orbit, that range between 0.5 m and 2 m or 3–10 parts in 10^8. Ephemerides based on regional tracking obviously will be less accurate over parts of the orbit not tracked so this is not really a fair comparison. Accuracies over the region tracked appear to be a factor of three to ten better depending on the strength of the viewing geometry (Wu et al., 1988). This would place the regional ephemeris accuracy in the range of 1–2 parts in 10^8 (Lichten and Bertiger, 1989), perhaps a factor of two larger than one would expect with perfect tracking accuracies. This is probably largely due to the use of carrier phase data without resolution of cycle ambiguities, which causes a weakening of orbit accuracy analogous to that experienced in the eastern baseline component. In fact, covariance studies with ambiguity resolved carrier phase measurements show that data errors in this case are less significant and, in fact, the terrestrial reference frame then limits ephemeris accuracies. Ephemeris errors resulting from a priori fiducial location errors tend to be magnified by an order of magnitude, particularly over regions outside of the fiducial network. As mentioned earlier, the use of very accurate pseudorange in addition to carrier phase should significantly improve ephemeris accuracies either by improving the a posteriori values of the cycle ambiguities, or by directly facilitating carrier cycle ambiguity resolution, or by tying cycle ambiguities together from day-to-day. In addition, the current success at resolving cycle ambiguities by optimal network design (Blewitt, 1989; Dong and Bock, 1989) also augurs well for success on transcontinental fiducial networks; but networks spanning ocean basins may be a bit more challenging.

The Global Framework. With the full deployment of the Block II constellation in the early 1990s one is likely to see an alternative approach to GPS reference frame maintenance, which uses the dynamical framework provided by the GPS constellation. For lack of a better name I have dubbed it the Global Framework. A dynamical approach to reference system maintenance does not, strictly speaking, require fiducial information. Instead, it relies on the dynamical consistence of the GPS constellation for the reference frame as embodied in the GPS ephemerides, in the set of derived station coordinates of a globally distributed GPS tracking system, in the Earth orientation parameters obtained from a global monitoring system, and in the transformations between the conventional inertial system and the conventional terrestrial system.

Why would the Global Framework be preferred by some over the fiducial approach? It would have some disadvantages, notably its reference system would not be under the total control of individual users. To make it work an international organization would be required to maintain and operate it. That in itself will be challenging. One reason to prefer it is its potential stability and amenability to conventional standards for its configuration and operation, and for the distribution and timeliness of its data products. Another is its potential for enhanced accuracies, less so over regions spanned by robust fiducial networks; but substantial increase in accuracy would be realized over regions that have sparse fiducial coverage such as third world and oceanic regions (Freymueller and Golombek, 1988). Global tracking would strongly control ephemeris error build-up over continental areas that may be small enough to permit cycle ambiguity resolution on these long baselines. Another compelling reason will be the cost advantages to its users if the reference system can be maintained and the tracking system operated through international agreements and funding.

Operationally, this dynamical approach is similar to the way the USAF maintains the operational GPS orbits and obtains the broadcast ephemerides (and the way LAGEOS orbits are maintained by SLR). However, we are talking here about something rather more ambitious. The ephemerides, the tracking station locations, and the relevant Earth orientation information are generated from continuous tracking, both carrier phase and high accuracy pseudorange, by a set of globally distributed tracking stations. These stations are closely enough spaced to have ample satellite mutual visibility for eliminating clock errors, something the current USAF network would have trouble with (see Figure 6). About a dozen stations would suffice if they are roughly uniformly distributed. With assumptions about tracking and modeling accuracies comparable to those described herein, and with realistic modeling of non-gravitational accelerations and antenna phase center variations on the satellites, covariance studies predict that such a system would yield GPS ephemeris accuracies after two days of tracking that are below 1 decimeter and station location accuracies of 1–2 cm, comparable to that of VLBI and SLR (Melbourne et al., 1988). Thus, the GPS tracking system would be capable of determining its own station locations and need not rely on the a priori information supplied by VLBI and SLR. The system would also solve for the offset of its reference frame from the geocentre and it would determine its own scale. Figure 7 shows that the geocentric offset is determined to better than 5 cm accuracy after only 1 day of tracking (Wu and Malla, 1988). In short, this system appears capable of providing on a continuous basis a reference frame whose accuracy may well approach 1 part in 10^9 by the mid-1990s.

A wild card for the Global Framework will be the advent of the low Earth orbiter (LEO) bearing a high accuracy GPS receiver that is concurrently tracking. Two examples are TOPEX/Poseidon in 1992 and the Eos platforms planned for the 1990s. GPS tracking from the LEO not only enables its positioning, it also provides an additional GPS tracking source for the ground network. The LEO mediates the transfer of tracking information between ground stations by simultaneously observing GPS satellites that are not co-visible from the ground stations, thereby expanding the range of tracking intervals over which concurrent observations (and hence clock cancellation) can be made. Also, the LEO to ground station baseline has on average a large "vertical" component and is corrupted by tropospheric errors at only one end of the baseline (see Figure 8). These result in improved accuracy of the vertical components of the baselines within the ground network by a factor of two in certain cases and it reduces the sensitivity of the vertical components of the ground baselines to mis-modeled tropospheric delay.

In its extreme form the only non-GPS observational information that the Global Framework would need is the longitude of a single GPS tracking station and UTI to tie down the degree of freedom associated with an unobservable rotation of the combined GPS constellation and Earth system. This is exactly how SLR works. Polar motion could be detected through the diurnal rotation of the Earth. Obviously, it would use conventional standards such as the nutation series and the constant of precession. For user convenience, UTC would be supplied and a current tectonic plate motion model could be included as part of the reference system. Also, one should collocate the GPS tracking stations at VLBI sites for operational and logistical reasons including use of their ultra-stable frequency standards and timing. (Using a linear model for clock variability instead of a white noise model improves the data noise component of baseline estimate accuracies by upwards to 40%.) Earth orientation information from an Earth monitoring system such as the International Earth Rotation Service (IERS) should be incorporated to maintain alignment with the Conventional Terrestrial System; and it would be foolish not to use the a priori site location information provided from the VLBI and SLR systems to minimize the effects of systematic errors. But this information need not dominate the solution set for station locations as it does in the fiducial approach. In fact, there is not compelling reason for 100% congruency between the GPS and VLBI/SLR frames. We can and no doubt will live with constant offsets between these frames without undermining plate tectonic and crustal deformation results. Offsets that vary with time will be the troublesome ones that should be minimized.

Operating in the SA Environment. As a postscript, I would like to add a few words about the prospects of operating high accuracy geodetic systems in the Block II era when the Selective Availability will be on. The DoD's current plan is to have SA operating on all Block II satellites except perhaps four, which would be used for time transfer functions. The levels at which SA might be operating and the technical details of its implementation will be found in a classified document issued by the USAF NAVSTAR/GPS Joint Program Office (Anon., 1989). SA alters the broadcast ephemeris, the clock correction relative to GPS time, and the P-code clock frequency (10.23 MHz) in such a way to degrade point positioning (and velocity) accuracies of unauthorized users of the P-code to about 100 m 2 drms, about an order of magnitude worse than its nominal accuracy. For authorized users, that is, those having access to the Precise Positioning Service (PPS), their receivers would be equipped with a decryption capability that would enable real-time corrections for SA and the ability to track the P-code when the AS function is on.

For high accuracy users involved in non-real time relative positioning, such as the geodetic community, the effects of SA are entirely different. In the first place, these users will be insensitive to errors in the broadcast ephemerides and clocks because they are not used for high accuracy non-real time data processing. SA dithers GPS P-code clock frequency in a pseudo random fashion but the clock epoch varies within certain bounds as prescribed by the 100 m 2 drms point position accuracy requirement. This means that the effects of SA clock frequency dithering manifest themselves as GPS satellite clock epoch errors and, potentially, as a mean error in the length scale when the carrier phase and pseudorange measurements are converted from units of light-sec to centimeters. A mean scale error could arise if the mean clock frequency, averaged over the entire observation set in both time and satellites observed, differed significantly from 10.23 MHz. (The L1 and L2 carrier frequencies are coherent with the P-code clock frequency and hence a variation on the latter produces identical proportional variations on the former.) The length scale error, which causes a baseline error that is proportional to length, is likely (but not certainly) to be negligible for most geodetic applications. For very noisy frequency standards aboard the satellite there could be small effects arising from the finite speed of light, which leads to the so-called retarded baseline effect. During the delay in arrival of a common wavefront at the two ends of the baseline ($\leq$20 msec), frequency instabilities of 0.001 ppm can lead to millimeter-sized effects; but these levels are very unlikely in nominal conditions. The clock epoch errors can be isolated for all GPS satellites co-visible from two or more ground receivers; they would be treated just like clock errors are treated today and would have no effect on baseline recovery accuracies.

Another aspect of SA is its potential impact on the carrier tracking integrity of all receivers not operating with a PPS capability and on subsequent editing and data compression procedures. Qualitatively, SA frequency dithering appears to the user as if the GPS satellite had a very noisy frequency standard driving the P-code clock frequency. This means that frequency dithering could limit the tracking loop bandwidth or the length of time over which carrier tracking can be integrated without potential loss of lock. This is unlikely to be a problem in most applications and in most conditions under which SA would be operating; but users should familiarize themselves with (Anon., 1989) to assure themselves. Obviously, receiver architectures using the P-code to recover the L2 carrier will have better threshold characteristics than codeless designs. For post-processing such as data editing and smoothing, SA would limit the maximum tracking sample interval without risking carrier cycle dropouts if a single-station/single-satellite data processing strategy were followed.

One could recover the clock errors, if desired. The fiducial sites are particularly suited for this; they frequently can provide H-maser timing to the GPS receiver; also, the fiducial baseline vectors are known a priori and thus, their solutions in the estimation process should be essentially invariant to any excursions in mean clock frequency. For example, one could treat the SA frequency dithering as a piecewise constant stochastic process and the clock epochs as a pseudorandom walk process. Using standard data processing operations on the tracking data information matrices such as double differencing, Householder or Givens transformations, triangularization, etc., the clock and frequency parameters could thereby be isolated from the geodetic and GPS ephemeris parameters and solved individually for their estimates relative to a master clock and frequency standard at a fiducial site. Thus, one could recover, at least in some lumped form, the clock epoch variations that result from SA or from any other sources of clock instability. From SNR and onboard oscillator stability considerations, one concludes that it is not feasible to recover the clock epoch state frequently enough to determine the actual SA-induced clock frequency transitions, which, in any case, would be treading on dangerous ground security-wise if it were done. Even the lumped values, say over several minutes to hours, might be sensitive if publicly released without an appropriate delay. It should be clear, however, that any agency, U.S. or foreign, could set up a fiducial network coupled with a geostationary communications satellite that would disseminate the SA corrections and improved ephemeris and clock information to its regional users in virtual real-time. Such a regional differential real-time positioning system would readily provide meter-level navigation accuracy.

To summarize, what are the prospects for GPS system accuracy in the 1–3 mm range by the end of the 1990s? My opinion is that they are very bright. If history provides any precedent, we should take a lesson from it. The system accuracies of VLBI and SLR have improved at a rate of one order of magnitude per decade for the past 20 years since the Williamstown Conference (Kaula, 1969), which first galvanized NASA into creating major development and scientific programs for these techniques. Besides their historical records, there are a number of innovations on the horizon for these systems that should perpetuate this trend in accuracy. For GPS, these innovations include:

a. high performance receivers and antenna/backplanes for sub-millimeter instrumental accuracy in carrier phase and centimeter-level accuracy in smoothed pseudorange (including multipath), high performance all-digital WVRs capable of millimeter accuracy in zenith delay.
b. improved tropospheric modelling including a better understanding of the spectral content of its variability, better calibration, and better understanding of site and seasonal dependence.
c. improved modeling of environmental factors that may cause variability in ground monument location, e.g., seasonal variability of the ground water table,
d. improved calibration of GPS satellite multipath,
e. more robust and accurate reference systems including a global tracking system for timely dissemination of reference frame data products as well as the tracking data,
f. conventional standardization of procedures and of the configuration of data acquisition and information systems, and
g. use of low Earth orbiters flying GPS receivers in conjunction with ground arrays to reduce the sensitivity of baseline estimates to tropospheric mis-modeling, to improve the vertical components, and to enhance the temporal resolution of baseline variability.

My view is that GPS system accuracy will probably continue to match future VLBI performance, particularly in regional applications, and that it will do so at considerably less cost.

Table 1 provides a rather speculative and personal view of where GPS system accuracies might lie at the end of the century. Error budgets are provided in two columns: one for today and one for a decade hence. These figures represent the final error contribution (1-sigma) of each line item to the horizontal baseline components, after all calibrations, modeling and estimation procedures have been applied using state-of-the-art procedures. If present trends continue, I would expect the vertical component to be about three times worse, but the LEO satellites carrying GPS receivers might enable one who uses this tracking information to reduce it to only a factor of two. It should be noted that an error contribution from pseudorange has not been directly included; I have assumed that its role is to facilitate cycle ambiguity resolution. One should discriminate between short-term baseline error arising from observation sessions ranging from minutes to hours and to that attainable with long-term averaging over days; this has been reflected in Table 1.

PROSPECTS FOR LOW COST GPS SYSTEMS

While GPS capital equipment costs are already an order of magnitude lower than those of VLBI and SLR, we are likely to realize another order of magnitude reduction over the next five years; that is nearly a 40% reduction per year. Thus, we face a scenario in the 1990s where GPS capital equipment will become a minor cost driver for GPS systems. Two other elements are major cost drivers in GPS now and will become more so unless significant improvements are made in their efficiency. These are the workforce costs associated with measurement campaigns and with data analysis. It is illustrative to show how costs for the GPS geodetic systems activities at JPL are currently distributed. For JPL, which NASA views as its principal R&D center for developing GPS technology for geodesy and satellite tracking, the FY88 budget distribution for GPS geodesy was as follows:

Campaigns: Planning, Preparation, Field Operations---------	25%
Data Analysis: Systems Analysis, Baseline Recovery---------	31%
Technology Development: Hardware & Software----------------	30%
Procurements: Field Equipment, Computers-------------------	9%
Management:--	5%

For this fiscal year JPL launched one major campaign, the CASA UNO campaign in Central and South America, which also included a global tracking experiment (Neilan et al., 1989). This campaign accumulated approximately 600 station-days of observations over a three-week period and involved 43 GPS receivers. This breakdown shows that nearly two-thirds of JPL's budget was directed in FY88 at supporting scientific applications including Campaigns, Data Analysis and Procurements. Our procurement costs were low because our strategy has been to borrow rather than buy GPS receivers on account of the rapidly moving technology. The point is that, if we are going to do more science with the GPS in the future, we need to invest in R&D that not only reduces the costs of data acquisition equipment, but also the costs of field campaigns and data analysis. I would like to discuss the prospects for getting each of these three cost drivers lower in the 1990s.

The keys to major reductions in campaign and data analysis costs are reducing the capital cost of the GPS receiver and expanding its functionality. A low-cost receiver means that you can buy and deploy a large number of them quadratically increasing the number of baselines recovered in a given time and correspondingly reducing labor costs per baseline. It also means that you can leave them in the field unattended. Expanded functionality means that you can do more things automatically in the field and less at home, and you can do them with fewer people both in the field and at home, thereby greatly reducing data processing costs.

Lowering the Cost of GPS Receivers. Major advances in VLSI and solid state technology made during the 1980s are pivotal in reducing the capital costs of GPS receivers in the 1990s. These advances will result in greatly reduced component and fabrication costs. The TI-4100, the workhorse of the 1980s, is derived from late 1970s technology. It has a lot of analog circuitry. For transistors it uses bi-polar TTL technology extensively, which, by present standards, is very power consumptive, drawing power in both of its logic states. The multiplexing architecture of the TI-4100 baseband processor may have resulted as much from their designers trying to keep the total power requirement of the receiver down as from their desire to minimize costs and possible inter-channel biases inherent in non-multiplexing analog schemes. But a decade makes a big difference in solid state electronics. State-of-the-art CMOS devices consume an order of magnitude less power at the same clock speed as their 1970s counterparts, mainly because of their compactness and because their power consumption is confined to state transitions. Also, they run a lot faster. Gate array chips providing semi-custom VLSI contain as many as 100,000 logic gates today and they are growing; their use in 1980 was virtually nil. Equally important is the availability today of powerful computer aided tools (CAD/CAM) for designing these complex chips and validating their performance before committing the design to fabrication at a foundry, and for their manufacture.

The upshot of these and other innovations, notably the commercial availability of much more power, faster, and cheap microprocessors, is that the top-of-the-line geodetic-quality receiver by the early 1990s will be entirely solid state with the exception of its power supply and antenna; it will be virtually all-digital, and it will have unprecedented computational power. Its all-digital baseband processors will have more parallel channels than GPS signals in the sky and hence, no multiplexing and no inter-channel biases; it will feature high performance microprocessors with clock speeds in the 30 MHz range and upwards (e.g., Motorola 68030) with integrated floating point units, and it will include 10–100 Megabyte memories, both volatile and nonvolatile. In short, the GPS receiver of the 1990s is going to look and operate more like a computer than like a traditional receiver (and its price history will also resemble that of a computer, a curve of very steep descent). It will be compact, light, and cheap.

Because I am familiar with the Turbo-Rogue, an advanced all-digital GPS receiver under development at JPL, I would like to use it for illustrative purposes; however, there are a number of other fine receivers with various features under development within Industry that could be cited as well. The Turbo-Rogue, which is a highly compact field receiver designed for unattended operations, is based on the Rogue architecture (Thomas, 1988). A high level block diagram is shown in Figure 9; some of its performance specifications are included. This receiver will provide comparable accuracy in pseudorange and carrier phase but it will have dramatically expanded on-site computational and functional capabilities. If our design goals are met, the receiver will weigh about 10 kg including the antenna/backplane assembly; it will require about 25 watts of power, and it will cost about $10K. If our schedules hold, it will be available in early 1991.

I would like to give two examples from the Turbo-Rogue development to illustrate how these design goals of low weight, power, and cost are going to be achieved. The first deals with the front end of a GPS receiver, the RF section after the preamplifier and before the baseband processor within which the signal is down-converted successively from RF through IF to an appropriately filtered baseband frequency, and then sampled and digitized. Heretofore, this section has been analog and included synthesizers, mixers, filters, a sampler, and an analog to digital converter (see Figure 10). It has been power consumptive, bulky, expensive, and a potential source of instability and performance degradation through dispersive phase variations from the analog filters.

The Turbo-Rogue front end will bypass all of that with a newly developed Gallium Arsenide custom-designed chip that allows direct digital sampling of a very broad band (600 MHz) RF signal which is passed to the baseband processor with intervening filters (Thomas et al., 1988). Figure 11 diagrams the high level organization of this digital front end chip (DFE) showing its four major functions: a RF sampler and A/D conversion section, a reference clock synthesizer section, a control logic section, and a digital processing section. The input is the RF analog signal; the output is a pair of 4-bit digital samples, one in phase and one in quadrature, each at a controlled rate that can range from 10 to 600 megasamples per second. Through software control the device will operate over a wide range of RF frequencies including the band covering the GPS frequencies at L1, L2, and L3. It uses a unique fractional rate or commensurate sampling scheme on the RF that is controllable by software and that can satisfy the Nyquist sampling criterion for the input bandwidth. Being virtually all-digital, it avoids nearly all of the error sources associated with analog systems. Its power consumption is about 6 W, roughly 25% of the analog front end on the Rogue. Its reproduction cost is about \$100 instead of \$20,000 associated with high quality analog front ends. Its volume is 200 cm^3 compared to approximately 32,000 cm^3. With this device the antenna/backplane, the preamp, and the DFE, all relatively low-cost and low-power items, can be physically located together (with, say, a solar-based power supply) at a remote but desirable site (e.g., low multipath, good field of view); its digital samples can be sent by a fiber optic link to the receiver, which may be kilometers away. A single receiver could service an array of antennas via multiple fiber optic links.

The second example is a highly compacted gate array that is under development for the Turbo-Rogue baseband functions including doppler modeling and stopping, code generation, cross-correlation, and accumulator processing. The Rogue baseband processor (Figure 9) incorporates a three-chip set of CMOS gate arrays, a shared Motorola 68020 CPU @ 12 MHz and a number of memory units to provide all of the GPS signal processing logic for a single satellite including C/A, P_1,P_2,L_1 and L_2. It will also provide a codeless-derived measurement of P_1–P_2 and L_1–L_2 for ionospheric calibration in the event that AS is on. These three VLSI chips were fabricated with 2.0 micron technology and comprise about 18,000 gates of digital logic. A Motorola 68020 has about 12,000 gates. An eight-satellite Rogue baseband processor requires 24 of these chips plus five Motorola 68020 CPU chips; four CPUs are for correlation functions and one CPU is for executive functions. Each chip costs about \$150. A total of nine boards is needed to support these functions in the Rogue: one CPU board and one correlation board per pair of satellites and one CPU executive board for the entire baseband processor. The power consumption of all this logic is about 80 W.

The Turbo-Rogue baseband processor inherits most of the digital design features of the Rogue receiver. However, by taking advantage of recent advances in semi-custom VLSI technology, the processing hardware becomes more compact. The Turbo-chip uses about half of the logic gates on a highly compacted array with 129.000 gates, which is a sub-micron high-speed CMOS device. Each chip will handle all the GPS signal processing for two satellites in addition to taking over a number of tasks previously done by the CPU. Therefore, only four Turbo-chips and two CPUs will be required for eight satellites, at a cost of about \$300 each. Although chip costs are not significantly less, fabrication costs decrease significantly due to fewer CPUs and fewer semi-custom chips; there is only one board in the Turbo-Rogue for baseband processing instead of nine. These advances alone reduce the volume, power consumption, and cost of the Turbo-Rogue baseband processor by a factor of ten.

Prospects for Greater Functionality: The Smart GPS Receiver. One of the two CPUs in the Turbo-Rogue, which is likely to be a Motorola 68030 @ 25 MHz or possibly an 88000 series, is the host processor for the receiver. It provides the executive control for the baseband processor and for its interfaces with the front end, the timing reference, and its output functions. The host CPU also provides all user interface and outside communications functions. It performs all of the tracking data processing functions required by the receiver: initialization and point positioning/timing, time tagging, data editing, data compression, sampling, cycle dropout recovery, receiver diagnostics, and management of ancillary calibration data (e.g., meteorology). It also provides the receiver interface for data transmission and remote command and control. The host CPU can be aided by special purpose hardware performing repetitive tasks that would otherwise be done with software.

The host CPU is central to the concept of the “smart receiver”, a GPS receiver that is highly autonomous and adaptive, and one that can be used in a wide variety of high performance applications. Here are some of the operational features that I would expect smart GPS receivers of the 1990s to have:

1. Extensive on-board processing with special purpose hardware,
2. Adaptive data editing, compression and sampling strategies,
3. Extensive data quality checks and receiver diagnostics
4. Pseudorange-based cycle ambiguity estimation,
5. Carrier range-based point positioning and timing,
6. Real-time pseudorange-based carrier cycle dropout recovery,
7. Autonomous operation with remote data links and control,
8. Cold start acquisition in seconds,
9. Kinematic operations with real-time recovery from dropouts,
10. Operations under high dynamic variability without dropout.

Reduced Operations Costs. These features and others will lead to markedly lower field operations costs as well as reduced analysis costs. These operations costs are $400–$700/station-day today, depending on the difficulty of the campaign and on the unit labor costs (e.g., graduate students versus professional surveyors). The smart receiver with real-time adaptive capabilities will greatly improve productivity in the field: more sites surveyed in less time. Kinematic surveying alone will revolutionize the field of surveying. The single most important field operations cost-reduction feature of receivers of the 1990s will be their ability for autonomous operation. This feature plus the low unit receiver cost means that networks of permanently installed, continuously operating and remotely monitored receivers can be maintained with virtually no human intervention in the field.

Reduced Data Processing Costs. A number of features of smart receivers will greatly relieve post-processing chores and therefore reduce the costs of data analysis. The presentation of very clean phase-connected carrier data, pseudorange, and integrated calibration data as a result of extensive on-site processing will greatly expedite initial post-processing. Automated on-site data compression will also reduce the post-processing load. The availability of ambiguity-resolved carrier phase is another benefit. The availability of reliable and accurate ephemerides would also lead to cost reductions. Probably the biggest factor in reducing data analysis costs is establishing standardization across many system elements: hardware and software configurations; data formats; procedures in site selection, monumentation, data acquisition and data processing; tracking network; reference system; scenarios; archiving; and so on. Nothing takes more time in data analysis than encountering and dealing with the unexpected. Through this kind of discipline and technology advances, the prospects in the 1990s for substantial efficiencies in data analysis are very bright. End-to-end data processing from a major field campaign lasting several days now takes several months to complete; it should take only days. Indeed future arrays operating continuously must do so without accumulating a backlog.

FUTURE APPLICATIONS OF GPS IN GEODESY AND SATELLITE TRACKING

The scientific goals in geophysics and physical oceanography will profoundly influence technology directions in the 1990s for GPS-based geodetic and satellite tracking applications (Mueller et al., 1989). Many of these goals can be transformed into geodetic measurement and/or tracking system requirements, and summarized as follows:

1. Relative velocity vectors of crustal blocks at 1 mm/yr accuracies are required in plate boundary deformation zones with spatial and temporal resolutions ranging over 10–1000 km and from hours up to years. Departures from long-term averages based on tectonic models, including transients, are required.
2. Relative positioning accuracies on land approaching 1 mm are required for intraplate deformation studies and refinement of constraints from tectonic models.
3. Earth rotation vector accuracies of 0.1 mas, measured down to sub-diurnal periods, are required.
4. Seafloor position accuracies of 1 cm are required to support deformation studies in this difficult environment. Potential experiments include monitoring the detailed strain field around spreading centers, subduction zones including the fore and back-arc environment, strike slip zones such as the southern California borderland, and major marine transform faults.
5. A global gravity model with accuracies in the 1 mgal region and 100 km resolution is required.

6. Global ocean circulation studies require satellite altimetry missions that yield ocean topography at 1 cm accuracy with spatial resolution down to 10 km and temporal resolutions of one week spanning several years; comparable accuracies are required for the spectral components of the ocean geoid to derive ocean circulation models from the topography. For altimetry missions this requirement also translates into 1 cm orbit determination accuracies in the radial components of the satellite orbit.

One of the most challenging of the geophysical goals is No.1 pertaining to regional deformation; this requires the deployment of geodetic arrays with unprecedented spatial and temporal resolutions and with very high accuracies. The extensiveness of the measurement programs implied by this requirement, if carried out on a global basis, would lead to enormous costs if carried out with SLR and VLBI. This leaves only the GPS as the economically viable space technique because of its low cost and high accuracy potential.

For GPS these requirements can be grouped into the following applications categories:

1. Arrays for monitoring crustal deformation rates,
2. Networks for maintenance of terrestrial reference systems including GPS ephemerides and fiducial site coordinates plate motions and Earth orientation variability;
3. Tracking and navigation systems for satellite dynamics and for satellite and/or air-borne remote sensing applications;
4. In-situ applications: kinematic surveying, translocation, etc.

I would like to briefly discuss the various GPS applications as they fall into these four broad categories. Most of the applications of GPS to date have been in array form; but their density is relatively sparse, they are infrequently operated, and they are very labor-intensive. The continuous array will redress many of these limitations and by year 2000 one would expect to find dozens of these arrays globally dispersed.

<u>Continuously Operating Remotely Monitored Arrays</u>. Certainly the continuously operating array is an idea whose time has come (Ladd, 1987; Shimada et al., 1988). Here, an array of unattended receivers is connected via a communications link to a remote data processing or network Center. The receivers track continuously and the tracking and ancillary data are collected via the link and processed at a real-time rate by the Center. The Center also controls the elements of the array and monitors their performance. The low cost, high accuracy and smart GPS receiver, discussed in Section III, is required for these arrays. Low cost means that dense arrays are economically feasible; high accuracy means that very weak geophysical signatures might be discernible; and “smart” means that the autonomous and adaptive

capabilities of the array receivers make unattended operations feasible and render the enormous real-time data processing tasks economical. Specific examples of on-site receiver data processing and management functions are editing, phase-connecting, calibrating, compressing, and data transmitting. P-code pseudorange is essential for some of these tasks. The receiver could automatically adjust its sample rate upon detection of anomalous doppler rates. Post-processing by the Center would be greatly streamlined over current off-line capabilities, which run at roughly one-tenth the real-time rate. Powerful parallel processing architecture will be commonplace; sophisticated post-filter editors will be used for combining separate network information, e.g., yesterday's solutions and information matrices with today's, or information from one array with that from another or from a reference or global tracking network, the Center will provide highly interactive capabilities for users and efficient and automated data management and archival systems. The potential applications for continuously operating arrays include:

1. In California, arrays containing 100–1000 elements for real-time monitoring of pre-seismic, coseismic and post-seismic strain with densification in problem areas;
2. Arrays for remotely monitoring volcanic activity;
3. Arrays of tethered and free-floating ocean buoys for ocean circulation, sea level and tidal studies;
4. Sparse arrays, e.g., inter-island geodesy over oceanic basins;
5. Sea floor geodesy: for positioning acoustic-based ocean platforms and tying sea floor monuments to a GPS network (Spiess, 1987);

Skeptics of the utility of continuous tracking contend that it might be compromised by water vapor variability. On the other hand, continuous tracking will certainly yield a great deal of information about the spectral content of tropospheric variability and should lead to better strategies for reaching the shortest temporal resolution of potential geophysical signatures.

Kinematic Surveying. Kinematic surveying depends on maintaining cycle ambiguity-resolved carrier phase in the field so that carrier range-based geometric point positioning can be continuously achieved under kinematic conditions (Remondi, 1985; Mader et al., 1989). Its accuracy potential is at the millimeter level for relative positioning. It is in its infancy today and susceptible to signal outages and loss of phase connection. But with the coming advent of the smart receiver with P-code pseudorange and carrier phase capability, I would expect kinematic surveying to be widely practices in the 1990s. Its potential for extraordinary cost effectiveness will lead to a wide range of applications with accuracy requirements ranging from 1 ppm (i.e., standard surveying) to 0.01 ppm or better. Its theater will be virtually unlimited in extent. (The smart receiver with good ephemerides should enable the user to transfer cycle ambiguity resolution from one satellite to another upon rise or set, and from one day to the next.) Kinematic surveying opens the way for obtaining geodetic measurements at multiple sites around a primary site all within one observation session; there will be no need in the future to dwell at a primary site for multiple days, or even for hours. Future kinematic surveying will also be effected with a single roving GPS receiver operating in conjunction with a remote array, thereby enabling virtual baselines among sites occupied by the rover to be measured with carrier range accuracy using only the one receiver.

GPS Global Tracking System. This topic has already been partially addressed previously under Reference Systems. Here, I will briefly describe the management and operation of a global tracking system, and enumerate its major uses and advantages. In a sense, a global tracking system is another array but on a global scale; it will have many of the features of a continuously operating array and incorporate many of its efficiencies. It will coexist with other global tracking systems such as the five-station USAF network for monitor and control of the GPS and for the broadcast ephemerides, timing, health status, and so on. For its surveying and mapping programs, the Defense Mapping Agency maintains an additional five tracking stations that complement the USAF sites. Other civilian users for surveying, transportation and navigation are likely to generate their own systems, but these accuracy requirements probably do not exceed 0.1 ppm and in many instances are in the 1–10 ppm range. Here, we are addressing a system that will maintain the terrestrial reference system for GPS at accuracy levels approaching 0.001 ppm. It will also feature GPS tracking from low Earth orbiters to augment the ground network. It will be reliable, its products timely, and its configuration and operation standardized by international agreement through organizations such as the IUGG/IAG. Figure 12 provides a flow diagram showing the major elements of an internationally organized global tracking system and the different classes of users. The System would have four major elements:

1. An International Governing Board:
2. An Operational Center and possibly sub-Centers functioning on a regional basis;
3. A network of tracking stations hosted by the participating country/organization;
4. One or more Computational Centers.

The GPS Global Tracking System would deliver tracking data, ephemeris and reference frame data products, information matrices, and calibrations. The users, if they were custom users, might further process the tracking data or the ephemeris solutions and information matrices to refine the accuracy and time resolution of solutions from their own networks. They might return some of their products to the Computational Centers for further integration. For example, the IERS would use these products to further strengthen its Earth orientation monitoring function. Tracking information from continuous arrays and from LEO's carrying GPS receivers might be sent to the computational centers to enhance the accuracy of their solutions. The standard user would use the ephemeris products and station coordinates as they stand. For these users the Global Tracking System would provide significant cost-savings as well as accuracy, stability, and reliability.

Is implementing, maintaining, and operating another civilian global tracking network worth the cost? Can nearly equivalent accuracy and stability be achieved by less ambitious measures? If GPS applications were strictly regional arrays with strong fiducial constraints, the utility of a global network would be dubious. A priori ephemeris values and covariances supplied to such a region appear to improve overall system accuracy; but this improvement is marginal and tends to apply only to the longer baselines if multi-pass regional tracking is obtained (Lichten et al., 1989). On the other hand, preliminary results from the global tracking network used in the CASA UNO experiment suggest that it facilitates cycle ambiguity resolution on the continental-sized baselines of that experiment. If this is true, it would lead to significant accuracy advances on these baselines. In summary, the major advantages of a global tracking system are:

1. Cost savings through reduced operations and computation tasks;
2. Standardization of the Terrestrial Reference System;
3. Standardization of the data management and information systems;
4. Timely tracking and ephemeris products through a world-wide distribution systems;
5. Increased accuracy, particularly for regions with weak fiducial constraints and for oceanic basins;
6. Shortened resolution time for regional arrays; i.e., shortened tracking time for a regional array to achieve a given accuracy level;
7. Complementary Earth orientation information for the IERS;
8. Ground-tracking network for differential GPS applications to Earth satellite missions.
9. Enhanced accuracy and temporal resolution from concurrent tracking by LEOs.

Although the GPS and the SLR depend on VLBI for maintaining the inertial reference system and for eliminating possible longitude/right ascension drifts inherent in dynamical systems, the introduction of GPS into Earth orientation monitoring will fundamentally change the mix of these three techniques for this application. GPS tracking, with its effective ranging accuracy at the centimeter level and range change accuracy at the millimeter level, will provide complementary and accurate short period information, and longer period information of comparable accuracy.

Satellite Precision Orbit Determination (POD). By placing a high performance flight-rated GPS receiver aboard an Earth satellite and using a globally distributed ground-tracking network, one can achieve sub-decimeter positioning of the satellite in the terrestrial frame defined by the ground network (Yunck et al., 1985). The first mission to use this approach as a flight experiment will be TOPEX/Poseidon, a NASA/CNES oceanographic satellite to be launched in 1992 (Melbourne and Davis, 1987; Carson et al., 1988). Here, the requirement for high accuracy in radial position stems from its use of precise satellite altimetry. Its three-year mission is to map the ocean topography to decimeter accuracy or better with mesoscale resolution; this information coupled with an improved ocean geoid, will enable determination of geostrophic currents and will lead to better understanding of both variable and mean ocean circulation from mesoscales up to global scales. Obviously, the requirement for decimeter accuracy altimetry or better implies the need for decimeter or better orbit accuracy. The GPS is likely to be the best way to reach sub-decimeter accuracies. Future remote sensing missions of the late 1990s such as Eos will also carry GPS receivers. Ultimately, centimeter accuracy should be achieved using precision orbit determination strategies that synergistically combine carrier phase and pseudorange. A critical component in GPS-based satellite POD is the GPS Global Tracking System; the ground stations of this network must concurrently track the GPS satellites to achieve these orbit accuracies in differential positioning.

Platform Positioning. In addition to positioning, GPS will provide attitude information for those remote sensing missions accomplished either by satellite or by aircraft. Three or more GPS antennas multiplexed to a receiver will provide the attitude, typically to 100 μrad accuracy. For dynamic platforms such as marine platform for sea floor geodesy or an Earth satellite platform such as NASA's planned Earth Observation System (Eos), P-code pseudorange will be a vital aid to carrier phase. In the case of Eos, the GPS will also be used for recovery of stratospheric temperatures by tracking GPS satellites during their occultation phases. Also, ionospheric tomography studies will be undertaken as well as satellite-based real-time geodesy (Melbourne, et al. 1988).

By combining kinematic tracking, multiplexed antenna arrays for attitude determination, and gravimeters, one should be able to mount airborne and shipborne experiments for regional gravity logging with accuracies and productivity that will be unprecedented.

Synthetic Aperture Radar (SAR) Missions. In addition to platform attitude, the GPS will provide positioning to an accuracy that is a small fraction of the radar wavelength, an important measurement for SARs. This positioning accuracy also allows radar data from successive overflights to be coherently combined. For target areas remaining invariant between overflights one obtains synthesized radar images of very high resolution equivalent to an aperture the size of the separation of the overflights. Indeed, the loss of coherence within a target area between overflights can also provide useful information.

Gravity Recovery with GPS-based Satellite Geodesy. A satellite-borne GPS receiver operating differentially in conjunction with a ground network can provide significant improvement in our geopotential models. Gravity perturbations, particularly as signatures manifested in the highly precise carrier phase-tracking, will be detected. Their determination will be enhanced by successive overflights of the same geographical area. This will be first performed with TOPEX/Poseidon to improve its overall orbit determination accuracy and to improve the ocean geoid. Because the GPS provides continuous and virtually omni-directional tracking, its gravity information will complement existing gravity models, which are currently less accurate in certain geographical regions as a result of a paucity of laser-tracking data from earlier missions over oceanic regions and most of Asia. However, TOPEX/Poseidon will fly at 1334 km, thus its resolution of the spectral components of gravity will be limited to wavenumbers of about 25 or lower, plus a few resonance terms (Yunck et al., 1985). Its results will also be vulnerable to mis-modeled non-gravitational accelerations.

Gravity Probe-B, a proposed NASA mission to be launched in the mid-1990s, offers much better prospects. Principally a mission to detect gyroscopic effects predicted by general relativity, it will be drag-free and fly in a polar orbit at 600 km altitude (Proceedings of SPIE, 1986). By also flying a GPS receiver, which is now planned, it should recover valuable gravity information up to a wavenumber of about 60. This maximum wavenumber falls short of the 200–300 that could be achieved with the Gravity Research Mission using a micron accuracy microwave satellite-to-satellite tracking technique, or alternatively, with super-conducting three-axis gravity gradiometer. But missions with this kind of technology are not likely to fly until near the end of this century; so these earlier but less precise results are likely to serve as interim measures. Figure 13 provides an estimate of the gravity recovery potential of GP-B for a six-month mission.

Another mission being developed by ESA, Aristoteles, will fly a gravity gradiometer and may also carry GPS receiver. This satellite will be placed in a 200–250 km orbit which may yield additional resolution. Unfortunately, Aristoteles is not currently scheduled to carry a drag-free or drag-compensated system; thus non-gravitational atmospheric drag and radiation pressure effects are likely to corrupt the gravity coefficient estimates, but the magnitude of this corruption versus spectral component is uncertain at the present.

SUMMARY

Most of this discussion has focused on the state of affairs in high accuracy differential applications of the GPS as they are now and will be a few years ahead; viewed from a distance, CY1995 is much easier to see than CY2000. For example, it is likely that other navigation satellite systems will play a complementary role in geodesy, such as GLONASS of the Soviet Union, and also flight systems such as PRARE and DORIS; but the details are less clear. There are also a myriad of surveying and navigation applications for GPS that are outside of our realm; these will undergo explosive growth in the next several years. But for our class of GPS applications, these important milestones will almost certainly be passed by CY2000 if present trends continue:

1. System accuracy for geodetic applications should improve by a factor of five. For long period resolutions where the tropospheric error contribution will be less, system accuracies should approach 1 mm in the horizontal components of regional baselines. For short period resolutions, it may be possible to surpass 3 mm.
2. Data acquisition hardware costs should drop by over an order of magnitude with top-of-the-line GPS receiver costs of less than $3000 in 1988 dollars; indeed the standard C/A-only receiver for coarse positioning will probably cost less than today's scientific calculator.
3. All-digital WVR instrumentation will be widespread. Its cost will be below $10,000 using MMIC technology and phased array antennas. By employing highly stable but small ovens their calibration system should enable measurements of sky temperature to 0.1 K accuracies.
4. Data processing and field operations costs should be over an order of magnitude lower than today's costs.
5. Continuously operating and remotely monitored arrays with unattended operations and automated data processing should become widespread for a broad class of applications including regional and global arrays, seafloor geodesy, satellite tracking for remote sensing missions, and ocean and ice circulation.
6. Kinematic surveying will have revolutionized surveying methodology and reduced costs by over an order of magnitude. It will enable the practical realization of regional footprints around primary geodetic sites for monitoring local strain distribution.
7. An internationally sponsored and cooperative GPS global tracking network will be operating to maintain a terrestrial reference system approaching an accuracy of 0.001 ppm, to support Earth orientation monitoring, and to support ground programs and Earth satellite missions requiring high accuracy positioning.
8. Long-life low Earth orbiters such as TOPEX/Poseidon and the Eos platforms will bear GPS receivers to augment the ground-tracking network and to further improve system accuracies.

Finally, having dwelled at length on the attributes of GPS for scientific applications, it would be remiss not to cite at least one of its deficiencies. The designers of the GPS were concerned with developing this nation's next generation satellite navigation system, not with its scientific applications. The P-code signal structure was designed to produce sub-dekameter navigation accuracies, not sub-centimeter geodetic accuracies. The effective wavelength of the P-code is over two orders of magnitude longer than the carrier wavelengths. For carrier cycle ambiguity resolution with high confidence, the P-code measurement accuracy in its averaged form must be below 1 cm, or about 0.02% of its effective wavelength. This is a burdensome requirement that is difficult to reliably meet today. One improvement is to add another tone or spread spectrum component in the GPS signal structure, coherent with 10.23 MHz P-code clock frequency but having a wider frequency base. An increase of at least a factor of five would lead to a ranging system with greatly reduced susceptibility to multipath, which with due care, would virtually assure 100% success in resolving ambiguities independent of network configuration or extent. Another very promising option is to use the already existing L3 carrier for navigation purposes instead of just for telemetry (Melbourne, 1985). It is plausible that by CY2000 the DoD will have begun implementing the GPS Block III satellites. Once can hope that they might reflect a few improvements promoted by the high precision civilian community.

ACKNOWLEDGEMENTS

I wish to express my appreciation to several individuals for their varied contributions to this work in the form of discussions and critiques, figures and technical data. They are: Sassan Bassiri, Geoff Blewitt, Tim Dixon, Michael Janssen, Steve Lichten, Tom Meehan, Ruth Neilan, Benno Rayhrer, John Scheid, and Larry Young, all from JPL. This work was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract from the National Aeronautics and Space Administration.

REFERENCES

Anon., ICD-GPS-207: NAVSTAR GPS Selective Availability and Anti-Spoofing Receiver Design Requirements, Current Issue, NAVSTAR/GPS Joint Program Office, USAF Space Command, El Segundo, CA, 1989.

Bender, P., Atmospheric Refractivity Uncertainties, 1987.

Beutler, G., W. Gurtner, M. Rothacher, and I. Schildknecht, Evaluation of the March 1985 High Precision Baseline (HPBL) Test: Fiducial Point Concept Versus Free Network Solutions, EOS Trans., AGU, Vol. 67, p. 911, 1986.

Blewitt, G., S. Lichten, P. Kroger, M. Kornreich, U. Lindquister, L. Skrumeda, and W. Bertiger, Accuracy and Long-Term Stability of GPS Baseline Estimation, EOS Tran. AGU, Vol. 69, No. 44, p. 1151, 1988.

Blewitt G., Carrier Phase Ambiguity Resolution for The Global Positioning System Applied to Geodetic Baselines up to 2000 KM, J. Geophys. Res., in press, 1989.

Bock, Y., R. Abbot, C. Counselman III, and R. King, A Demonstration of 1–2 Parts in 10 Accuracy Using GPS, Bull. Geod., Vol 60, pp. 241–254, 1986.

Carson, L., R. Davis, G. Geier, L. Hailey, G. Huth, T. Munson, Design and Predicted Performance of the GPS Demonstration Receiver for the NASA TOPEX Satellite, Proc. IEEE PLANS 88, Position, Location, and Navigation Symposium, Orlando, FL, 1988.

Clark, T., D. Gordon, W. Himwich, C. Ma, A. Mallama, and J. Ryan, Determination of Relative Site Motions in the Western United States Using Mark III Very Long Baseline Interferometry, J. of Geophys. Res., Vol 92, No. B12, 1987.

Clark, T., Private communication, 1988.

Clynch, J., and B. Renfro, Evaluation of Ionospheric Residual Range Error Model, Proc. Third International Geodetic Symposium on Satellite Doppler Positioning, Vol. 1, p. 517, Las Cruces, NM, 1982.

Davidson, J., C. Thornton, C. Vegos, L. Young, and T. Yunck, The March 1985 Demonstration of the Fiducial Network Concept for GPS Geodesy: Preliminary Report, Proc. First International Symposium on Precise Positioning with the Global Positioning System, Positioning with GPS-1985, Vol. 1, p. 603, Rockville, MD, 1985.

Dong, D., and B. Bock, GPS Network Analysis with Phase Ambiguity Resolution Applied to Crustal Deformation Studies in California, J. Geophys. Res., in press, 1989.

Elgered, G., J. Davis, T. Herring, and I. Shapiro, Methods of Correction for the “Wet” Atmosphere in Estimating Baseline Lengths from VLBI , Proc. IAU Symp. No. 129, May 10–15, 1987, p. 543, Cambridge, MA, 1988.

Freymueller, J. and M. Golombek, Geometry and Treatment of Fiducial Networks: Effect on GPS Baseline Precision in South America, Geophys. Res. Lett., Vol. 15, No. 13, p. 1467–1469, 1988.

Hatch, R., The Synergism of GPS Code and Carrier Measurements, Proc. Third International Geodetic Symposium on Satellite Doppler Positioning, Vol. 3, p. 1213, Las Cruces, NM, Feb. 8–12, 1982.

Herring, T., J. Davis., I. Shapiro, Geodesy by Radio Interferometry, the Application of Kalman Filtering to the Analysis of VLBI Data, J. Geophys. Res., to be submitted, 1989.

Jannsen, M., A New Instrument for the Determination of Radio Path Delay Due to Atmospheric Water Vapor, IEEE Trans. on Geo. Science and Remote Sens., Vol. GE-23, No. 4, p. 485, 1985.

Kaula, W., (Ed.), The Terrestrial Environment Solid-Earth and Ocean Physics: Application of Space and Astronomic Techniques, Report of a Study at Williamstown, Mass. to NASA, August 1969, sponsored by NASA and MIT.

Ladd, J., Continuous Monitoring of Deformation with GPS, Proc. Deformation Measurements Workshop, MIT, Oct. 31-Nov. 1, 1986, p. 416–436, Cambridge, MA, 1987.

Lichten, S., and J. Border, Strategies for High-Precision Global Positioning System Orbit Determination, J. of Geophys. Res., Vol. 92, No. B12, p. 12751–12762, 1987.

Lichten, S., W. Bertiger, and U. Lindqwister, The Effect of Fiducial Network Strategy on High-Accuracy GPS Orbit and Baseline Determination , Proc. from Fifth International Geodetic Symp. on Satellite Positioning, March 13–17, Las Cruces, NM, in press, 1989.

Lichten, S., and W. Bertiger, Demonstration of Sub-Meter GPS Orbit Determination and 1.5 Parts in 108 Three-Dimensional Baseline Accuracy , submitted to Bulletin Geodesique, in press, 1989.

Lindqwister, U., G. Blewitt, W. Bertiger, Future of GPS Network Processing , EOS, Vol. 69, No. 44, p. 1152, 1988.

Mader, J., W. Carter, B. Douglas, W. Krabill, Decimeter Level Aircraft Positioning with GPS Carrier-Phase Measurements, Bulletin Geodesique, in press, 1989.

Melbourne, W., Series X Cycle Ambiguity Resolution Using Bandwidth Synthesis and “Angle Synthesis,” IOM #3300-82-119, Jet Propulsion Laboratory, 1982.

Melbourne, W., The Case for Ranging in GPS-Based Geodetic Systems, Proc. First International Symposium on Precise Positioning with the Global Positioning Systems, Positioning with GPS-1985, Vol. 1, p. 373, Rockville, MD., 1985.

Melbourne, W., and E. S. Davis, GPS-Based Precision Orbit Determination: A TOPEX Flight Experiment, AAS/AIAA Astrodynamics Specialist Conference, Paper AAS 87–430, Kalispell, Montana, 1987

Melbourne, W., G. Blewitt, S. Lichten, R. Malla, R. Neilan, S. Wu, and B. Schutz, Establishing a Global GPS Tracking System for Fiducial and Ephemeris Production, Paper No. G21A-02, AGU Spring 1988, 1988.

Melbourne, W., T. Yunck, L. Young, B. Hager, G. Lindal, C-H Liu, G. Born, GPS Geoscience Instrument for EOS and Space Station Global, Atmospheric Temperature Profiling, Acoustic Gravity Wave analysis, Ionospheric Tomography and Precise Tracking Using the Global Positioning System, Jet Propulsion Laboratory proposal, Investigation, Technical, Data and Management Plans, Vol. 1, July 1988.

Mueller, I., S. Zerbine, J. Dickey, G. Freeman, C. Goad, W. Kaula, W. Melbourne, C. Reigber, R. Schutz, D. Turcotte, (Editors), The Interdisciplinary Role of Space Geodesy, Proc. of the International Workshop, Institute Nazionale di Geofisica in Erice, Sicily, Italy, July 1988, Springer-Verlag, in press, 1989.

Neilan, R., T. Dixon, J. Kellogg, T. Meehan, W. Melbourne, J. Scheid, J. Stowell, Operational Aspects of CASA UNO '88-The First Large Scale International GPS Geodetic Network, IEEE Trans. on Instrumentation and Measurement, in press, April 1989.

Remondi, B., Performing Centimeter-Level Surveys in Seconds with GPS Carrier Phase: Initial Results, NOAA Tech Memorandum NOS NGS-43, Rockville, MD, Oct. 1985.

Rogers, A., The Accuracy of Position Determination by VLBI Expected by the Year 2000, in Geodesy in the Year 2000, NRC Committee on Geodesy Nat. Acad. Press, Washington, DC, 1987.

Shimada, S., S. Sekiguchi, T. Eguchi, Y. Okada, and Y. Fujinawa, Fixed-Point GPS Simultaneous Baseline Determination Network in Kanto Tokai District, Central Japan, AGU Chapman Conference on GPS: Measurements for Geodynamics, Sept. 19–23, 1988, Fort Lauderdale, FL, 1988.

Sims, M, Phase Center Variation in the Geodetic TI4100 GPS Receiver System's Conical Spiral Antenna, First International Symposium on Precise Positioning with the Global Positioning System, Positioning with GPS-1985, Vol. 1, p. 227, Rockville, MD, 1985.

Spiess, F., Seafloor Geodesy by the Year 2000, Nov. 1987.

Thomas, J., Functional Description of Signal Processing in the Rogue GPS Receiver, JPL Publ. 88-15, 1988.

Thomas, J, B. Rayhrer, L. Young, A Sampling Down-Converter for RF Signals, NASA New Tech. Report #NPO-17530, 1988.

Tralli, D., T. Dixon, and S. Stephens, The Effect of Wet Tropospheric Path Delays on Estimation of Geodetic Baselines in the Gulf of California Using the Global Positioning System, J. Geophys. Res., 93, p. 6545–6557, 1988.

Tranquilla, J., Multipath and Imaging Problems in GPS Receiver Antennas , Proceedings of the Fourth International Geodetic Symposium on Satellite Positioning, p. 557–571, 1986.

Treuhaft, R., G. Lanyi, The Effect of the Dynamic Wet Tropospheric on Radio Interferometric Measurements , Radio Sci., Vol. 22, No. 2, pp. 251–265, 1987.

Wu, S., and R. Malla, Determination of a Geocentric Coordinate Frame for GPS Measurements, AIAA/AAS Astrodynamics Conference, AIAA 88-4210-CP, Minneapolis, Minnesota, 1988

Wu, S., W. Melbourne, T. Yunck, Impact of Tracking Network Variation on GPS Orbit Determination, AIAA Paper #88-0573, AIAA 26th Aerospace Sciences Meeting, Reno, Nevada, 1988.

Young, L., R. Neilan, F. Bletzacker, GPS Satellite Multipath: An Experimental Investigation, First International Symposium on Precise Positioning with the Global Positioning System: Positioning with GPS-1985, Vol. 1, p. 423, Rockville, MD., 1985, also Young, L., Private communication, 1988.

Yunck, T., W. Melbourne, and C. Thornton, GPS-Based Satellite Tracking System for Precise Positioning, IEEE Trans. Geosci. and Remote Sens., GE-23(4), p. 450–457., 1985.

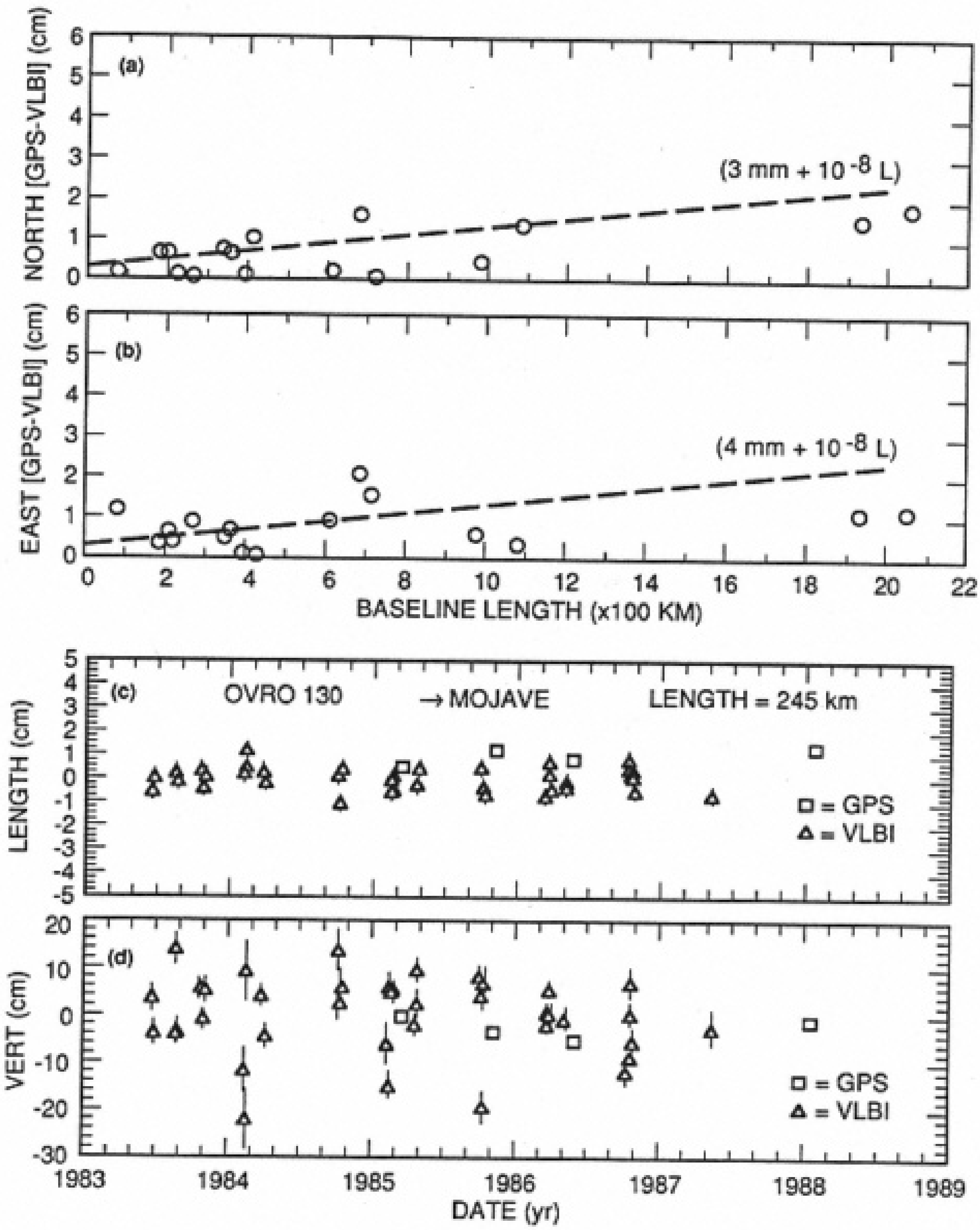


Figure 1.
Agreement Between VLBI and GPS. a) and b) Difference in Horizontal Components of Baselines Versus Baseline Length; Continental U.S. Campaigns of March and November, 1985, June, 1986, and January, 1988. c) and d) VLBI and GPS Length and Vertical Results Over a Five Year Period; OVRO/Mojave Baseline.

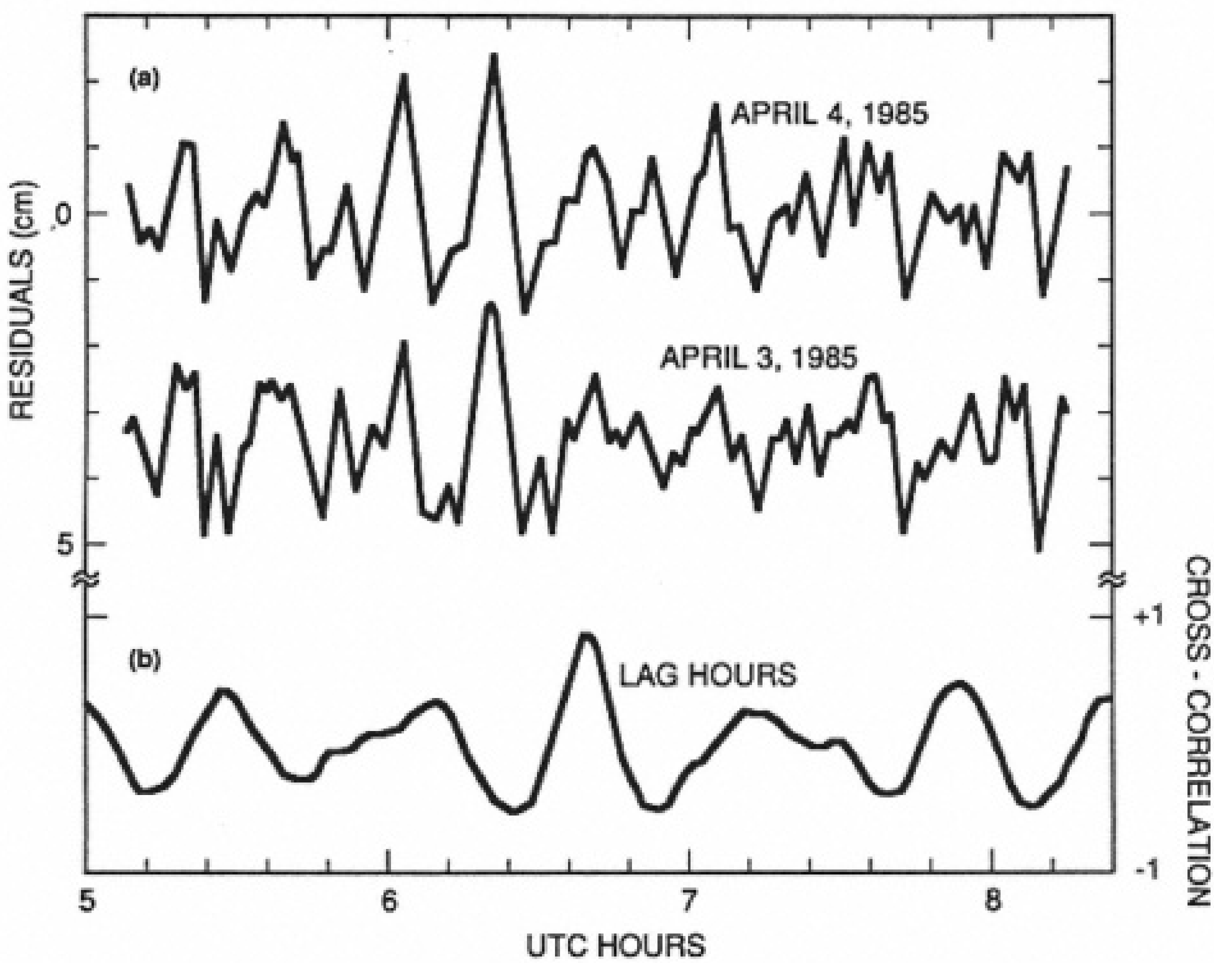


Figure 2.
Carrier Multipath. a) Double Differenced Ionosphere Corrected, Post-Fit, and 2 Minute Gaussian Smoothed Carrier Phase Residuals Using a Pair of TI-4100 Receivers with the Standard TI Antenna Observing SV 9 and SV 11 from Mojave and OVRO. Use of Ground Plane Could Reduce Multipath Variability Significantly. b) Cross-Correlation of April 3 and April 4 Residuals Versus Lag.

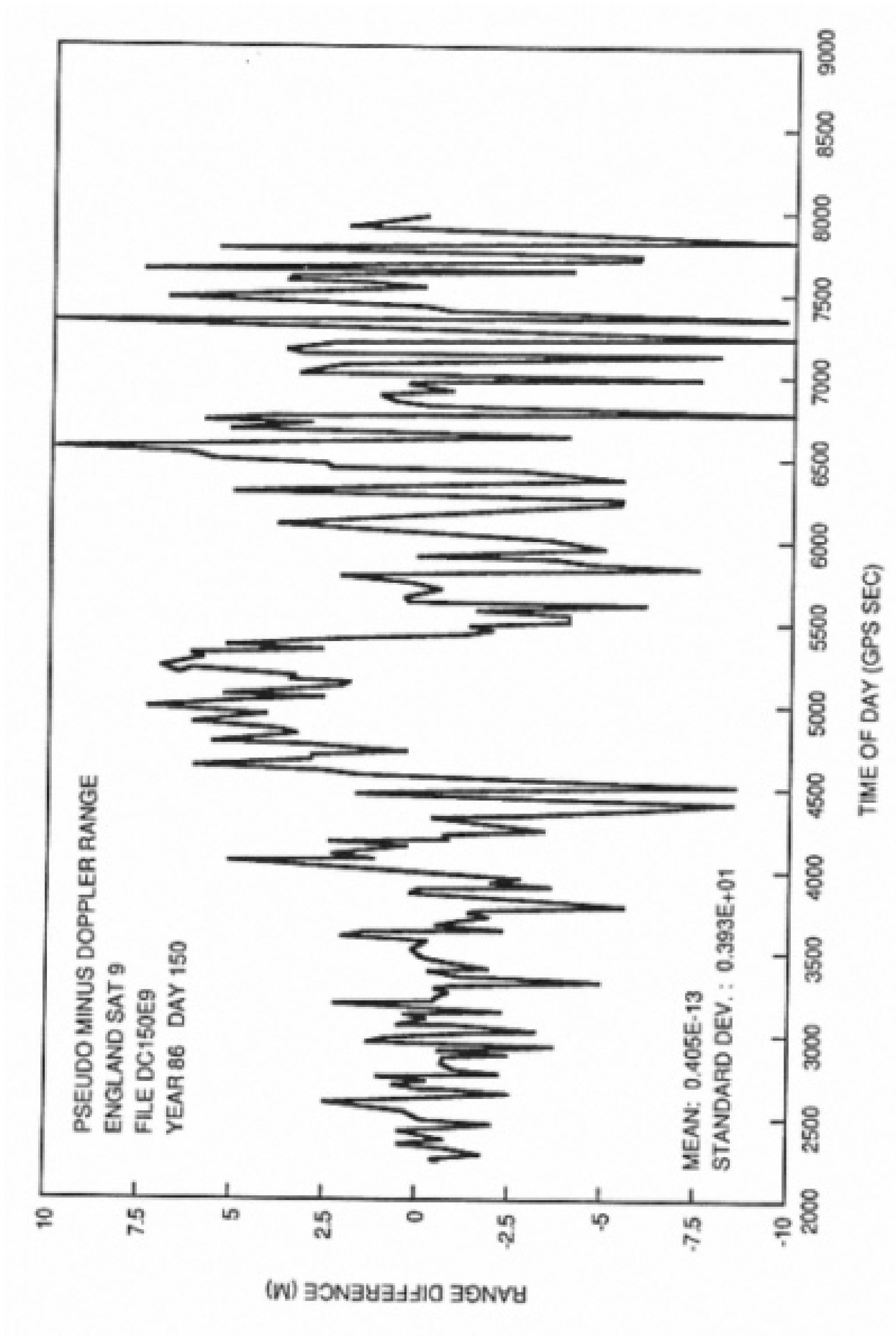


Figure 3.
P-code Pseudorange Multipath Using a TI-4100 in a Typical Environment.

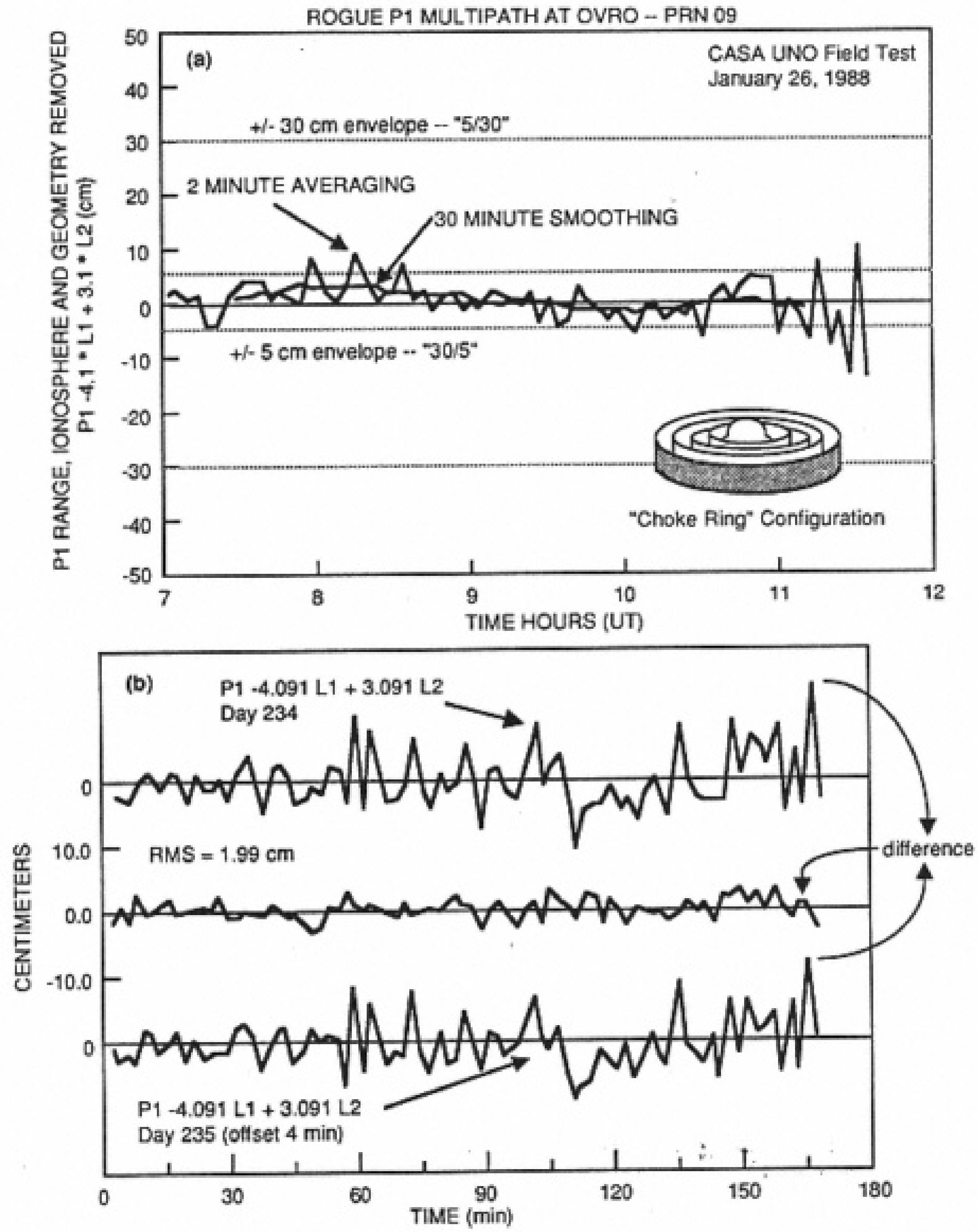


Figure 4.
P-code pseudorange multipath obtained by a Rogue receiver operating with a Dorne & Margolin antenna coupled with a quarter wavelenth choke-ring backplane designed by JPL; a) Ordinate is the linear combination of Equations 1 and 2 that eliminates non-dispersive and ionospheric delays and should be a constant in absence of multipath and SNR errors. Variability is a strong measure of P-code multipath. b) P-code pseudorange multipath on two successive days; residuals are 2 minute averages.

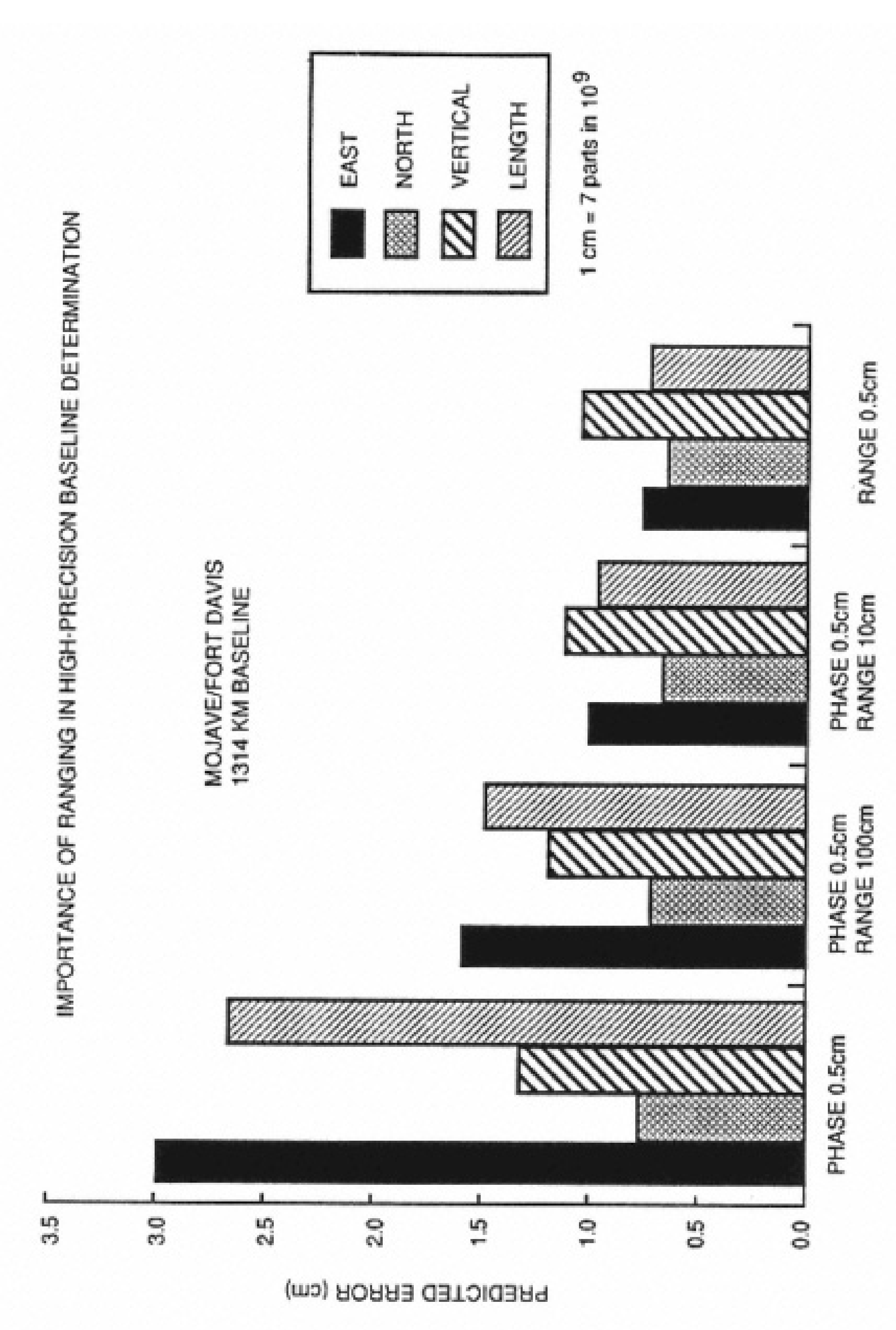


Figure 5.
Predicted Baseline Recovery Accuracy Using Carrier Phase Tracking Over a Single Day Versus Accuracy of Ranging. Carrier Range is Ambiguity-Resolved Carrier Phase. Covariance Assumptions Are Found in Ref. 29.

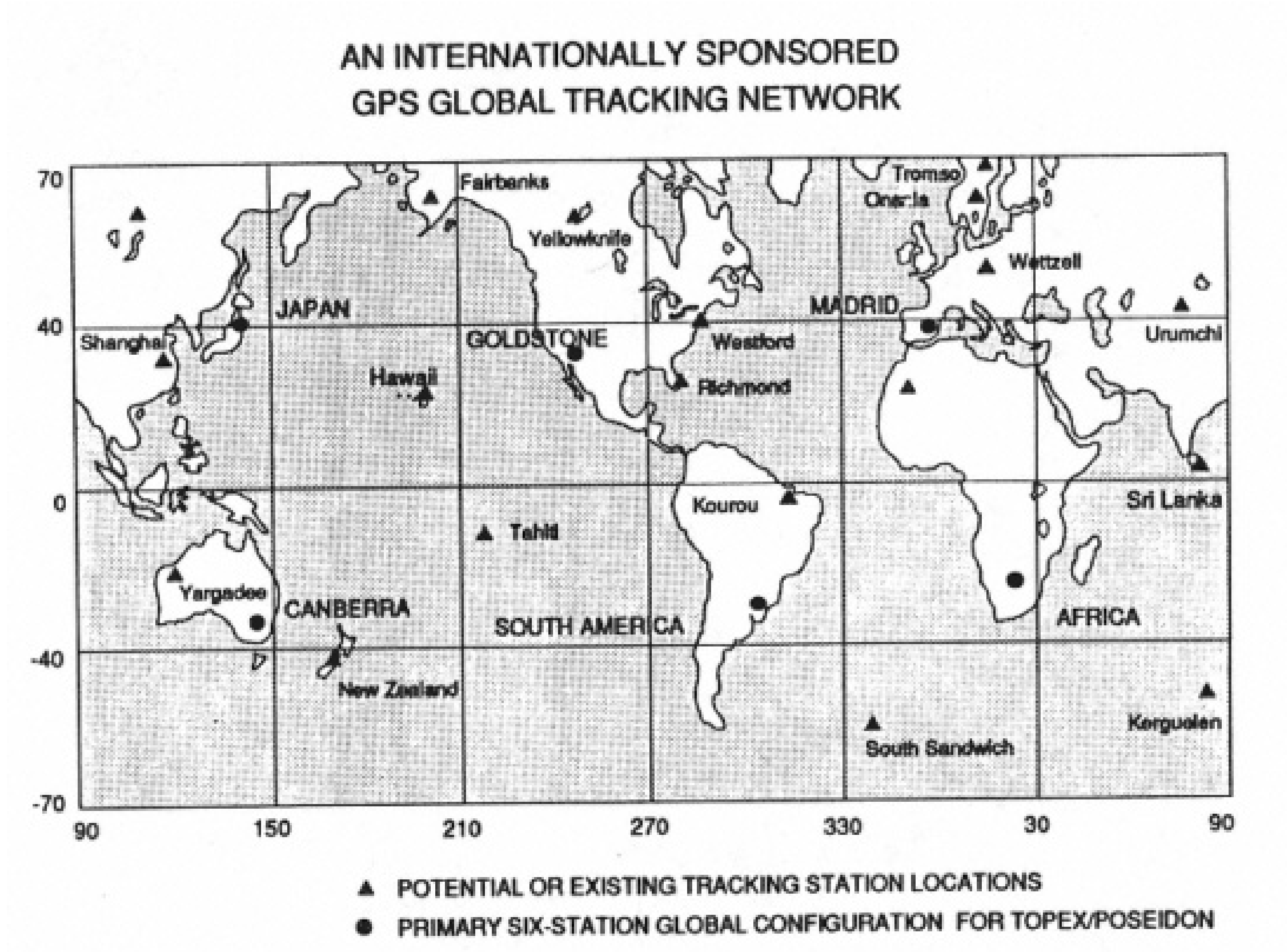


Figure 6.
Probable Locations of GPS Global Tracking Stations, Circa 1992. USAF and DMA Monitor Stations (Not Shown) are Operational in 1988. TOPEX/Poseidon Network Planned for 1991.

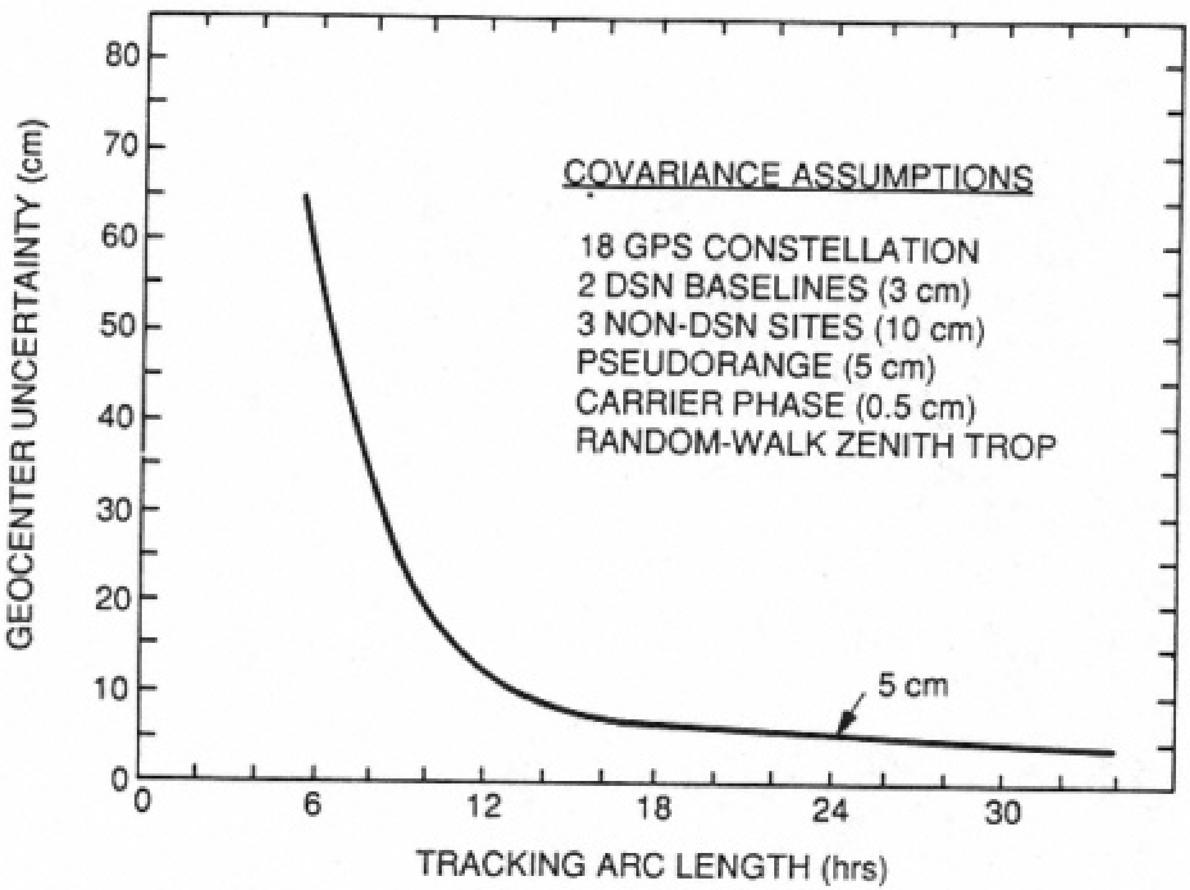


Figure 7.
Predicted Accuracy Versus Tracking Time for Recovery of Geocentric Offset in Reference Frame Using Full 21-Satellite Block II Constellation and a 12-Station Global Tracking Network. See [30] for Additional Covariance Assumptions

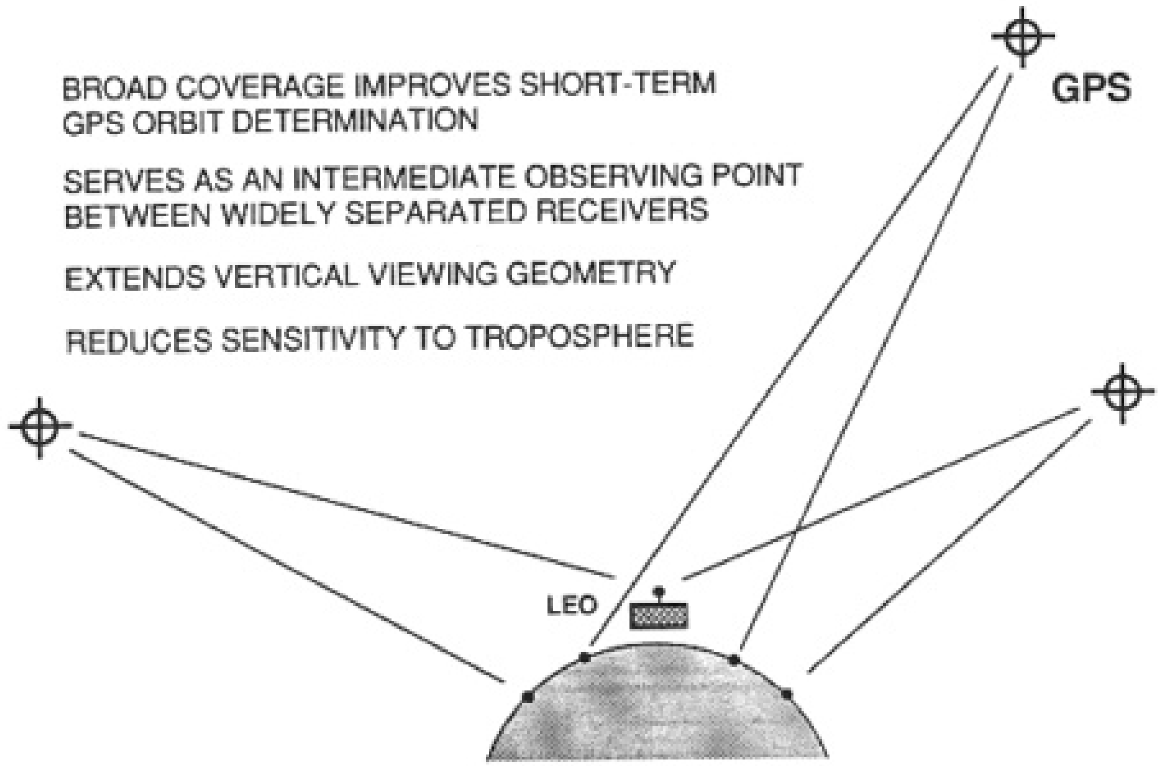


Figure 8.
Utility of LEO in enhancing accuracies of baselines and GPS ephemerides using a global tracking network.

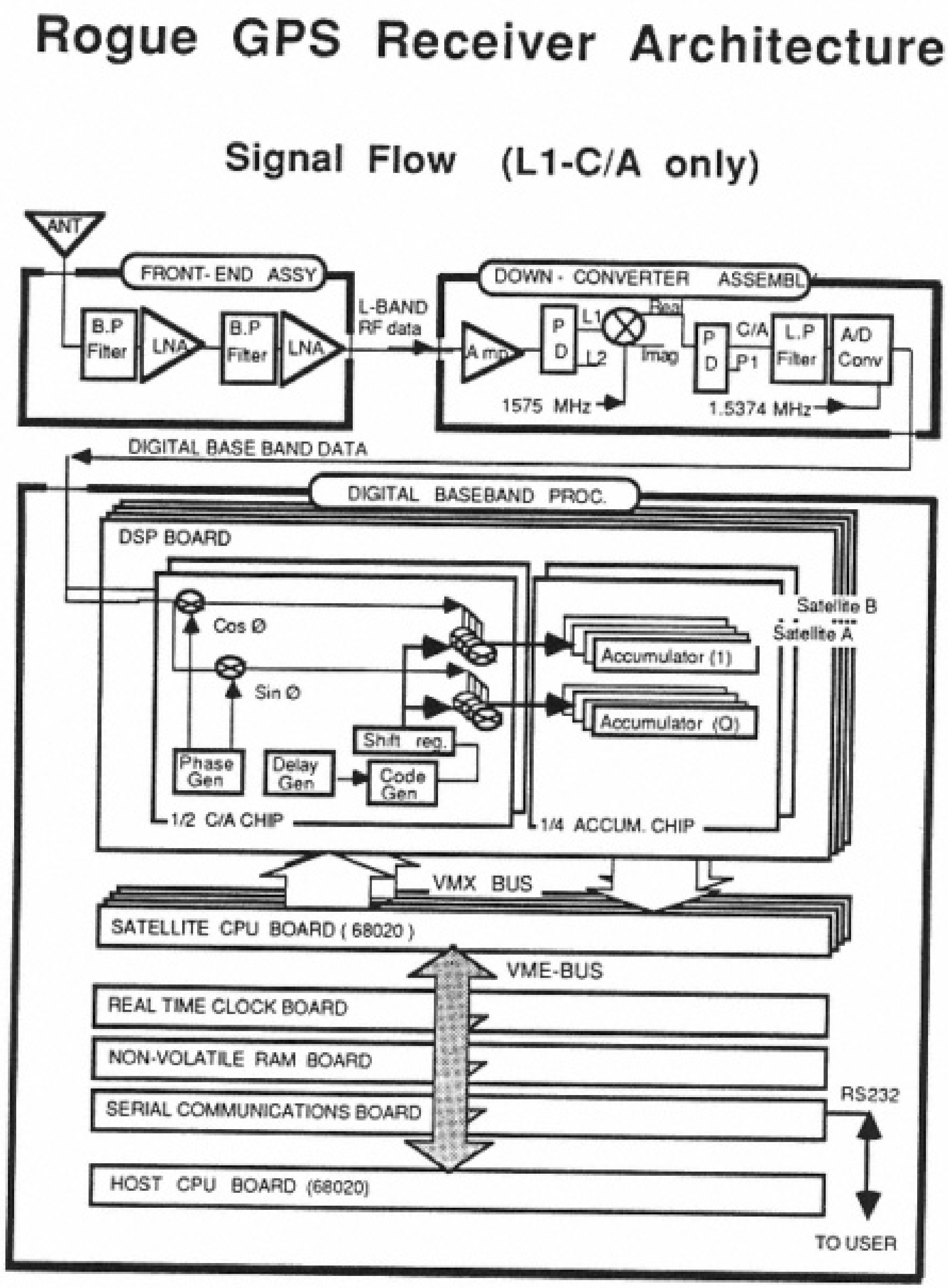


Fig. 9a).
Rogue GPS Receiver Architecture. Only the C/A signal processing is shown. In actuality there are five processing streams for C/A, L1, L2, P1, and P2.

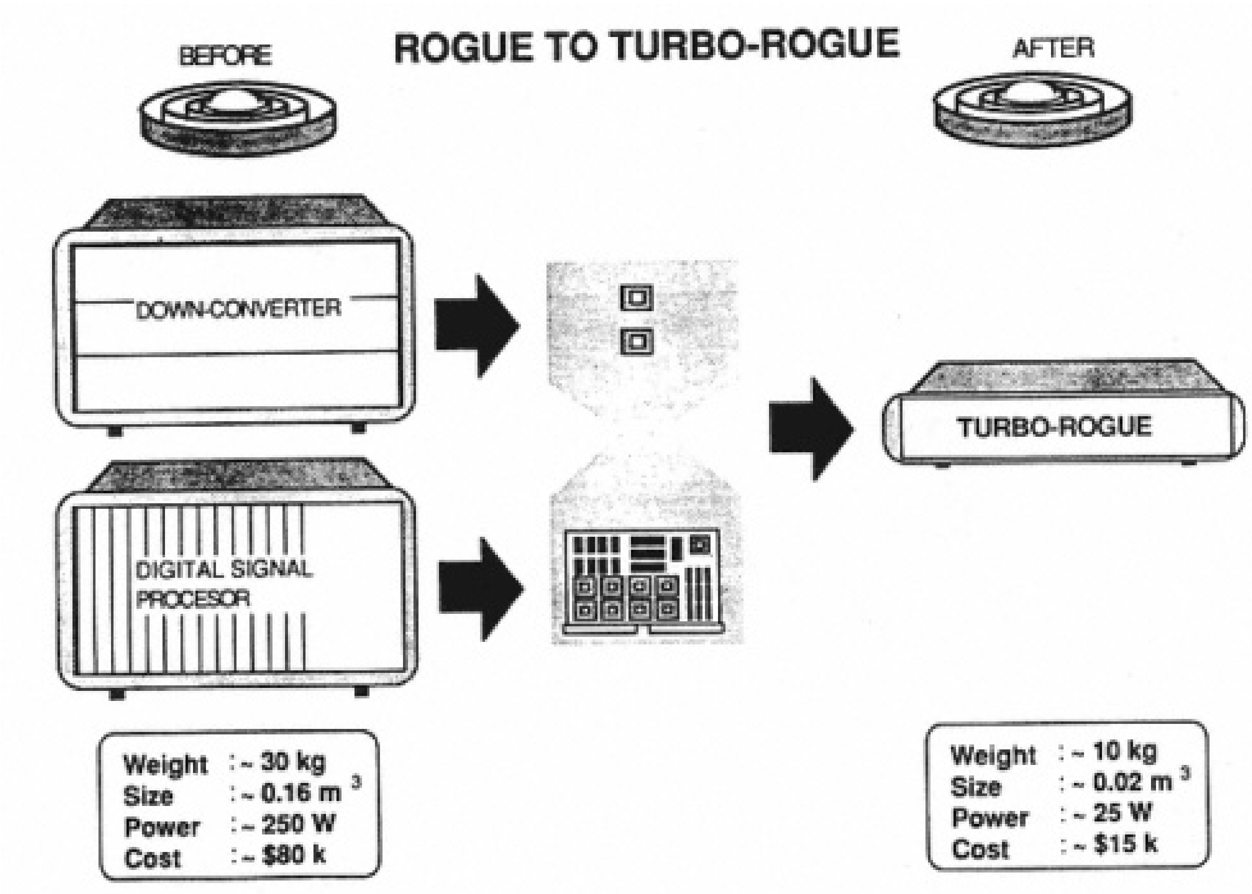


Fig. 9b).
The Rogue to Turbo-Rogue Transition utilizing the DFE chip for RF down conversion to baseband digital sampling and the turbo-chip for baseband signal processing.

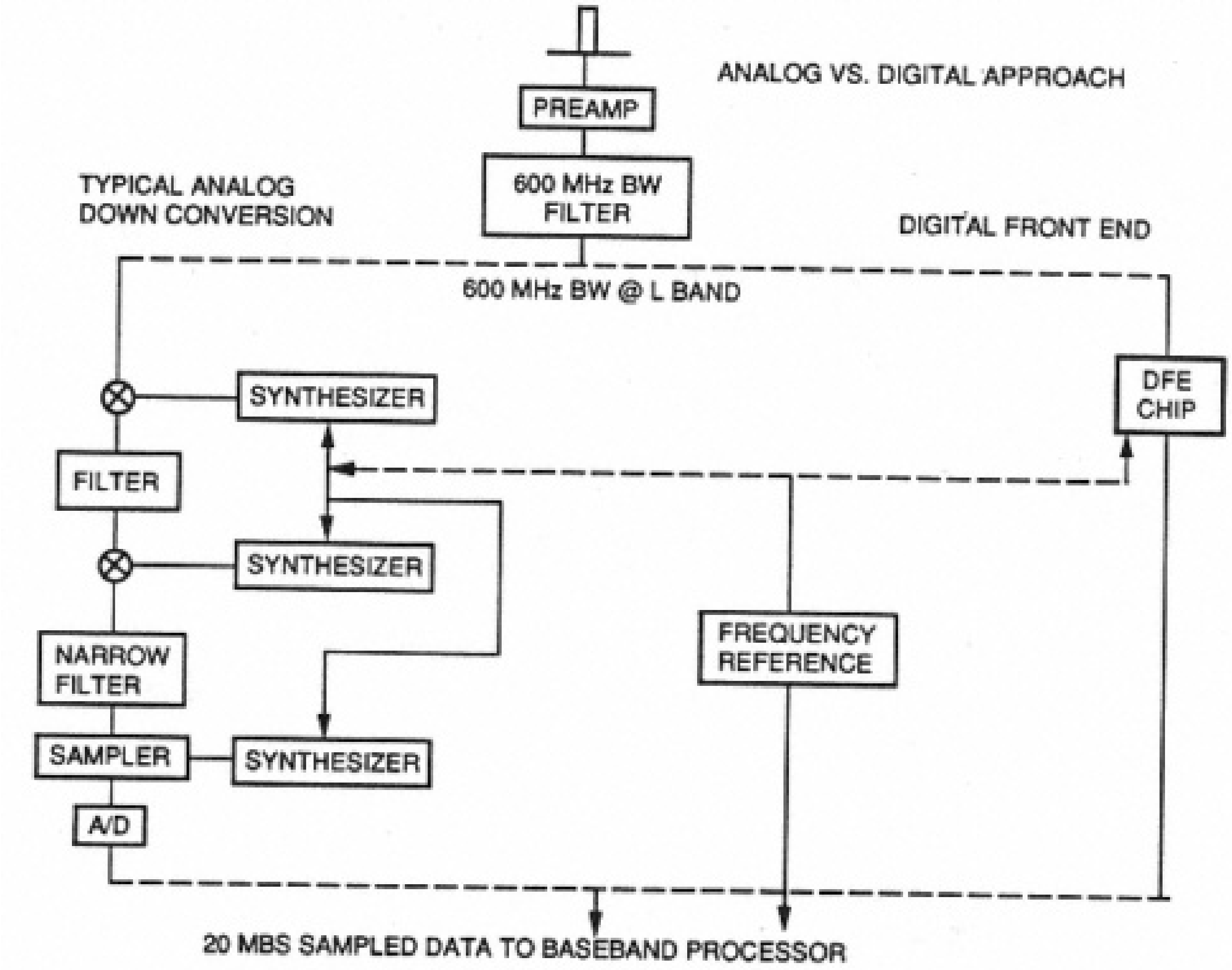
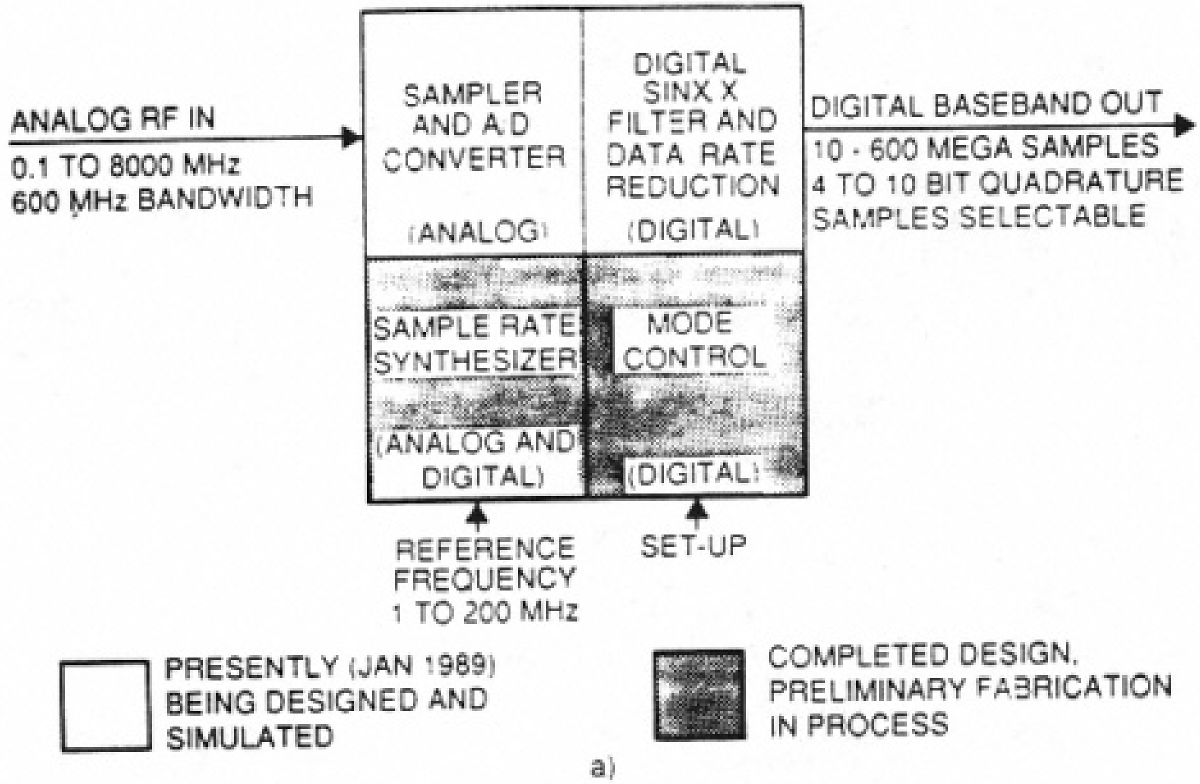


Figure 10.
Functional Schematic of an Analog Front End for RF Down-Conversion to Baseband, A/D Conversion, and Digital Sampling. Digital Front End Eliminates Analog Components that are Expensive and Sources of Error.



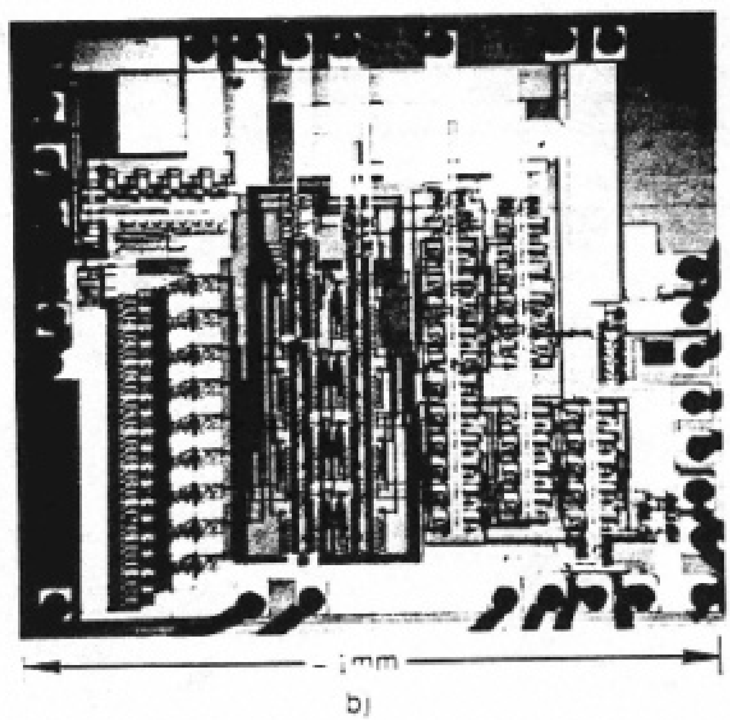


Figure 11.
a) Functional Layout of Digital Front End Chip for Converting RF Analog to Baseband Digital Samples. b) Photo of Completed Sections of the DFE Chip.

Figure 12.
Organization Diagram for an Internationally Operated Civilian GPS Global Tracking System for High Accuracy Applications

Internationally Operated Civilian
for High Accuracy Applications

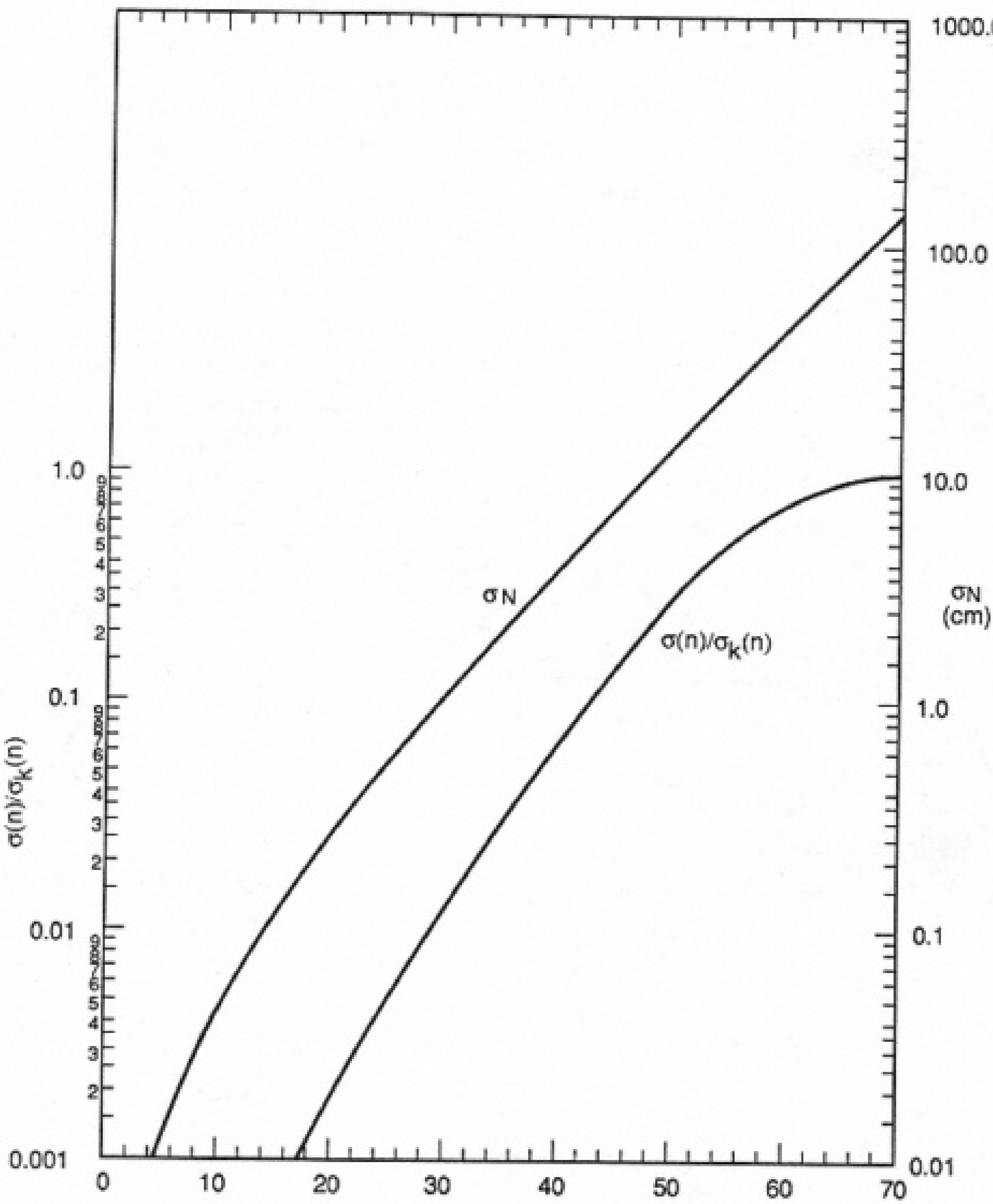


Figure 13.
Predicted Gravity Recovery from GP-B Using a TOPEX/Poseidon Class GPS Flight Receiver in Conjunction with a GPS Global Tracking System. Mission Duration is 6 Months. σ(n) is Based on Combined Tracking and A Priori Kaula Model Information. $\sigma_k(n)$ is Based on Kaula Model.