Gravitational Physics: Exploring the Structure of Space and Time

Committee on Gravitational Physics, National Research Council

ISBN: 0-309-51455-X, 126 pages, 6 x 9, (1999)

This PDF is available from the National Academies Press at: http://www.nap.edu/catalog/9680.html

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

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

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

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

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

Gravitational Physics

Exploring the Structure of Space and Time

Committee on Gravitational Physics
Board on Physics and Astronomy
Commission on Physical Sciences, Mathematics, and Applications
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

Front cover: Gravitational waves are ripples in the curvature of space and time that propagate with the speed of light through otherwise empty space. Mass in motion is the source of gravitational waves. The figure shows the predicted gravitational wave pattern from a pair of neutron stars or black holes spiraling inward toward a final merger. The figure shows one polarization of the waves as seen by observers stationed throughout the plane of the orbit at the moment of final merger. The waves measured far away were emitted during the earlier steady inspiral of the objects about one another, while the peak at the center comes from the final merger. The reception of gravitational waves in the next decade would not only confirm one of the most basic predictions of Einstein's general relativity, but also provide a new window on the universe. (Courtesy of Patrick R. Brady, Institute for Theoretical Physics, University of California at Santa Barbara, and the University of Wisconsin-Milwaukee.)

Printed in the United States of America

COMMITTEE ON GRAVITATIONAL PHYSICS

BOARD ON PHYSICS AND ASTRONOMY

ROBERT C. DYNES, University of California at San Diego, *Chair*
ROBERT C. RICHARDSON, Cornell University, *Vice Chair*
STEVEN CHU, Stanford University
VAL FITCH, Princeton University
IVAR GIAEVER, Rensselaer Polytechnic Institute
RICHARD D. HAZELTINE, University of Texas at Austin
JOHN HUCHRA, Harvard-Smithsonian Center for Astrophysics
JOHN C. MATHER, NASA Goddard Space Flight Center
R.G. HAMISH ROBERTSON, University of Washington
JOSEPH H. TAYLOR, Princeton University
KATHLEEN C. TAYLOR, General Motors Research and Development Center
J. ANTHONY TYSON, Lucent Technologies
GEORGE WHITESIDES, Harvard University

DONALD C. SHAPERO, Director
ROBERT L. RIEMER, Associate Director
KEVIN AYLESWORTH, Program Officer
JOEL R. PARRIOTT, Program Officer
NATASHA CASEY, Senior Administrative Associate
GRACE WANG, Senior Project Associate
MICHAEL LU, Project Assistant

iv

COMMISSION ON PHYSICAL SCIENCES, MATHEMATICS, AND APPLICATIONS

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 Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

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

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

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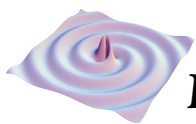
Preface

The Committee on Gravitational Physics (CGP) was organized by the National Research Council's (NRC's) Board on Physics and Astronomy (BPA) as part of the decadal survey *Physics in a New Era.* The committee's main charges were (1) to assess the achievements in gravitational physics over the last decade and (2) to identify the most promising opportunities for research in the next decade and describe the resources necessary to realize those opportunities. This report fulfills those charges.

As is made clear in the report, the field of gravitational physics has significant overlaps with astrophysics, elementary-particle physics, and cosmology, areas that have been or will be assessed by the NRC. Elementary-particle physics is the subject of a separate volume of the current physics survey, *Elementary-Particle Physics—Revealing the Secrets of Energy and Matter* (National Academy Press, Washington, D.C., 1998). Cosmology is discussed in *Cosmology: A Research Briefing* (National Academy Press, Washington, D.C., 1995). Astrophysical phenomena in which gravitation plays a key role were considered in the NRC study *A New Science Strategy for Space Astronomy and Astrophysics* (National Academy Press, Washington, D.C., 1997) and will be a part of the NRC's Astronomy and Astrophysics Survey now under way. Reports with overlapping content and emphases are to be expected because of emerging interdisciplinary areas of physics. Naturally, each of these reports makes its recommendations from the perspective of the subfield of physics involved. This report sets priorities and makes recommendations based on the committee's assessment of the impact of opportunities for research in gravitational physics.

As part of its task, the CGP reevaluated the estimates of the event rate for a number of sources of gravitational waves that might be received by the LIGO gravitational wave detector in the next decade in the light of current theoretical and observational understanding. These estimates are reported in the addendum to Section I of Chapter 3. The discussion given there should be regarded as the output of the entire committee, but we would be remiss if we did not also acknowledge that the detailed analysis is the work of three of us—Ramesh Narayan, Joseph Taylor, and David Spergel.

The CGP was helped in its tasks by input from many sources, some organized by the committee and some submitted by members of the gravitational physics community in response to various requests for input. The CGP's activities, in which the BPA staff headed by Don Shapero and Roc Riemer assisted greatly, are described in Appendix A.

The committee's work was supported by grants from the National Aeronautics and Space Administration, the National Science Foundation, and the U.S. Department of Energy. We thank them for this support.

James B. Hartle, *Chair*
Committee on Gravitational Physics
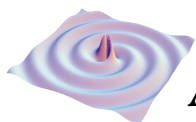
Acknowledgment of Reviewers

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

Mitchell C. Begelman, University of Colorado,
James E. Faller, University of Colorado,
J. Ross Macdonald, University of North Carolina at Chapel Hill,
Riley D. Newman, University of California at Irvine,
Kenneth Nordtvedt, Northwest Analysis,
Andrew Eben Strominger, Harvard University,
J. Anthony Tyson, Lucent Technologies,
Robert M. Wald, University of Chicago, and
Edward Witten, Princeton University.

Although the individuals listed above have provided many constructive comments and suggestions, the responsibility for the final content of this report rests solely with the authoring committee and the NRC.

Contents

xi

Executive Summary

Gravity is one of the four fundamental forces of nature. It is an immediate fact of everyday experience, yet it presents us with some of the deepest theoretical and experimental challenges in contemporary physics. Gravity is the weakest of the four fundamental forces, but, because it is a *universal* attraction between all forms of energy, it governs the structure of matter on the largest scales of space and time, including the structure of the universe itself. As one of the fundamental interactions, gravity is central to the quest for a unified theory of all forces, whose simplicity would emerge at very high energies or, equivalently, at very small distances.

Gravitational physics is thus a *two-frontier science*. On the large scales of astrophysics and cosmology it is central to the understanding of some of the most exotic phenomena in the universe—black holes, pulsars, quasars, the final destiny of stars, and the propagating ripples in the geometry of spacetime called gravitational waves. On the smallest scales it is concerned with the quantized geometry of spacetime, the unification of all forces, and the quantum initial state of the universe. Its two-frontier nature means that gravitational physics is a cross-disciplinary science overlapping astrophysics and cosmology on large scales and elementary-particle and quantum physics on small scales.

The theory that bridges this enormous range of scales is Einstein's 1915 general theory of relativity. The key ideas of general relativity are that gravity is the geometry of four-dimensional spacetime, that mass produces spacetime curvature while curvature determines the motion of mass, and that all freely falling bodies follow paths independent of their mass (an idea that is called the principle of equivalence).

When gravitational fields are weak and vary only slowly with time, the effects of general relativity are well approximated by Newton's 300-year-old theory of gravity. However, general relativity predicts qualitatively new phenomena when gravitational fields are strong, are rapidly varying, or can accumulate over vast spans of space or time. Black holes, gravitational waves, closed universes, and the big bang are some examples. Further, when the principles of classical general relativity are united with quantum theory, quantum uncertainties can be expected in the geometry of spacetime itself. The focus of modern gravitational physics has naturally been on exploring such relativistic and quantum phenomena.

ACHIEVEMENTS—A SHORT LIST

Gravitational physics is one of the oldest subjects in physics. Yet the expansion of opportunities in both experiment and theory has made it one of the most rapidly changing areas of science today. A short list of some of the important achievements of the past decade illustrates this point:

• The confirmation of the existence of gravitational waves by the observed shortening of the orbital period of a binary pulsar.

• The detection of the fluctuations in the cosmic background radiation (the light from the big bang) that are the origin of today's galaxies, stars, and planets.

• The development of a new generation of high-precision tests (to parts in a thousand billion) of the equivalence principle that underlies general relativity, and the verification of general relativity's weak-field predictions to better than parts in a thousand.

• The identification of candidate black holes in x-ray binary stars and in the centers of galaxies. Black holes are no longer a theorist's dream; they are central to the explanation of many of astronomy's most dramatic phenomena.

• The use of gravitational lensing as a practical astronomical tool to investigate the structure of galaxies and to search for the dark matter in the universe.

• The increasing use of large-scale numerical simulations to solve Einstein's difficult nonlinear equations. These simulations can predict the effects of strong gravity that will be seen in the next generation of observations of gravitational phenomena.

• The discovery of "critical phenomena" in gravitational collapse analogous to those that occur in transitions between different states of matter.

• The development of string theory and the quantum theory of geometry as promising candidates for the union of quantum mechanics and general relativity.

• The first descriptions of the quantum states of black holes.

• The development of powerful mathematical tools to study the physical regimes where Einstein's theory can break down.

OPPORTUNITIES

The Committee on Gravitational Physics (CGP) foresees that the transformation of the science of gravitational physics will accelerate in the next decade, driven by new experimental, observational, and theoretical opportunities. A single theme runs through the most important of these opportunities: *the exploration of strong gravitational fields*. Among the specific opportunities the CGP believes could be realized in the next decade if appropriate resources are made available are the following:

• The first direct detection of gravitational waves by the worldwide network of gravitational wave detectors now under construction.

• The first direct observation of black holes by the characteristic gravitational radiation they emit in the last stages of their formation.

• The use of gravitational waves to probe the universe of complex astronomical phenomena by the decoding of the details of the gravitational wave signals from particular sources.

• The continuing transformation of cosmology into a data-driven science by the wealth of measurements expected from new cosmic background radiation satellites, new telescopes in space and on the ground, and new systematic surveys of the large-scale arrangements of the galaxies.

• The first unambiguous determination of the basic parameters that characterize our universe, its age and fate, the matter of which it is made, how much of that matter there is, and the curvature of space on large scales.

• The unambiguous measurement of the value of the cosmological constant, with profound implications for our understanding of the fate of the universe, and also for particle physics and quantum gravity.

• The use of gamma-ray, x-ray, optical, infrared, and radio telescopes on Earth and in space to detect new black holes in orbit about companion stars and to explore the extraordinary properties of the geometry of space in the vicinity of black holes that are predicted by general relativity.

• The measurement of the dragging of inertial frames due to the rotation of Earth at the 1 percent level by the Gravity Probe B mission scheduled for launch in 2000.

• Dramatically improved tests of the equivalence principle that underlies general relativity.

• The understanding of the predictions of Einstein's theory in dynamical, strong-field, realistic situations through the implementation of powerful numerical simulations and sophisticated mathematical techniques untrammeled by weak-field assumptions, special symmetries, or other approximations.

• The development of current ideas in string theory and the quantum theory of geometry to achieve a finite, workable union of quantum mechanics, gravity, and the other forces of nature, potentially resulting in a fundamentally new view

of space and time. The application of this new theory to predict the outcome of black hole evaporation and the nature of the big bang singularity.

• The continued development within quantum gravity of a theory of the quantum initial condition of the universe capable of making testable predictions of cosmological observations today.

If these opportunities are realized, the CGP expects the next decade of research in gravitational physics to be characterized by (1) a much closer integration of gravitational physics with astrophysics, cosmology, and elementary-particle physics, (2) much larger experiments yielding much more data and requiring international collaboration, (3) a much closer relationship between theory and experiment, and (4) a much wider, more important role for computation in gravitational physics.

GOALS FOR GRAVITATIONAL PHYSICS
IN THE NEXT DECADE

In light of such opportunities, the CGP identified the following unordered list of highest-priority goals for gravitational physics:

- *Receive gravitational waves and use them to study regions of strong gravity.*
- *Explore the extreme conditions near the surface of black holes.*
- *Measure the geometry of the universe and test relativistic gravity on cosmological scales; explore the beginning of the universe.*
- *Test the limits of Einstein's general relativity and explore for new physics.*
- *Unify gravity and quantum theory.*

In making this list, the CGP assumed that the scientific objectives of a number of projects now under way will be achieved, e.g., Gravity Probe B, construction of the Laser Interferometer Gravitational-Wave Observatory (LIGO), the Chandra X-ray satellite, and the MAP cosmic background satellite. Although fully endorsed by the CGP, these projects do not appear in its recommendations.

RECOMMENDATIONS

The CGP makes several recommendations for reaching these goals. The four areas of recommended actions are listed in priority order, with the highest-priority area given first. The recommendations within each of the four categories have equal weight.

1. Gravitational Waves

The search for gravitational waves divides naturally into the high-frequency gravitational wave window (above a few hertz) accessible by experiments on Earth, and the low-frequency gravitational wave window (below a few hertz) accessible only from space. Both windows are important, and the CGP has not prioritized one over the other. The highest priority is to pursue *both* of these sources of information.

The High-Frequency Gravitational Wave Window

- *Carry out the first phase of LIGO scientific operations.*
- *Enhance the capability of LIGO beyond the first phase of operations, with the goal of detecting the coalescence of neutron star binaries.*
- *Support technology development that will provide the foundation for future improvements in LIGO's sensitivity.*

The Low-Frequency Gravitational Wave Window

- *Develop a space-based laser interferometer facility able to detect the gravitational waves produced by merging supermassive black holes.*

2. Classical and Quantum Theory of Strong Gravitational Fields

- *Support the continued development of analytic and numerical tools to obtain and interpret strong-field solutions of Einstein's equations.*
- *Support research in quantum gravity, to build on the exciting recent progress in this area.*

3. Precision Measurements

- *Dramatically improve tests of the equivalence principle and of the gravitational inverse square law.*
- *Continue to improve experimental testing of general relativity, making use of available technology, astronomical capabilities, and space opportunities.*

4. Astronomical Observations

The astronomical observations recommended below have strong arguments for support from astronomy and astrophysics. The ones listed are those that the CGP expects will have the greatest impact on *gravitational* physics in the next decade.

- *Use gamma-ray, x-ray, optical, infrared, and radio telescopes on Earth and in space to study the environment near black holes.*
- *Measure the temperature and polarization fluctuations of the cosmic background radiation from arcminute scales to scales of tens of degrees.*
- *Search for additional relativistic binary systems.*
- *Launch all-sky gamma-ray and x-ray burst detectors capable of detecting the electromagnetic counterparts to LIGO events.*
- *Use astronomical observations of supernovae and gravitational lenses to infer the distribution of dark matter and to measure the cosmological constant.*

If these recommendations are implemented, the CGP believes that the next decade in gravitational physics could see as significant a transformation of the field as occurred in the late 1960s and early 1970s. This transformation will take the subject further into the arena of strong gravitational fields, with stronger coupling from experiment than ever before, leading to a deeper understanding of the central place of gravitational physics in resolving the fundamental questions of contemporary physics.

Chapter 1
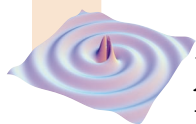
Introduction, Overview, and Recommendations

I. GRAVITATION: A TWO-FRONTIER SCIENCE

Of the four fundamental forces of nature, gravity has been studied the longest, yet gravitational physics is one of the most rapidly changing areas of science today. Gravity is an immediate fact of everyday experience, yet presents us with some of the deepest theoretical and experimental challenges of contemporary physics. Gravitational physics has given us some of the most accurately tested principles in the history of science, yet gravitational waves—one of its most basic predictions—have never been detected by a receiver on Earth. Gravitational physics is concerned with some of the most exotic phenomena in the universe—black holes, pulsars, quasars, the big bang, the final destiny of stars, gravitational waves, the microscopic structure of space and time, and the unification of all forces—challenges to understanding that have captured the imaginations of physicists and lay persons alike. Yet gravitational physics is also concerned with the minute departures of the motion of the planets from the laws laid down by Newton, and is a necessary ingredient in the operation of the Global Positioning System used every day. The challenges of gravitational physics have been the central concerns of some of the most famous 20th-century scientists—Albert Einstein, S. Chandrasekhar, Robert Dicke, Stephen Hawking, and Roger Penrose to mention just a few examples. As the Committee on Gravitational Physics (CGP) outlines below, the past decade has seen major achievements in gravitational physics. The next decade promises to be even more exciting, yielding revolutionary insights. This report reviews past accomplishments in the emerg-

7

ing field of gravitational physics, describes opportunities for future research, and recommends priorities for the most promising of these.

Gravity is the weakest of the four fundamental forces. The gravitational force between the proton and electron is 10^{40} (1 followed by 40 zeros) times smaller than the electric force that binds these particles together in atoms. However, gravity is a *universal* force. *All* forms of matter and energy attract each other gravitationally, and that interaction is unscreened—there is no negative "gravitational charge" to cancel the attraction. It is therefore gravity that governs the structure of matter on the largest scales of space and time and thus the structure of the universe itself. Gravity is also central to the quest for a unified theory of all forces whose simplicity would emerge at very high energies or very small distances. Gravity is the last force to be included in contemporary unified theories, yet many of the ideas for these "final theories" come from gravitational physics. Indeed, it would not be an exaggeration to say that many frontier problems in elementary-particle physics originate in gravitational physics.

Gravitational physics is thus a *two-frontier science*. Its important applications lie on both the very largest and the very smallest distance scales that are considered in today's physics. (See Figure 1.1.) On the largest scales, gravity is linked to astrophysics and cosmology. On the smallest scales, it is tied to elementary-particle and quantum physics. These frontiers are not disjoint; they become one in the early universe at the time of the big bang where the whole of today's observable universe was compressed into a minuscule volume.

II. ACHIEVEMENTS OF THE PAST DECADE

The theory of gravity proposed by Isaac Newton more than 300 years ago provided a unified explanation of how objects fall and how planets orbit the Sun. But Newton's theory is not consistent with Einstein's 1905 principle of special relativity. In 1915, Einstein proposed a new, relativistic theory of gravity—general relativity. When gravity is weak—for example, on Earth or elsewhere in the solar system—general relativity's corrections to Newton's theory are tiny. But general relativity also predicts new strong-gravity phenomena such as gravitational waves, black holes, and the big bang that are quantitatively and qualitatively different from those accounted for in Newtonian gravity. Modern gravitational physics focuses on these new phenomena and on high-precision tests of general relativity.

The basic formulation of general relativity was complete in 1915 and was almost immediately confirmed by tests in the solar system—the precession of the orbit of Mercury and the bending of light by the Sun. Over the ensuing decades theoretical analyses deepened the understanding of the theory and exhibited the richness and variety of its predictions. But, except perhaps for cosmology, the theory had little observational impact until the middle 1960s. Then, developments on several different fronts led to a renaissance in gravitational physics that
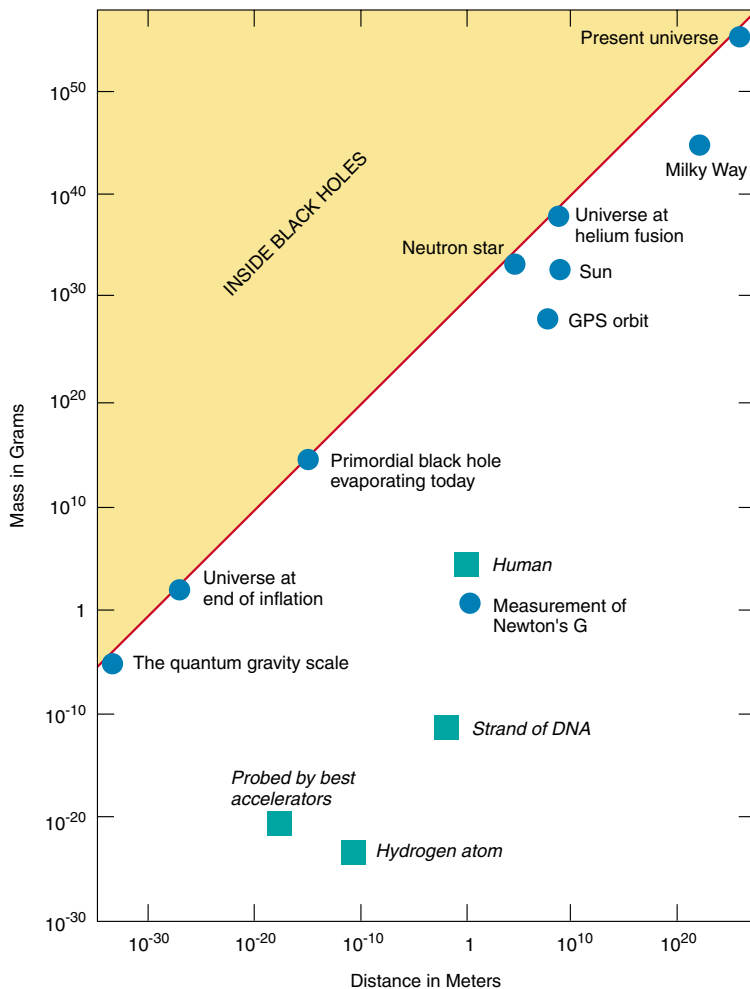
FIGURE 1.1 Gravitational physics deals with phenomena on scales of distance and mass ranging from the microscopic to the cosmic—the largest range of scales considered in contemporary physics. There are phenomena for which relativistic gravity is important over this whole range of scales. Representative ones are indicated by filled circles; other illustrative phenomena in which gravitation plays little role are shown by filled squares and italics. Phenomena above the diagonal line are unobservable, because they take place inside black holes. Phenomena close to the diagonal line are in the strong-gravity regime. The largest scales are the frontier of astrophysics; the smallest, of elementary-particle physics. Scales referring to the universe at various moments in its history denote the size of the volume light could travel across since the big bang, and the mass inside that volume, if the universe always had the expansion rate it had at that moment.

continues today. First, the discoveries of pulsars, quasars, and galactic x-ray sources revealed for the first time astrophysical phenomena for whose understanding relativistic gravity was essential. At the same time, the theory was subjected to increasingly varied, accurate, detailed, and systematic tests of its predictions for the weak gravitational field of the solar system. General relativity emerged from these tests confirmed in a wide domain. Today it is the only serious contender for a classical relativistic theory of gravity. Indeed, in certain areas of physics, the curvature of spacetime has become a realistic concern or a tool to be exploited. Examples include accounting for the effects of spacetime curvature in the operation of the Global Positioning System, correcting for the bending by the Sun of the light from quasars used to precisely monitor the rotation of Earth, the use of gravitational lenses to measure the properties of galaxies and cosmological parameters, and the use of general relativity to measure the masses of binary neutron stars.

While these astrophysical and experimental developments were taking place on large length scales, progress toward relativistic gravity was being made at the smallest distances considered by physics. The concerns of elementary-particle physics were moving to higher and higher energies, or equivalently to shorter and shorter distances—another regime where relativistic gravity is important. Progress was made toward a unified theory of the strong, electromagnetic, and weak forces. Gravitational physics became the next frontier of particle physics, and the unification of gravity with quantum mechanics and the other forces of nature is today a major challenge of theoretical physics.

The past decade saw many achievements in gravitational physics. Any short list of highlights would include the following:

• The confirmation of the existence of gravitational waves by the detailed analysis of the shortening of the orbital period of the Hulse-Taylor binary pulsar, showing that the radiated power in gravitational waves agrees with the prediction of general relativity to within a third of a percent. The 1993 Nobel Prize in physics was awarded to Russell Hulse and Joseph Taylor for discovering this pulsar system.

• The accurate measurements of the cosmic background radiation—the omnipresent light from the hot big bang that has cooled to a little under 3 degrees above absolute zero in the subsequent expansion of the universe. The observations verified detailed predictions of the character of the radiation from the hot big bang. They also revealed for the first time the tiny fluctuations that arose from minute early irregularities that grew under the attractive force of gravity to become the galaxies, stars, and planets of today. These measurements have given scientists the most detailed picture of the early universe yet available.

• The development of a new generation of high-precision tests (to parts in a thousand billion) of the equivalence principle that underlies general relativity, and the verification of general relativity's weak-field predictions to better than

parts in a thousand. The new techniques provide high sensitivity to interactions that violate the equivalence principle with ranges from infinity down to a centimeter, and sharply constrain speculations in particle and cosmological physics.

• The identification of candidate black holes in two major classes of astronomical objects: double stars called x-ray binaries, where black hole candidates of a few solar masses have been found, and the centers of galaxies, where compact objects with masses up to a billion solar masses or more have been discovered. Black holes are no longer a theorist's dream; they are central to the explanation of many of astronomy's most dramatic phenomena.

• The use of gravitational lensing as a practical astronomical tool to investigate the structure of galaxies and galactic clusters, and to search for dark matter in the universe. Thus, one of the first experimental verifications of general relativity—the deflection of light by mass—was put to practical use.

• The increasing use of large-scale numerical simulations to solve Einstein's difficult nonlinear equations. These simulations can predict the effects of strong gravity that will be seen in the next generation of experiments.

• The use of numerical simulations of gravitational collapse to discover "critical phenomena" associated with the onset of black hole formation. These critical phenomena are analogous to those that occur in transitions between different states of matter.

• The development of string theory and the quantum theory of geometry as promising candidates for a finite, workable theory that unifies quantum mechanics and general relativity.

• The first descriptions, in the above theories, of the quantum states of black holes. The demonstration within string theory that the topology of space can change. The analysis, without recourse to weak-field approximations, of quantum gravity effects in the context of the quantum theory of geometry.

• The development of powerful mathematical tools to study the physical regimes where Einstein's theory can break down. Under special assumptions, it was shown that this can occur only at an initial big bang, inside a black hole, or at a final "big crunch," thus supporting the cosmic censorship conjecture that these are the *only* places where the theory breaks down.

In addition to these scientific achievements, the past decade saw the start or continuation of experimental projects whose results will shape the field in the next decade. Notable were the final preparation of the Gravity Probe B mission to measure the minute twisting of the spacetime geometry ("dragging of inertial frames" effect) caused by Earth's rotation, and the start of construction for the Laser Interferometer Gravitational-Wave Observatory (LIGO) and other large-scale gravitational wave detectors. These gravitational wave receivers will open a new window on the universe by being sensitive enough to see the gravitational waves expected to be produced by astrophysical sources.

III. OPPORTUNITIES FOR THE NEXT DECADE

The transformation of the science of gravitational physics will accelerate in the next decade, driven by new experimental, observational, and theoretical opportunities. It would therefore be most accurate to think of gravitational physics as an emerging new area of physics despite its long history. In subsequent sections the CGP discusses many exciting opportunities, but a single theme runs through most of them: *the exploration of strong gravitational fields.* Until now our direct evidence of general relativity has been through weak-field effects in the solar system and ground-based experiments. To be sure, physicists have convincing evidence for strong gravitational effects such as black holes and the big bang, but in nothing like the detail expected in the next decade.

In the following the CGP lists opportunities that *could* be realized in the next decade. Whether these opportunities *will* be realized depends largely on the availability of funding, and on the fortunes of observational and theoretical discovery.

• The first direct detection of gravitational waves by the worldwide network of gravitational wave detectors now under construction.

• The first direct observation of black holes by the characteristic gravitational radiation they emit in the last stages of their formation.

• The use of gravitational waves to probe the universe of complex astronomical phenomena by the decoding of the details of the gravitational wave signals from particular sources.

• The continuing transformation of cosmology into a data-driven science by the wealth of measurements expected from new cosmic background radiation satellites, new telescopes in space and on the ground, and new systematic surveys of the large-scale arrangements of the galaxies.

• The first unambiguous determination of the basic parameters that characterize our universe, its age and fate, the matter of which it is made, how much of that matter there is, and the curvature of space on large scales.

• The unambiguous measurement of the value of the cosmological constant, with profound implications for our understanding of the fate of the universe, and also for particle physics and quantum gravity.

• The use of gamma-ray, x-ray, optical, infrared, and radio telescopes on Earth and in space to detect new black holes in orbit about companion stars and to explore the extraordinary properties of the geometry of space in the vicinity of black holes that are predicted by general relativity.

• The measurement of the dragging of inertial frames due to the rotation of Earth at the 1 percent level by the Gravity Probe B mission scheduled for launch in 2000.

• Dramatically improved tests of the equivalence principle that underlies general relativity.

• The understanding of the predictions of Einstein's theory in dynamical, strong-field, realistic situations through the implementation of powerful numerical simulations and sophisticated mathematical techniques untrammeled by weak-field assumptions, special symmetries, or other approximations.

• The development of current ideas in string theory and the quantum theory of geometry to achieve a finite, workable union of quantum mechanics, gravity, and the other forces of nature, potentially resulting in a fundamentally new view of space and time. The application of this new theory to predict the outcome of black hole evaporation and the nature of the big bang singularity.

• The continued development within quantum gravity of a theory of the quantum initial condition of the universe capable of making testable predictions of cosmological observations today.

If these opportunities are realized, the CGP expects the next decade of research in gravitational physics to be characterized by the following features:

• *A much closer integration of gravitational physics with other areas of science.* On the frontier of the largest scales the CGP expects gravitational physics to become increasingly integrated with astrophysics and cosmology as more phenomena for which relativistic gravity is important become accessible to detailed observation and theoretical analysis. This will be ensured by the new data from the worldwide network of gravitational wave detectors now under construction, from the cosmic background radiation satellites now planned, and from new gamma-ray, x-ray, optical, infrared, and radio telescopes on Earth and in space. The CGP expects these phenomena to yield increasingly accurate tests and demonstrations of strong-field gravitational theory. On the frontier of the smallest scales the committee expects the integration of quantum gravity with elementary-particle physics to continue. Gravity is a key ingredient in any unified theory of all forces, and conversely that unified theory is one source of a manageable theory of quantum gravitational phenomena.

• *Much larger experiments yielding much more data.* Again the ground-based gravitational wave detectors now under construction are enough to ensure this. Gravitational wave detectors and other experiments in space will only accelerate the trend. International collaborations are likely to be required to realize the full potential of these experimental possibilities.

• *A much closer relationship between theory and experiment.* The experiments now under way require theoretical analysis at a level of detail, depth, and coordination only now being appreciated. The CGP expects that the next decade will see the emergence of a new cadre of gravitational phenomenologists focused on using fundamental theory to analyze data from experiment.

• *A much wider, more important role for computation in gravitational physics.* Understanding actual phenomena requires realistic solutions to Einstein's equation incorporating realistic properties of the matter (fluid, gas) sources. This

means large-scale numerical simulations carried out by teams of theorists employing state-of-the-art computers.

Chapter 2 of this report contains a brief description of general relativity and key phenomena in gravitational physics. In Chapter 3 the CGP analyzes the achievements of the past and opportunities for the future in gravitational waves, black holes, cosmology, testing general relativity, and quantum gravity. The CGP's recommendations arising from this analysis of the most promising scientific opportunities to pursue are described immediately below.

IV. GOALS AND RECOMMENDATIONS FOR GRAVITATIONAL PHYSICS

The scientific opportunities summarized above and described in detail in Chapter 3 are many and varied. In this section, the CGP sets out what it believes are the highest-priority goals for gravitational physics in the next decade and makes recommendations on how to achieve these goals.

Basis for the Goals and Priorities

The CGP based its goals and priorities for gravitational physics on its assessment of the scientific impact on the field that would follow from realizing these goals in the next decade. The committee has not shrunk from the challenge of making these assessments across the entire subject of gravitational physics. Thus expensive efforts (e.g., gravitational wave detectors) are prioritized along with inexpensive ones (e.g., theoretical research in quantum gravity). The reader wishing to construct sublists, of expensive projects for example, should have no difficulty doing so.

In this discussion the CGP assumes that the scientific objectives of a number of projects now under way will be achieved. These are the Gravity Probe B experiment (now with a definite launch window in 2000), construction of the LIGO gravitational wave detector (now nearing completion in time for an initial data run starting in 2002), the Chandra X-ray satellite, which was launched in July 1999, and the MAP cosmic background satellite currently under construction, to be launched late in 2000. Although fully endorsed by the CGP, these projects do not appear in its recommendations.

The CGP focused on assessing the *scientific* opportunities that will be presented by the next decade. It did not attempt a detailed assessment of the technical readiness of any of the large projects proposed. That task should be undertaken by appropriate committees at appropriate junctures. The CGP does not therefore mention by name specific unapproved projects that have been proposed for realizing these scientific opportunities. Rather, it describes the important scientific goals and measurement objectives.

As can be expected in a science as cross-disciplinary as gravitational physics, many of the projects entering into the CGP's priorities have strong arguments for support from related areas of physics and astronomy. These arguments were taken into account, but the CGP's list of priorities reflects its view of the projects' potential impact on gravitational physics.

Goals

The CGP believes that the most important goals for gravitational physics are those on the following *unordered* list. These goals constitute the CGP's long-term vision for the field. The list is ambitious. Some of these goals could take longer than a decade to realize depending on the availability of funding, the adequacy of technology development, and the fortunes of observational and theoretical discovery.

- *Receive gravitational waves and use them to study regions of strong gravity.*

Study of the Hulse-Taylor binary pulsar 1913+16 proved that gravitational waves exist, but the discovery is still incomplete, in the same way that neutrinos needed to be detected even after their existence was proved from the study of beta decay. Reception of gravitational waves will allow the precise comparison of their properties with those predicted by general relativity. However, beyond these tests, gravitational waves provide a window into regions of strong and rapidly varying gravity in the universe that are largely invisible using electromagnetic signals. The strongest waves come from the most extreme and catastrophic events in the universe and can provide important clues to the nature of those events. Supernova explosions, stellar and black hole collisions, and the big bang are all examples. Gravitational waves can provide unique signatures for the existence of black holes. They can also be used to test the validity of general relativity. A worldwide network of gravitational wave observatories is poised to begin exploiting this new astronomical window.

- *Explore the extreme conditions near the surface of black holes.*

Astronomers have discovered black hole candidates with masses several times that of the Sun in binary star systems, and up to a billion times larger in the centers of galaxies. Radio, optical, x-ray, and gamma-ray observations of the candidates imply dense concentrations of matter in very small regions of space. Einstein's equations of relativity, together with our understanding of the properties of matter, do not allow any viable interpretation of the observations other than that the objects are black holes. However, there is not yet direct confirmation of the black hole nature of the candidates. Much can be learned from the

detailed study of the environment near the black hole surface, using electromagnetic observations. Gravitational waves from the damped vibrations of newborn or disturbed black holes can supply even better probes.

- *Measure the geometry of the universe and test relativistic gravity on cosmological scales; explore the beginning of the universe.*

General relativity, together with the observation that the universe is expanding, implies that the universe began in a big bang. Yet the overall geometry, material content, and ultimate fate of the universe are still open questions. A number of astronomical tools, including observations of the microwave background radiation, studies of distant supernovae, measurements of the large-scale distribution of galaxies, and studies of gravitational lenses, provide ways to measure the geometry of the universe and the value of the cosmological constant. Extending our understanding of physics to enormous densities, temperatures, and curvatures in the earliest moments of the universe is one of the great challenges of theoretical physics. Not only is a quantum theory of gravity needed, but also a theory of the universe's quantum initial state.

- *Test the limits of Einstein's general relativity and explore for new physics.*

General relativity has passed all experimental tests performed to date. Yet there is now a growing expectation that deviations from the predictions of pure general relativity will occur at some level, mediated by interactions of hitherto unseen elementary particles. High-precision experiments in terrestrial laboratories or in space can place constraints on such interactions, and possibly detect them.

- *Unify gravity and quantum theory.*

Despite the outstanding success of general relativity, this theory is not able to describe the strongest gravitational fields in the universe, such as the earliest moments of the big bang, or the ultimate fate of a star that collapses to form a black hole. To describe these situations, a quantum extension of Einstein's theory is needed. The significant progress over the past decade has given hope that this long-sought theory may soon be completed. Its implications—from cosmology and black hole physics, to a new understanding of space and time, to a possible unification of all of the known forces and particles in nature—are enormous.

Recommendations

Listed below are the CGP's specific recommendations for research in the next decade to reach these goals. This list of four is ordered with the highest-priority area of recommended actions given first. The recommendations within each of the four categories have equal weight.

1. Gravitational Waves

As is described in more detail in Chapter 3, the search for gravitational waves divides naturally into the high-frequency gravitational wave window (above a few hertz) accessible by experiments on Earth, and the low-frequency gravitational wave window (below a few hertz) accessible only from space. The CGP did not attempt to prioritize one of these windows over the other. Both are important. While there are perhaps more currently known sources accessible from the low-frequency window, the high-frequency window is the one that most clearly will open up in the next decade. The highest priority is to pursue *both* of these sources of information.

The High-Frequency Gravitational Wave Window

- *Carry out the first phase of LIGO scientific operations.*
- *Enhance the capability of LIGO beyond the first phase of operations, with the goal of detecting the coalescence of neutron star binaries.*
- *Support technology development that will provide the foundation for future improvements in LIGO's sensitivity.*

The main U.S. opportunity for the direct detection of gravitational waves in the next decade lies in the Laser Interferometer Gravitational-Wave Observatory. The LIGO detectors are sensitive to waves with frequencies of several kilohertz down to 50 Hz in their initial data run, extending downward toward 10 Hz after they are upgraded. In the high-frequency window are several candidate sources of gravitational waves whose detection would contribute important new astronomical and physical information. These include inspiraling and merging binary systems of black holes or neutron stars, gravitational collapse of stellar cores in supernova events, unstable oscillations of newly formed neutron stars, and a random background of waves, possibly from processes in the early universe. The discovery of waves from binary neutron star inspirals can reveal information about the nature of matter at supernuclear densities and could shed light on the origin of gamma-ray bursts, while waves from merging double black holes could show how event horizons coalesce and provide proof of their existence. Detection of gravitational waves from pulsars would reveal whether or not their surfaces are distorted and provide key clues as to their internal structure.

The CGP recommends support for the initial operation of LIGO. It recommends support for *sustained* development of the technology necessary to upgrade LIGO to a sensitivity necessary to detect neutron star binary coalescences. In particular the CGP supports the mid-decade enhancement of LIGO's sensitivity by reasonable extrapolations of existing technology. (See Table 1.1.) If further improvements by deployment of new technology involve a large increase in costs, the CGP recommends that the project be reviewed when the funding step is required. The review should consider developments in detector sensitivity, detected sources, and current astrophysical understanding.

The Low-Frequency Gravitational Wave Window

• *Develop a space-based laser interferometer facility able to detect the gravitational waves produced by merging supermassive black holes.*

Gravitational waves below a few hertz provide a window on the universe that is different from that studied by LIGO, much as the universe seen in radio waves differs from that seen in visible light. Seismic noise makes this low-frequency window inaccessible to ground-based observations; observations from space are required. In this window we should be able to detect gravitational waves from known binary stars and from the merging and formation of supermassive black holes, and to search for waves from the earliest moments following the big bang. The detection of any of these gravitational waves will meet important goals in both physics and astronomy. In astronomy the low-frequency gravitational wave window offers the possibility of detecting objects that can be seen in no other way, such as supermassive black holes; probing the interiors of some of the most energetic events in the universe, such as those occurring in quasars and active galactic nuclei; and investigating the collisions of galaxies in epochs close to the time of their formation. In physics, observations in this window would allow precision tests of the properties of gravitational waves, tests of strong-field theories of the production of these waves, detailed confirmations of the predicted properties of black holes in general relativity, and observational tests of the theory of gravitational collapse. Limits on the gravitational waves from the big bang would constrain the physics of the fundamental interactions at the ultrahigh energies realized in the early universe. For these reasons the CGP supports the development of key technologies aiming at the deployment of such an interferometer late in the first decade of the 21st century.

19

TABLE 1.1 Projected LIGO Development Stages and Funding Action Recommended by the Committee on Gravitational Physics

Stage	Description	Approximate Start Date	Maximum Distance at Which NS-NS Mergers Will Be Detectable (in megaparsecs)	NS-NS Merger Event Rate[a]	CGP Recommendation
Initial operation	Initial data run	2002	15-20 Mpc	None expected	Support
Mid-decade upgrade	Improved sensitivity by reasonable extrapolations of existing technology (e.g., high-Q test masses,[b] low-loss optics)	2004-2005	100-200 Mpc	One per year (200 Mpc)	Support
Late-decade upgrade	New technologies (e.g., cryogenic masses, new materials)	2008 (?)	400-600 Mpc	One per month (600 Mpc)	Support research and development; review proposed improvements before implementation

[a]The CGP's estimates of the rate for neutron star-neutron star (NS-NS) mergers are presented in the addendum to Section I of Chapter 3.
[b]High-Q test masses have a high mechanical quality factor and low energy dissipation.

2. Classical and Quantum Theory of Strong Gravitational Fields

• *Support the continued development of analytic and numerical tools to obtain and interpret strong-field solutions of Einstein's equations.*

Full exploitation of the opportunities of the coming decade in strong-field gravitational physics will require a deeper grasp of the underlying theory than we currently possess. A detailed understanding of solutions to Einstein's equations will be necessary to understand regions of strong gravity, including black holes. In addition to analytic techniques, computational approaches will be essential, because Einstein's equations are too complicated to yield entirely to analytic methods.

The investigation of gravitational theory should be two-pronged. On the one hand, gravitational physicists must aim at achieving a fundamental understanding of general relativity. On the other, there are important astrophysically relevant questions that need reliable answers if the goals outlined above are to be fully met. The most urgent of these is to understand quantitatively the outcome of black hole and neutron star collisions. If these calculations are to supply predictions of the gravitational waves produced by such events by the time LIGO is on line, an expanded effort is required, including adequate human resources and increased access to supercomputer facilities.

• *Support research in quantum gravity, to build on the exciting recent progress in this area.*

The most fundamental questions about space, time, the nature of the big bang, and the interior of black holes cannot be answered within classical general relativity. They require a quantum theory of gravity. Recent developments in string theory and the quantum theory of geometry have brought us closer to constructing this new fundamental physical theory. It is already possible to answer some long-standing questions about the quantum properties of black holes. In the next decade, the CGP recommends a concerted effort to complete the construction of this new theory, possibly by combining some of the ideas in these two approaches. Applications of this theory to the nature of the very early universe should be explored, possibly resulting in modifications of our current ideas about a very early epoch of ultrarapid expansion of the universe (inflation) and shedding light on its initial state.

The potential implications of such a theory are extraordinary. At the microscopic level, it may be necessary to abandon such basic notions as the spacetime continuum, causality, and locality. The universe may have extra "hidden" spatial dimensions. Fundamental entities may be extended objects like strings rather than point particles.

Much of the recent progress in quantum gravity has occurred through a confluence of ideas from gravitational physics, elementary-particle physics, and mathematics. Fostering close contacts between these communities (for example through joint research, conferences, and schools) is vital for continued progress.

3. Precision Measurements

- *Dramatically improve tests of the equivalence principle and of the gravitational inverse square law.*

The equivalence principle is one of the foundations of general relativity, and any violation requires new physical interactions that could also modify the inverse square law, which is satisfied by general relativity in its Newtonian limit. Quantum theories of gravity, as well as some cosmological theories, could produce apparent violations of the principle at some level. New experiments carried out in terrestrial laboratories and in space can improve the precision, explore much shorter length scales, and test the effects of exotic forms of matter.

- *Continue to improve experimental testing of general relativity, making use of available technology, astronomical capabilities, and space opportunities.*

Modest investments in promising laboratory techniques or space missions could yield important improvements in experimental tests that could probe the limits of general relativity. Examples include continued lunar laser ranging, placing high-precision clocks on satellites, tracking of Earth-orbiting and interplanetary spacecraft, and binary pulsar observations.

4. Astronomical Observations

The astronomical observations recommended below have strong arguments for support from astronomy and astrophysics. The ones listed are those that the CGP expects will have the greatest impact on *gravitational* physics in the next decade.

- *Use gamma-ray, x-ray, optical, infrared, and radio telescopes on Earth and in space to study the environment near black holes.*

Such observations provide important insights into the extreme environments from which a broad spectrum of radiation is emitted and can potentially pin down basic properties of black hole candidates, such as their masses and spins. The observations may also lead to definitive proof of the black hole nature of the objects.

• *Measure the temperature and polarization fluctuations of the cosmic background radiation from arcminute scales to scales of tens of degrees.*

Microwave background observations measure variations in spacetime, from the scale of galaxies up to the scale of the visible universe. These ripples can be due either to fluctuations in the density of the universe or to gravitational waves with wavelength comparable to the size of the universe. The COBE satellite detected these temperature fluctuations at the largest angular scales. In the future, the MAP and Planck satellites, and ground-based and balloon-based experiments, will map these fluctuations at finer scales. Gravitational waves and density fluctuations also generate polarization fluctuations whose amplitude is expected to be a few percent of the temperature fluctuations. Observations of these polarization fluctuations could lead to the detection of a stochastic background of gravitational waves from the early universe.

• *Search for additional relativistic binary systems.*

Astronomers have detected only perhaps a percent of the pulsars in our Galaxy. Future surveys may detect a pulsar orbiting a black hole. A black hole-pulsar binary system would be a powerful laboratory for gravitational physics, testing with high precision whether the orbital motion and gravitational wave generation of black holes conform to the general relativistic predictions.

• *Launch all-sky gamma-ray and x-ray burst detectors capable of detecting the electromagnetic counterparts to LIGO events.*

Cross-correlation of electromagnetic and gravitational signals will help to establish the reality of gravitational wave detections, and may immediately yield crucial clues to the nature of the emitting objects. For example, the enigmatic gamma-ray bursts might be explained if a gravitational wave burst is detected in coincidence. Similarly, supernova searches from ground-based telescopes and neutrino detectors could play a mutually reinforcing role with gravitational wave detectors. Since the first gravitational waves to be detected may well come from transient events, it is urgent to continue the development of the space-based electromagnetic observational capabilities that have already revealed a rich range of astronomical phenomena.
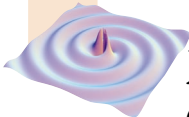
• *Use astronomical observations of supernovae and gravitational lenses to infer the distribution of dark matter and to measure the cosmological constant.*

Certain supernovae appear to be "standard candles"; their intrinsic brightness seems to be the same from case to case, and thus their distance from Earth can be determined from their apparent brightness. Observations of these super-

novae at different locations measure the relationship between distance and red-shift. Current observations of this kind suggest that the expansion rate of the universe is accelerating. This surprising result suggests the existence of a cosmo-logical constant whose value is of fundamental importance for physics. Future observations can help reduce both statistical and systematic errors in these re-sults.

Observations of gravitational lenses can map the distribution of dark matter. By observing lenses at different redshifts, astronomers can determine the evolu-tion of density fluctuations with redshift. Since the evolution depends on the composition of the universe, gravitational lens observations are an independent tool for determining cosmological parameters.

Chapter 2

Ideas and Phenomena of General Relativity

Not everyone who reads this report will be familiar with the beautiful and simple ideas that underlie Einstein's general relativity or with the vast range of phenomena for which gravitational physics is important. In this chapter the Committee on Gravitational Physics briefly describes some key ideas in general relativity and some key phenomena of gravitational physics that are discussed in Chapter 3.

I. KEY IDEAS IN GENERAL RELATIVITY

Gravity Is Geometry. Gravity is the geometry of four-dimensional spacetime. That is the central idea of Einstein's 1915 general theory of relativity—the classical theory of relativistic gravitation. It is not difficult to imagine a curved space. The curved surface of a sphere or a car fender are two-dimensional examples. But gravitational effects arise from the curvature of *four*-dimensional spacetime with three space dimensions and one time dimension. It is more difficult to imagine a notion of curvature involving time, but the Global Positioning System (described in Box 2.1) provides an everyday practical example of its implications.

In Newtonian physics two identically constructed clocks run at the same rate no matter what their positions in space. But in relativity a stationary clock above Earth's surface runs fast compared to a clock at the surface by 1 part in ten thousand billion for each kilometer in height. That tiny difference is the result of the curvature of space*time* produced by the mass of Earth—a small effect indeed, but large enough that the Global Positioning System would fail in a few minutes

24

BOX 2.1 General Relativity and Daily Life

There is no better illustration of the unpredictable application of fundamental science in daily life than the story of general relativity and the Global Positioning System (GPS). Built at a cost of more than $10 billion mainly for military navigation, the GPS has been rapidly transformed into a thriving, multibillion-dollar commercial industry. GPS is based on an array of 24 Earth-orbiting satellites, each carrying a precise atomic clock. With a hand-held GPS receiver that detects radio emissions from any of the satellites that happen to be overhead, a user can determine latitude, longitude, and altitude to an accuracy that currently can reach 50 feet, and local time to 50 billionths of a second. Apart from the obvious military uses, the GPS is finding applications in airplane navigation, wilderness recreation, sailing, and interstate trucking. Even Hollywood has met the GPS, pitting James Bond in "Tomorrow Never Dies" against an evil genius able to insert deliberate errors into the system and send British ships into harm's way.

Because the satellite clocks are moving in high-speed orbits and are far from Earth, they tick at different rates than clocks on the ground. Gravity and speed contribute comparable amounts to the total discrepancy. The offset is so large that, if left uncompensated, it would lead to navigational errors that would accumulate at a rate greater than 6 miles per day. In GPS, the relativity is accounted by electronic adjustments to the rates of the satellite clocks, and by mathematical corrections built into the computer chips that solve for the user's location.
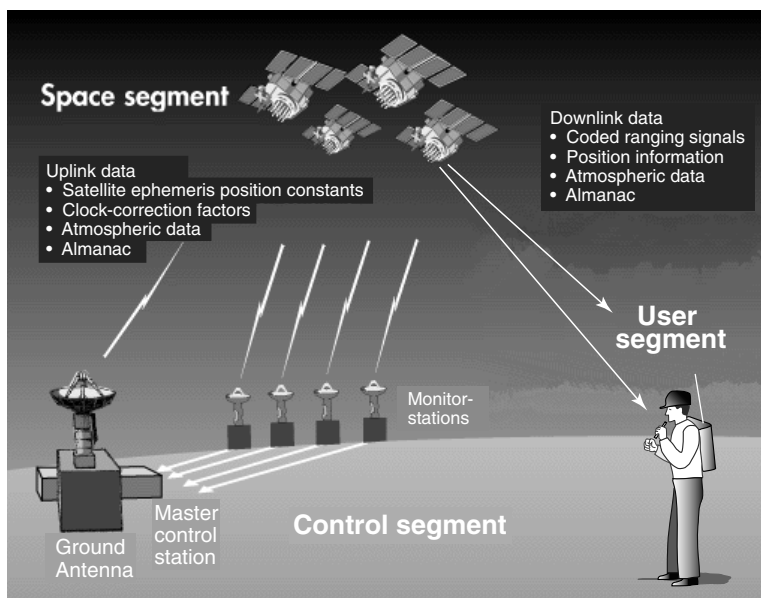
Schematic illustration of segments used in operation of the Global Positioning System. (Adapted from a figure courtesy of the Aerospace Corporation.)

BOX 2.2 Newtonian and Einstein Gravity Compared

In Newton's 300-year-old theory of gravity, a mass attracts other masses with a force of gravity that decreases as the inverse of the square of the distance between them. Masses move in response to the forces acting on them, including gravitational forces, according to Newton's laws of motion.

In Einstein's 1915 general theory of relativity, a mass curves the one time dimension and three space dimensions of spacetime according to Einstein's equation. The spacetime curvature is greatest near the mass and vanishes at a distance. Other masses move along the straightest possible paths in this curved spacetime. Einstein's theory thus expresses both the gravitational effect of mass and the response of mass to that effect in terms of the *geometry of spacetime*. The Newtonian idea of a gravitational force acting at a distance between bodies was replaced by the idea of a body moving in response to the curvature of spacetime.

In relativity, mass and energy are the same thing according to Einstein's famous $E = mc^2$ relation. Not only mass but also any form of energy will curve spacetime. Gravity itself carries energy, and even small propagating ripples in spacetime cause further curvature. The equations of Einstein's theory keep track of this complex feedback interrelationship between energy and curvature.

Newton's theory of gravity is not wrong. It is a correct approximation to Einstein's theory when spacetime curvature is small and the velocities of masses are much smaller than the velocity of light. The first general relativistic corrections beyond Newtonian theory (called "post-Newtonian") are responsible for small deviations to the motion of light and to the orbits of the planets from those predicted by Newton. Measurements of these deviations are among the most precise tests of general relativity.
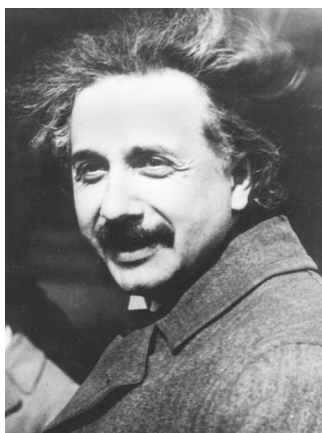
The founders of gravitational physics Isaac Newton (1642-1727) and Albert Einstein (1879-1955). (Courtesy of the American Institute of Physics Emilio Segrè Visual Archives.)

if this effect of the spacetime curvature implied by general relativity were not taken into account.

Mass Produces Spacetime Curvature, and Spacetime Curvature Determines the Motion of Mass. Einstein's equation makes a quantitative connection between mass and the amount of curvature of spacetime it produces. (See Box 2.2.) Just as Earth curves spacetime near its surface, so too does the Sun produce a slight curvature of spacetime in its vicinity. The curvatures produced near the surface of a black hole or a neutron star, or at the beginning of the universe, are much greater. These are realms of strong gravitational physics. According to general relativity, Earth follows an elliptical orbit about the Sun, not because it is attracted to the Sun by a gravitational force, but because it is following the straightest possible path through the spacetime that has been curved by the Sun.

The Principle of Equivalence. General relativity predicts that a tiny asteroid, or indeed any other body, could follow the same path around the Sun as Earth does. Each body is following a path determined by the geometry of spacetime, not by its mass. This universality of free fall—called the principle of equivalence—is one of the foundations of general relativity. It is one of the most accurately tested predictions in all of physics. The equality of accelerations of different bodies in the curved spacetime of the Sun has been verified to a few parts in a thousand billion. Were a violation of this equality ever detected it would signal either new physical interactions or a revision in our ideas about the nature of space, time, and gravity.

II. KEY PHENOMENA IN GRAVITATIONAL PHYSICS

Described below are some important phenomena in gravitational physics. S*trong gravitational physics* plays a central role in all these examples. The essential features of general relativity are present, and the Newtonian approximation is inadequate.

Gravitational Waves. Einstein's theory predicts that ripples in spacetime curvature can propagate with the speed of light through otherwise empty space— a gravitational wave. Mass in motion is the source of a gravitational wave. In turn, gravitational waves can be detected through the motion of masses produced as the ripple in spacetime curvature passes by. The weak coupling of mass to spacetime curvature means that an extraordinarily energetic, strong-gravity event, such as the coalescence of two massive stars, is required to produce gravitational waves copious enough to be detected by gravitational wave receivers now under construction. By contrast, the indirect detection of gravitational waves from the Hulse-Taylor binary pulsar system resulted from the observation of the minuscule shortening of the period of a pair of neutron stars orbiting about each other.

This weak coupling of gravity to matter is the reason that gravitational waves have not yet been detected directly. But this weak coupling also means that the universe is largely transparent to gravitational waves. Once produced, little is absorbed. A gravitational wave receiver could therefore enable researchers to see phenomena in the universe that are visible in no other way.

Black Holes. Perhaps no other concept in physics has made as deep an impact on public consciousness as has the black hole. General relativity predicts that a black hole is created whenever mass is compressed into a volume small enough that the gravitational pull at the surface is too large for anything to escape, no matter how fast it accelerates. The surface of a black hole—called its event horizon—is like a one-way membrane. Mass, information, and observers can fall into it, but nothing can emerge from it. Although black holes in nature are typically produced by complex gravitational collapse, such as gave rise to binary x-ray sources or as occurred at the centers of galaxies, general relativity predicts that they are remarkably simple objects completely characterized by just a few parameters.

Black holes exhibit many properties of ordinary objects: they have mass and spin and can have electric charge; they can oscillate, change shape, show tides, and emit gravitational radiation; they can exhibit electric polarizability, resistivity, eddy currents, and threaded magnetic fields; they can act as generators and engines for the most energetic phenomena in the universe. Yet all this richness of physics is described cleanly by the Einstein equation coupled to ordinary matter. As S. Chandrasekhar put it: "The black holes of nature are the most perfect macroscopic objects there are in the universe: the only elements in their construction are our concepts of space and time. And since the general theory of relativity provides only a single unique family of solutions for their descriptions, they are the simplest objects as well." (*The Mathematical Theory of Black Holes*, Oxford University Press, New York, 1992, p. 1.)

According to quantum mechanics, black holes exhibit yet more remarkable properties as participants in the second law of thermodynamics. They possess an entropy proportional to their area whose statistical mechanical origin is beginning to be understood. They radiate like blackbodies with a temperature inversely proportional to their mass. Thus, as they radiate they heat up and radiate even more. They may radiate completely away, producing for a brief explosive moment the strongest spacetime curvatures since the big bang.

The Universe and the Big Bang. Gravity governs the structure and evolution of the universe on the largest scales of space and time. This is true even though gravity is the weakest of the four fundamental forces. Gravitation is universal, acts at long range, and cannot be canceled since it has no negative "charges." Cosmology and gravitational physics are thus inextricably linked. From cosmo-

logical observations and Einstein's theory, the future fate of the universe can be extrapolated and its origins reconstructed.

Galaxies—the basic building blocks of the present universe—are distributed uniformly on the largest distance scales. They recede from one another in a way that shows that the universe is expanding. Observations show that the universe was even simpler earlier that it is now. The origin of the universe was an initial state of extremely high density, pressure, and spacetime curvature about 13 billion years ago—the big bang. Although extreme in these measures, the big bang was remarkably regular. It was an explosive event that happened everywhere in space at the same time, producing matter in nearly perfect thermal equilibrium. Such a uniform, expanding universe is described by solutions of Einstein's equation known as the Friedmann-Robertson-Walker (FRW) cosmological models. They are characterized by a few cosmological parameters whose values are the subject of ever refined observational searches. The simplest FRW models come in two varieties: models in which space is closed like the surface of a sphere, and models in which space is unlimited or open. The closed models end in a finite time in a "big crunch," whereas the open models expand forever. The closed models have a higher density than the open ones, and the dividing density between them is called the critical density to close the universe. The real universe cannot exhibit exactly the perfect uniformity of the FRW models, since small quantum fluctuations away from uniformity must have occurred. These tiny "seeds" grew by the action of gravitational attraction to form the galaxies and stars we see today.

Cosmic Backgrounds. Matter cooled as the universe expanded from its initial hot beginning. Protons and neutrons formed in the first microsecond after the start of the big bang; during the first few minutes they combined to form primordial nuclei, chiefly hydrogen and helium. About 300,000 years after the big bang, nuclei combined with electrons to make atoms. Once most of the electrons combined into atoms, matter was cool enough to be transparent to light. This light from the big bang has been propagating to us ever since. The subsequent expansion has cooled it to a temperature of only 2.73 degrees above absolute zero, but it still comes toward us from every direction, forming the cosmic background radiation. This light from the early universe is detectable by sensitive instruments on the ground and in space, giving the most compelling evidence for the big bang. Small variations of a few microdegrees that are observed in the temperature of the cosmic background radiation are evidence for the slight initial concentrations of density that grew to be the galaxies today. From the details of these fluctuations—their amplitude, angular distribution, and spectrum—we can learn a great deal about the universe. If we could observe the similar background of gravitational waves, we could see back to the earliest moments of the big bang.

Binary Pulsars. Gravitational physics is central to some of the most dramatic and large-scale phenomena in nature. The big bang, black holes, explosive gravitational collapse, quasars, pulsars, and x-ray sources are all examples. Yet because gravity couples universally to all matter, its effects are in principle observable in any physical system. Just as remarkable, just as beautiful, and just as confirming as the dramatic phenomena mentioned above are the minute, precisely observable predictions of relativistic gravity for the deviations of the paths of orbiting bodies from the laws of Newton. These effects have been observed with impressive accuracy in the solar system. They are observed even more cleanly in binary neutron stars—pairs of neutron stars orbiting about each other. Neutron stars are extraordinarily compact—somewhat more than the mass of the Sun in a radius of 10 kilometers. Spacetime in their vicinity is more highly curved than in any place in the universe other than the big bang and black holes, and binary neutron stars are therefore among the best laboratories for precision tests of general relativity. Their orbits can be observed when one of the neutron stars is a pulsar—a magnetized object whose rotation can be observed from the radio waves it emits, received at Earth as precisely periodic signals. More than 1000 pulsars are known, and 50 are in binary systems with neutron star or white dwarf companions. Many of them are extraordinarily accurate clocks. The rotational (spin) period of the Hulse-Taylor binary pulsar PSR1913+16, for example, is $0.059029997929613 \pm 0.000000000000007$ seconds. By noting the changes in this period induced by the pulsar's orbital motion over decades, the effects of general relativity can be precisely observed.

Singularities. Einstein's classical theory predicts the formation of a singularity in the interior of a massive body whose gravity collapses it to a sufficiently compact volume. A singularity is a region of the universe where a classical description breaks down because it predicts infinite spacetime curvatures or densities of matter. Singularities limit the predictive ability of classical general relativity and are therefore places where we can expect to find new physics. But there is considerable evidence that singularities produced in any realistic collapse are hidden inside the event horizons of black holes where they cannot interfere with the predictability of physics on the outside. The idea that this always happens is called the cosmic censorship conjecture. Proving or disproving it is one of the outstanding challenges of general relativity theory. The singularities of gravitational collapse may be hidden inside black holes, but Einstein's theory also predicts that the universe began in a singularity—the big bang—whose consequences are all about us.

The Small-Scale Structure of Space and Time. The union of the two most significant developments of 20th-century physics—general relativity and quantum theory—is one of the greatest challenges of contemporary theoretical physics. The result of this union—a quantum theory of gravity—will have implica-

tions as profound for our understanding of spacetime on small scales as Einstein's theory did for that understanding on large scales. The frontier of small scales for quantum gravitational phenomena is marked by the combination of the quantum of action $\hbar$, the velocity of light c, and Newton's gravitational constant G, called the Planck length:

$$\ell_p = \left(\hbar G / c^3\right)^{1/2} \approx 10^{-33}\,\text{cm}.$$

The corresponding Planck energy is about ten thousand trillion times greater than the energy reached by the world's largest accelerators. Yet these energies, with their accompanying enormous curvatures, occurred at the big bang and occur in the final stages of gravitational collapse. An understanding of spacetime (or whatever replaces it) on these extreme scales is necessary to understand these central phenomena of astrophysics. But the small-scale structure of spacetime is also central to the quest for a unified theory of the fundamental interactions, because it is only there, where gravity is as strong as any other force, that the full symmetry between these interactions is likely to emerge.

There are currently two main approaches to constructing this new theory. In one, Einstein's picture of gravity as spacetime geometry is fundamental so that a quantum theory of gravity brings with it a quantum theory of geometry. At the Planck scale, quantum excitations of geometry are structured like a branched polymer, and familiar quantities such as lengths and areas can assume only discrete values. It is only because the basic discrete unit of length—the Planck length—is so small that space can be approximated by a continuum under ordinary circumstances.

The other approach to quantum gravity unifies gravity with all the other forces and matter in a natural way. The basic idea is that although elementary particles appear point-like, they are actually excitations of a one-dimensional extended object called a *string*. One mode of oscillation of the string is a graviton, a quantum of gravity, while other modes are photons, electrons, quarks, and so on. Furthermore, the interactions between these particles are reproduced by a splitting and joining interaction between the strings. We thus obtain a compelling and beautiful unified picture of particles and their interactions, including gravity.

Both of these approaches are still being developed, and the coming decade promises a much more detailed picture of space and time at small scales.

Chapter 3
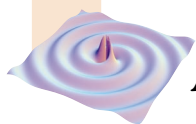
Achievements and Opportunities in Gravitational Physics

I. GRAVITATIONAL WAVES

Key Questions

It is a tenet of Einstein's 1905 special relativity that no information can be transmitted or carried in any way at a speed faster than that of light, an idea prefigured in Maxwell's earlier theory of electromagnetic waves. When general relativity was worked out by Einstein using special relativity as a base, it was natural that it should predict that moving masses would communicate their changed gravitational fields at the speed of light, through the propagation of gravitational waves. Gravitational waves are a quintessential relativistic strong-gravitational-field phenomenon—one that is completely absent in Newtonian gravitational theory.

In the years since general relativity was proposed, many of its predictions have been spectacularly verified, but a few key features still remain unconfirmed. (See Section IV of this chapter.) Remarkably, one idea yet to be fully checked is the feature most closely related to the principle of relativity: gravitational waves. Just as interestingly, the eventual detection of gravitational waves will probably provide the best possible way to verify the other most spectacular of the unverified predictions of general relativity: the existence of black holes.

There are practical reasons that the earliest relativistic idea about gravity might be among the last to be verified. Compared to electric or magnetic forces,

32

gravity is extremely weak. This means that it is much harder to construct a practical receiver for gravitational waves than it is to construct an electromagnetic (e.g., radio) receiver. Worse, construction of a gravitational wave generator (or transmitter) from laboratory-scale components is hopeless. Einstein himself thought gravitational waves might never be detected.

As is described below, new technology just coming on line is expected to be able to detect gravitational waves generated by the rapid motion of astronomical bodies whose masses are comparable to those of stars.

In a radio receiver, the reception of an electromagnetic wave begins with the acceleration of electrons in the receiver's antenna by the electric field component of the electromagnetic wave. Similarly, a gravitational wave will cause motions among a set of masses that are free to move. Only relative motions are meaningful, though, because all objects are required to have the same free fall motion by the principle of equivalence. (For more on this principle, see Box 3.3 in Section IV of this chapter.) The measurable effect of a gravitational wave is a distortion in the distances between a set of free masses, characterized by the fractional change in distances $\Delta L/L$. (It is traditional to refer to the wave amplitude $h \equiv 2\Delta L/L$.)

The strongest waves arriving regularly at Earth (say, several times per year) are expected to cause fractional length changes $\Delta L/L$ between pairs of detector masses separated by a distance L of no larger than about 1 part in 10^{21}. It is this fact that scales the technological challenge of detecting gravitational waves. Still, the reception and study of gravitational waves can help answer many key questions in basic physics and astrophysics:

• Do waves such as those predicted by Einstein propagate away from dynamic massive objects, and do they interact with test bodies in the way described by general relativity?

• Do gravitational waves propagate at the speed of light, and do they have the polarization that general relativity predicts?

• What is the nature of gravity in the strong-field regime where general relativity makes its most dramatic predictions?

• Do black holes exist? What are the properties of the highly relativistic spacetime just outside their horizons?

• Can we use neutron star and white dwarf binary systems to study gravitational physics?

• Are massive black hole binaries present in galactic centers?

• What is the state of matter inside neutron stars or in the collapsing cores of supernovae?

• What is the origin of gamma-ray bursts?

• Are there gravitational waves left over from the very early universe?

Achievements

The Binary Pulsar and the Emission of Gravitational Waves

For some years after Einstein's initial prediction, even the existence of gravitational waves was in doubt, at least for some. That is no longer the case, thanks to the remarkable work of radio astronomers Joseph Taylor and Russell Hulse. In 1974, they discovered the pulsar PSR1913+16. Its frequency varied with a period of 7 3/4 hours, revealing it to be a member of a binary system. (For this reason, it was called *the* Binary Pulsar; now about 50 other binary pulsars are known. For background on the Binary Pulsar see Chapter 2.) Over the succeeding years, careful measurement of the arrival times of the pulses of radio emission from the pulsar revealed the shape of the pulsar's orbit in unprecedented detail. By recognizing a variety of relativistic effects (including orbit precession, gravitational redshift, and the special-relativistic time-dilation), Taylor and his co-workers were able to show that both the pulsar and its companion were neutron stars with precisely measured masses around 1.4 times the mass of the Sun.

The most exciting result of these studies was the recognition that the motion of the pulsar around its companion could not be understood unless the dissipative reaction force associated with gravitational wave production was included. The two neutron stars, by virtue of their motion about one another, execute precisely the sort of motion that generates gravitational waves. Those waves carry away energy. Thus, the two stars must gradually fall closer to each other, with the result that their orbit steadily speeds up. The motion of the Binary Pulsar has shown that this orbital speedup is occurring in accordance with the rate predicted by general relativity, to a precision of a third of a percent. (See Figure 3.1.) For the discovery of this remarkable object, Hulse and Taylor were awarded the Nobel Prize in physics in 1993.

Experimental Searches, Ongoing and New

A gravitational wave interacts with matter by producing differential forces and thus relative motions between sets of masses. Experiments to detect gravitational waves involve setting up systems of a few test masses, then looking in as sensitive a way as possible for relative motions between them. Resonant detectors, based on the original idea pioneered by Joseph Weber in the 1960s, use a large single extended body such as a solid cylinder, whose ends may be thought of as separate masses being pulled apart or pushed together by the wave as it passes. The newer interferometric detectors use three or more small masses that are widely separated; a propagating laser beam is used to monitor their separations, which will be perturbed when a gravitational wave passes.
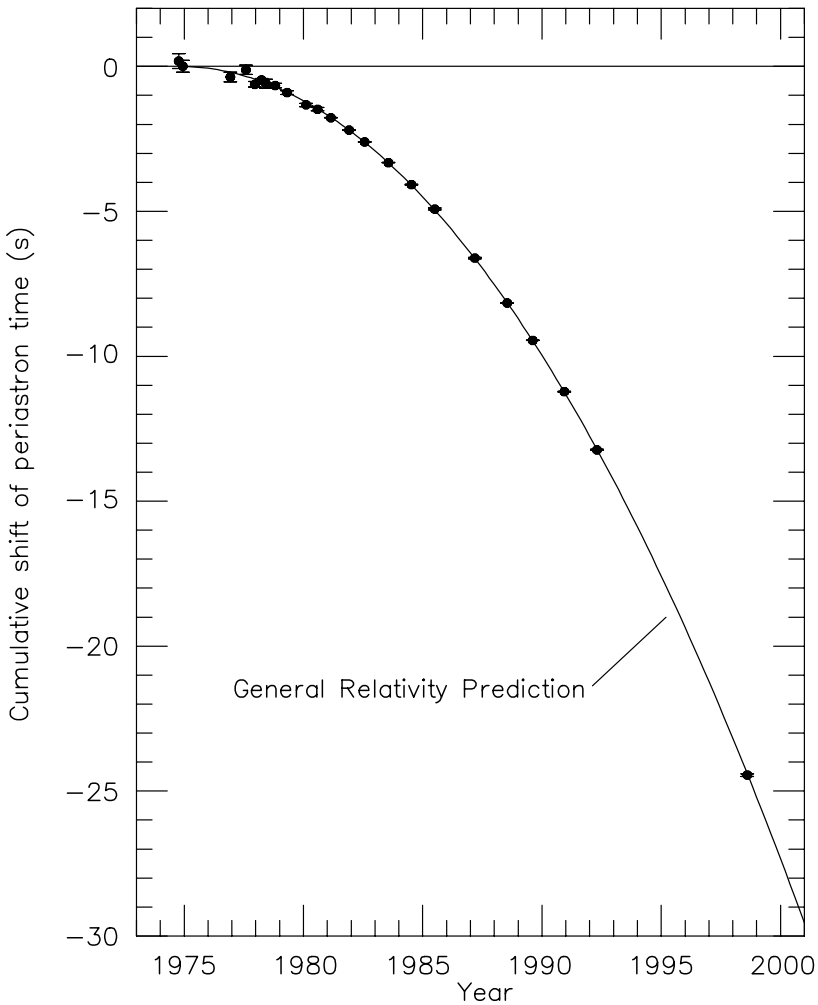
FIGURE 3.1 The orbital period of any body around another decreases because of the energy lost to gravitational radiation. That effect is strongest in highly relativistic systems such as the binary pulsar PSR1913+16. One measure of this decrease in orbital period is the steady shift over time of the time of the pulsar's closest approach (periastron) to its companion star. The figure above shows the cumulative value of this shift measured by J. Taylor and J. Weisberg at the Arecibo radio telescope in Puerto Rico over several decades. The points are their data points. The solid line is the shift predicted by general relativity. The agreement is better than a third of a percent. (Courtesy of J.H. Taylor and J.M. Weisberg; to be published.)

Resonant Detectors. The first detector to reach astrophysically interesting sensitivity was the ultracold resonant bar at Stanford University, operating at 4 K (degrees above absolute zero). In 1980, it operated with a sensitivity to short bursts with strain amplitudes ($\Delta L/L$) of around 1 part in 10^{18}. The first observations looking for coincident events in widely separated detectors were carried out in 1986. High-sensitivity coincidence observations were performed between an Italian detector and the Allegro detector at Louisiana State University (Figure 3.2) in 1991. This run determined the strongest upper limit yet on the flux of gravitational waves.

A new generation of resonant detectors has now begun operation, using dilution refrigerators to bring their several-ton resonators to temperatures of around 50 mK (5/100 degree above zero). The pioneers of this class are the Nautilus and Auriga detectors in Italy. They should eventually attain a sensitivity of about 1 part in 10^{20}.

Designs have been produced for resonant detectors that could reach sensitivities of parts in 10^{-21}. These detectors would extend the technology already developed for aluminum cylinders to spheres as much as 10 times more massive.

Interferometric Detectors. For many years, work on interferometers, directed at kilometer-scale devices that could achieve astrophysically motivated sensitivities, was devoted mainly to proof-of-principle devices and engineering tests. Finally, in the early 1990s large interferometer construction projects were approved in several countries around the world. The U.S. entry, the Laser Interferometer Gravitational-Wave Observatory (LIGO), will consist of two facilities—one in Hanford, Washington, and the other in Livingston, Louisiana, each of which will contain a Michelson interferometer of arm length 4 kilometers (Figure 3.3). (The Hanford site will also carry an interferometer of half that length for additional coincidence measurements.) When LIGO becomes operational in 2002, it is expected to be able to make unambiguous detections of waves with strains $\Delta L/L$ around 1 part in 10^{21}. Similar results are expected from the 3-kilometer VIRGO interferometer (a French-Italian project located near Pisa). The British-German GEO 600-meter interferometer near Hannover has the handicap of shorter arm length, but early application of advanced interferometer technology will allow it to be competitive in some frequency ranges, at least for a while. There is also a 300-meter interferometer called TAMA under construction near Tokyo, and an Australian project in the planning stage called ACIGA.

Theoretical Studies of Gravitational Wave Sources

During the last decade, the theoretical prediction of gravitational wave sources reached new levels of sophistication and promise. This effort was driven by progress in gravitational wave detectors and made possible by advances in numerical and analytic techniques for solving Einstein's equations. The ability to
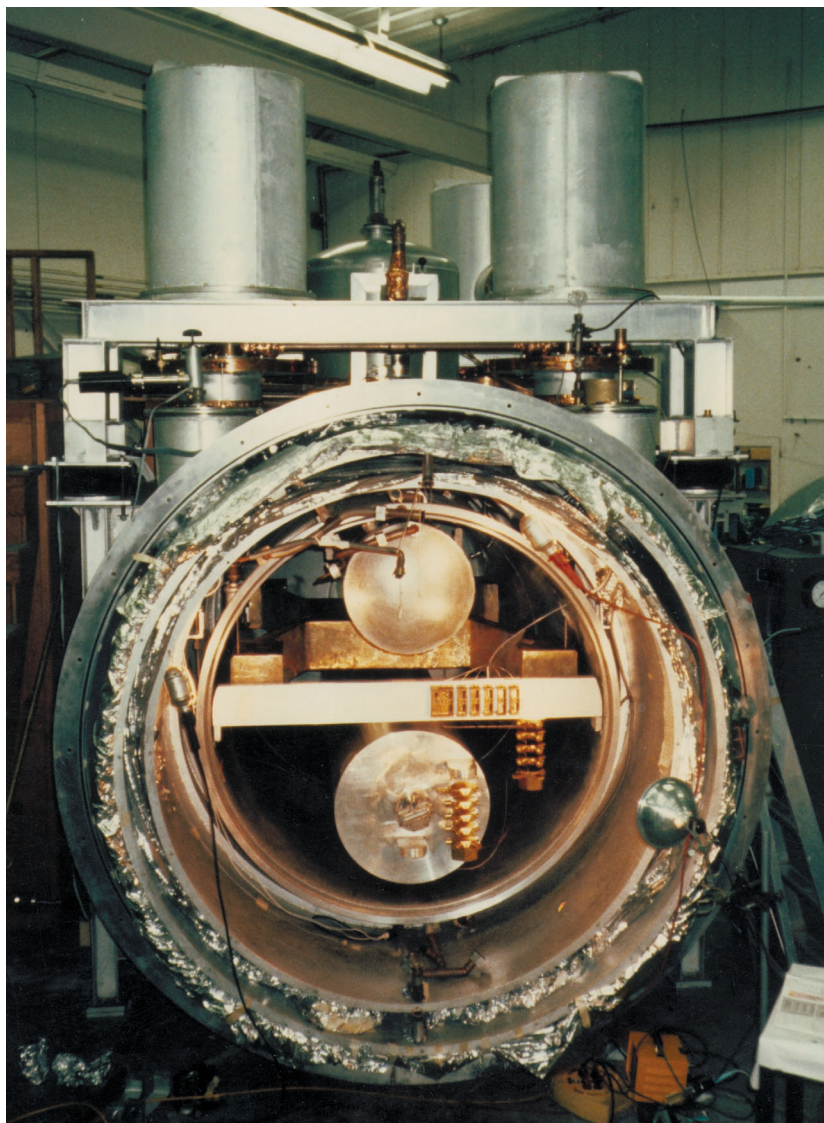
FIGURE 3.2 A view of the Allegro resonant bar gravitational wave detector at Louisiana State University. The photo shows the bar nestled in its cryogenic dewar, shortly before it was closed up. The end of the bar is visible as the circular structure in the lower half of the dewar. The rest of the internal components are structures for cooling the bar to 4.2 K, for bringing out the signal from the transducer on the bar's end, and for isolating the whole system from external disturbances. Since 1991 Allegro has functioned as the most sensitive continuously operating gravitational wave detector in the world. (Courtesy of Bill Hamilton, Louisiana State University Physics and Astronomy.)
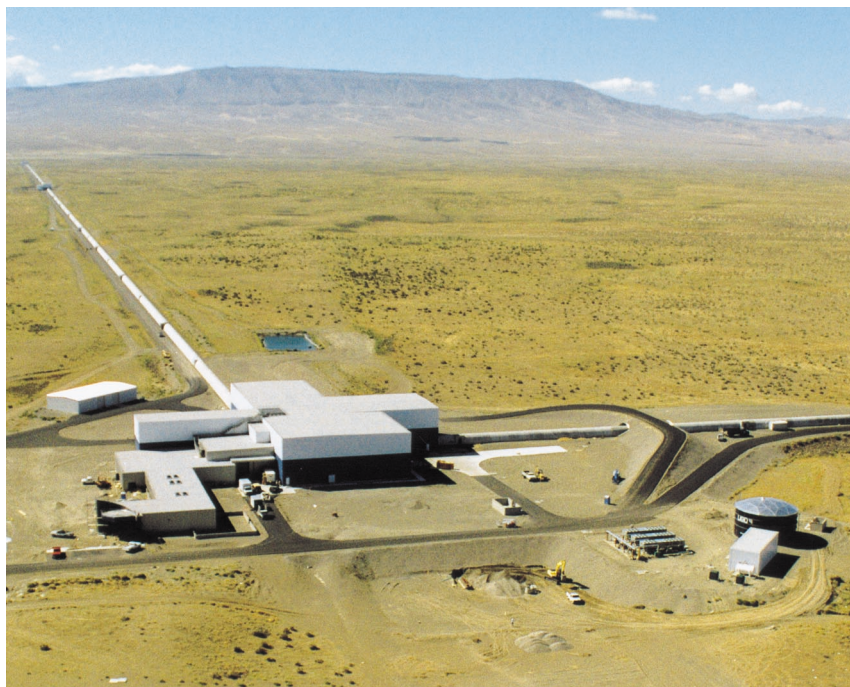
FIGURE 3.3 An aerial photograph of the Laser Interferometer Gravitational-Wave Observatory (LIGO) facility at Hanford, Washington, in 1998. The large building in the foreground holds the vertex of the interferometers, along with the lasers, input test masses, and the input and output optics. The building also contains the control room, experiment staging areas, laboratory space, and offices for the observatory staff. One of the 4-kilometer arms disappears out of the right-hand side of the photo. The other is seen stretching to the upper left. At the 2-kilometer point can be seen a smaller building holding the end mass of the half-length interferometer that will run in parallel to the 4-kilometer interferometer, providing a local check on the observations. (Courtesy of the LIGO Laboratory.)

carry out numerical simulations of gravitational collapse in three spatial dimensions (i.e., without restrictive symmetries), together with improvements in integrating realistic microphysics into the description of the collapsing stellar matter, gave results that demonstrated a remarkable sensitivity of the gravitational wave output from a supernova to the details of neutrino physics, hydrodynamics, and thermal physics. Similarly, large-scale numerical simulations of the merger and coalescence of double black hole or double neutron star systems are on the verge of achieving reliable results. (See the discussion under "Computational General Relativity" in Section II of this chapter.)

The theory of gravitational waves from the inspiral phase of double star systems was advanced using analytic tools based on the post-Newtonian technique, a method for approximating solutions of Einstein's equations by successive improvements on its first-order, Newtonian approximation. The results, carried to remarkably high order in successive steps, provided very accurate gravitational wave "templates," which will play a role in analysis of signals detected by LIGO-type gravitational wave detectors.

The theory of small perturbations of stars and black holes, initiated in the early 1960s, was taken to new levels of development. One result was the discovery of entirely unsuspected unstable modes of oscillation of rotating stars, which could be promising sources of gravitational radiation and could explain the rapid spin-down of newly formed neutron stars. Another was the development of a nearly complete description of the gravitational wave emission from a small mass orbiting a massive black hole and of the "ringing" modes of distorted black holes.

Finally, over the last 5 years a new method was developed to study the properties of gravitational waves emitted in the very final stages of black hole mergers. The method, called the close limit approximation, combines analytic and numerical approximation methods and has yielded fresh insights into the diverse physical processes responsible for various features of the emitted radiation. It paves the way for gravitational wave phenomenology along the lines of the standard quantum mechanical perturbation theory used in analyzing spectra in atomic physics.

Opportunities

Ground-based Reception of Gravitational Waves

The decade just ending saw the National Science Foundation make a substantial investment in the construction of the research facilities of the LIGO project (see Figure 3.3). The great opportunity of the coming decade is to exploit those facilities by operating receivers of sufficient sensitivity to detect the gravitational waves emitted by astronomical bodies. The expectation that gravitational waves will be detected during the coming decade represents one of the most exciting research opportunities of gravitational physics. The spectrum of gravitational waves expected from known sources is shown in Figure 3.4. Detection and study of those waves can address many of the key questions listed above.

Key Questions Addressed

• Do waves such as those predicted by Einstein propagate away from dynamic massive objects, and do they interact with test bodies in the way described by general relativity?
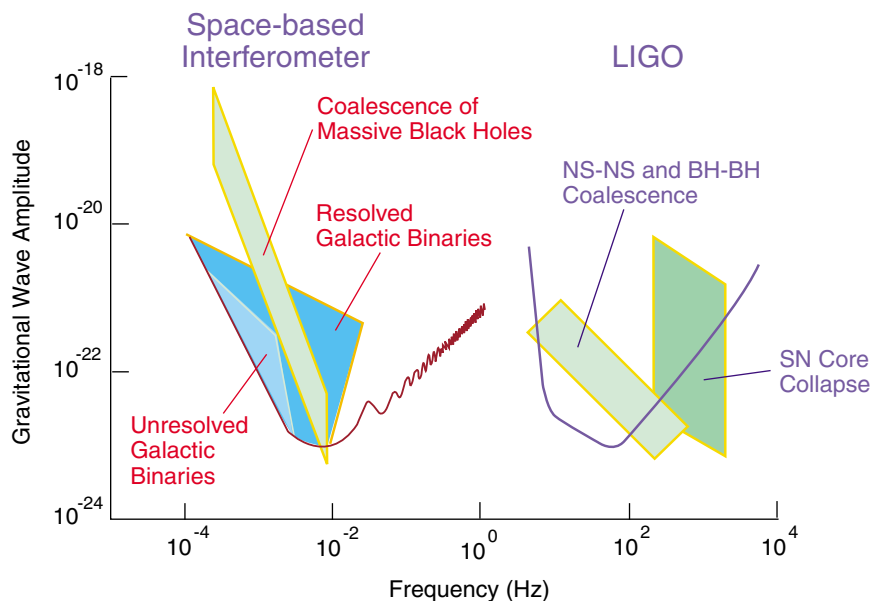
FIGURE 3.4 A schematic view of the gravitational wave spectrum, showing the project-
ed sensitivity of the advanced version of LIGO and of a proposed space-based interferom-
eter. LIGO and the space-based detector will each be able to look for gravitational wave
sources in a band a decade or two wide. LIGO will have its best sensitivity near 100 Hz,
while an instrument in space should be most sensitive near 10 mHz. The high-frequency
window accessible to LIGO is best suited for studying the signals from the coalescence of
neutron star (NS) binaries, and from binaries consisting of black holes (BH) with masses
around 10 times that of the Sun. Binaries of massive (10^6 solar masses) black holes, such
as are found in galactic nuclei, will be a primary target of a space-based interferometer.
Each detector should also be capable of finding the signals from a variety of other astro-
nomical objects, as described in the text. For example, a space-based detector would be
able to record the signals of many known binary star systems. SN, supernova. (Courtesy
of the Jet Propulsion Laboratory, California Institute of Technology.)

The work of Taylor and his collaborators tracking the orbit of the binary
pulsar PSR1913+16 established dramatically that gravitational waves were being
emitted by the binary neutron star system, with a rate of energy loss in agreement
with the predictions of general relativity. In a very real sense, that measurement
can be said to have "detected" the emission of gravitational waves.

But physicists' paradigm of establishing the existence of a wave phenom-
enon is the set of 19th-century experiments on electromagnetic waves performed
by Heinrich Hertz (1857-1894), which demonstrated not only energy loss in the
transmitter, but also (1) propagation across spatial intervals large compared to the

wavelength and (2) excitation of test bodies in a manner consistent with the field carried by the wave. We may never have the capability to construct and manipulate a gravitational wave transmitter, but can always rely on the existence of natural ones, as did the Hulse-Taylor binary pulsar experiment. However, receivers can be built that establish that a wave propagated across the space from the transmitter and that this wave interacted with test particles in the expected way. This is the fundamental role LIGO and the other gravitational wave detectors can be expected to play, when they successfully detect the arrival of gravitational waves of astrophysical origin.

• Do gravitational waves propagate at the speed of light, and do they have the polarization that general relativity predicts?

The speed of gravitational waves is unambiguously predicted by general relativity to be equal to the speed of light in vacuum. Similarly, general relativity states that the polarization of the waves should be strictly quadrupolar, although other gravitation theories predict some admixture of other polarizations in addition. The relativistic predictions are equivalent to the statements that the quantum of gravitation analogous to the photon—the graviton—is massless and has spin 2. (However, no detector is foreseen as being sensitive enough to detect individual gravitons.)

These features of gravitational waves can be checked once the waves are detected. The polarization is most directly measured by verifying that the signal strengths detected by receivers at different locations on Earth (hence with different orientations) agree with those predicted from the source's position on the sky (as determined by time delays). The propagation speed can be checked against that of light whenever the gravitational wave emission is accompanied by some electromagnetic counterpart; examples might include the optical flash of a supernova, or perhaps a gamma-ray burst.

• What is the nature of gravity in the strong-field regime where general relativity makes its most dramatic predictions?

Gravitational waves are emitted most strongly when large masses move at relativistic speeds in close proximity to one another, especially as those masses approach the degree of compactness of black holes (as in neutron stars or black holes themselves). These are inherently strong-field situations, with dynamics dramatically different than Newtonian theory would predict. The dynamics of the ultimate strong-field sources, black holes, are even more distinctive.

• Do black holes exist? What are the properties of the highly relativistic spacetime just outside their horizons?

Black holes are relativistic strong gravitational phenomena, and they show the most dramatic effects of strong gravity. The many aspects of the study of black holes are discussed in Section II of this chapter.

The "cleanest" test of the existence of black holes (and of the predictions of theory regarding strong-field gravity in general) would be the measurement of the gravitational waves emitted when a black hole is disturbed, or when it forms in a gravitational collapse or merger. These waves provide a direct probe of the dynamics of the region just outside the black hole's horizon. The spectrum of the quasi-normal modes (the "ringing" of a disturbed black hole) has been extensively studied theoretically and should be easily recognizable in a gravitational waveform. Its most distinctive feature is that all of the modes are highly damped. The damping mechanism is the gravitational wave emission process itself. The frequency of the signal is inversely proportional to the mass; a 10-solar-mass black hole has its fundamental resonance near 1 kHz. Detection of waves from the merger phase will also probe strong-field gravity effects. Interpretation of such signals will require numerical integration of the full Einstein equations. (See "Advances in Computational General Relativity" in Section II of this chapter.)

• What can we learn from the study of coalescing neutron star binaries?

The gravitational wave source whose signal is most securely predicted is the coalescence of neutron star binaries, of which the Hulse-Taylor binary pulsar is a prototype. The shrinkage of the orbit of such a system is driven by gravitational radiation until the two stars coalesce, through a process that can be calculated in precise detail until nearly the final moments. The gravitational luminosity of the system grows as the orbit shrinks, with the final signal at frequencies of several hundred hertz.

It is a key aim of the LIGO project to develop receivers capable of detecting several of these events per year, which means being able to detect them to distances of several hundred megaparsecs (Mpc). This is the goal of the planned upgrades to the initial LIGO interferometers, scheduled to start in the middle of the coming decade. (See Chapter 1, Table 1.1, which is based on the analysis described in the addendum to this section.)

When the signals from coalescing neutron star binaries are detected, we will probe strongly post-Newtonian orbital dynamics. Just as interesting will be what we will learn about the properties of nuclear matter, which will strongly influence the final phases of the orbit and ring-down of the coalesced star. It is a difficult problem to calculate the signals researchers should expect from the end of a coalescence event. This led to the formation of the NASA Neutron Star Grand Challenge effort in computational physics. (See the discussion on neutron stars under "Advances in Computational General Relativity" in Section II of this chapter for more details.)

- What can we learn about the state of matter inside neutron stars or in the collapsing cores of supernovae?

The essential truth of the core collapse model for Type II supernovae was demonstrated only quite recently, by the detection of neutrinos from Supernova 1987A. Gravitational waves offer another way to learn about the inner workings of the collapse. The gravitational waveform is proportional to the second time derivative of the core's quadrupole moment, so it will trace the history of the collapse directly. The gravitational waveform will emerge without any absorption by the outer layers of the star, unlike all other signals (including neutrinos). The strength of the signal (and thus the rate of detectable signals) depends on the degree to which the core is in fact endowed with a quadrupole moment. Recent calculations have begun to suggest the possibility of non-negligible departures from spherical symmetry, although the amplitudes predicted still would not make this a very strong signal. Whether the strong kicks that newborn pulsars appear to receive could be related to an asymmetric collapse is also uncertain.

- What is the origin of gamma-ray bursts?

Some of the most popular models for gamma-ray bursts involve the collision of two neutron stars. If these ideas are correct, then for some bursts there should be coincident gamma-ray and gravitational wave signals, presumably with a gravitational wave precursor from the inspiraling stars. Other popular models of gamma-ray bursts involve collisions of black holes or core collapse events inside massive stars; either of these classes of models also would generate coincident gravitational wave signals. Whether or not such models turn out to be correct, gravitational wave observations will make an important contribution to the understanding of the enigmatic gamma-ray bursts.

- Are there gravitational waves left over from the very early universe?

A gravitational wave background from the early universe, if it is detectable, would provide the earliest possible glimpse of the history of our universe. It is hard to predict the strength or the frequency of such signals and also hard to distinguish such signals from other sources of noise. Nevertheless, LIGO will search for this effect, as will a space-based detector. The issues are described in more detail below in "Space-based Reception of Gravitational Waves," in the discussion of key questions addressed by space-based detectors.

LIGO Operation and Sensitivity Enhancement. The LIGO initial data run (2002-2003) will represent a genuine milestone: its sensitivity of 10^{-21} will be nearly 3 orders of magnitude in amplitude (6 orders of magnitude in energy) more

sensitive than the best previous searches. Discovery of gravitational wave signals at this sensitivity could occur.

But the search for gravitational waves has to be viewed as a long-term effort. This is not because there is uncertainty about whether gravitational waves exist or about their theoretical underpinning. Gravitational waves have been detected indirectly through the analysis of the orbit of the Hulse-Taylor binary pulsar, in excellent agreement with the predictions of general relativity. Rather, there is uncertainty concerning the frequency of violent events in the universe which could produce gravitational radiation strong enough to be detected on Earth. Uncertainty in the abundance of sources means that no precise sensitivity threshold for the first detection can be specified. The only sensible strategy for success is to press ahead with two activities simultaneously: aggressive development of the technology for more sensitive interferometers, and the installation and operation of advanced interferometers as they become available.

It may be that the only sources of gravitational waves are the ones we already know about from previous (electromagnetic) observations. If in fact the universe holds so few surprises, then LIGO will first detect the signals from coalescing neutron star binaries. The CGP's estimates for the event rate for these is discussed in detail in the addendum to this section. Given these estimates, an improvement in sensitivity of around a factor of 10 from that characterizing the initial data run will probably be required to detect these signals regularly.

An R&D program aimed at achieving the indicated sensitivity enhancement has started. Designs are now laid out that would improve LIGO's sensitivity by around a factor of 10. Development of key technologies is under way. As shown in Table 1.1 in Chapter 1, LIGO's plans call for the implementation of these enhanced-sensitivity interferometers around the middle of the next decade.

Plans for even more sensitive interferometers have also begun to be formulated. This line of research is critical; its results may be required to guarantee the detection of gravitational wave signals, if the most pessimistic estimates of event rates are correct. Even in the happy event that signals are detected sooner, improvement in signal-to-noise ratio and in event rate will pay big dividends in the scientific value of gravitational wave observations.

Space-based Reception of Gravitational Waves

Key Questions Addressed. While LIGO's promise is great, its observations will necessarily be limited by terrestrial noise sources to a frequency band above about 10 Hz. A space-based interferometer can open the frequency band between roughly 10^{-4} Hz and almost 1 Hz. Completely different classes of gravitational wave sources will be observable at these lower frequencies.

Several key technologies necessary to implement such a gravitational wave detector still need development. Nevertheless, the promise of a space-based detector is enormous. At the sensitivity that can reasonably be expected, there

are numerous sources of gravitational waves that will be detected with signal-to-noise ratios of hundreds to thousands. The CGP discusses below four scientific opportunities that an appropriately designed space-based gravitational wave detector should allow us to achieve.

- Can we use white dwarf binaries to study gravitational physics?

The low-frequency gravitational wave band contains the signals from white dwarf binaries. These have the tremendous advantages that both their waveforms and their strengths can be confidently predicted. Their quadrupole moments can be calculated in a straightforward way, and their distances are well known from many years of astronomical study. Those signals ought to be relatively easily detected by the sorts of space-based detectors now being proposed. Thus, the cleanest possible tests of gravitational wave theory can be made by observing the waves emitted by a source of precisely known properties. Detectable signals will be numerous, so much so that, at the lowest frequencies, their periodic signals will overlap in a "confusion limit" that will constitute one of the chief noise sources.

- Are massive black hole binaries present in galactic centers?

The frequency band accessible from space contains the signals from black holes in the 10^3 to 10^6 solar mass range. The upper end of this mass range is characteristic of the black holes known to populate many galactic nuclei. Massive black hole binaries could result from the merger of two galaxies, each containing a single massive black hole. The role these black holes play in galactic activity, and the importance of galactic mergers, are two of the key questions of extragalactic astronomy. (For more on this, see "Supermassive Black Hole Mergers and Space-based Detectors" in the addendum to this section.) If a binary of two such black holes were to coalesce anywhere within the visible universe, the resulting gravitational wave signal should be detectable at very high signal-to-noise ratio by the planned space-based interferometers.

- What are the properties of the highly relativistic spacetime just outside black hole horizons?

One of the key ideas of the theory of black holes is that they are each entirely describable by three numbers: mass, spin, and charge. (In astrophysical situations, the charge is expected to be extremely small.) This is called the "no hair" theorem since it accords black holes so little individuality. These three numbers completely determine the nature of a black hole's spacetime. A space-based interferometer could measure the waves from black hole binary coalescences in such detail that any departure from the predictions of the "no hair" theorem can be sensitively tested.

• Are there gravitational waves left over from the very early universe?

Just as our ability to probe the interior of a supernova electromagnetically is frustrated by the opacity of the outer layers of the star, so too do our electromagnetic probes of the universe run into the opaque "surface of last scattering," when hydrogen was last ionized about 300,000 years after the big bang. Gravitational waves emitted before this epoch could be bathing Earth. The simple gravitational analog of the electromagnetic cosmic background radiation would be too weak to be detected. But some processes that might have occurred during inflation, or others associated with cosmic phase transitions, could lead to detectable levels of gravitational waves. Distinguishing such a background of waves from other sources of noise will be challenging. Still, no other astronomical tool even holds out the hope of enabling a look at the Planck-scale physics of quantum gravity. Thus gravitational wave observations could play a key role in supplying evidence for a theory of quantum gravity. (See the discussion of ideas about quantum gravity in Section V of this chapter.)

R&D Toward a Space-based Gravitational Wave Detector. Planning is under way for detectors able to explore the low-frequency window, which is accessible only from space. To make the detectors a reality, it will be necessary to carry out a program of technology development to ensure that the required performance specifications can be reached. Development of inertial sensors, drag-free attitude control, very long baseline optical interferometry, and ultrastable mechanical structures is required. When these technological developments have been achieved, then the scientific case for deploying a space-based gravitational wave detector will be overwhelmingly strong.

Related Opportunities

Gravitational Wave Theory. The effort to study and detect gravitational waves requires not only the effort of scores of experimental physicists and engineers, but also the help of many theoretical physicists as well. A multifaceted program of research in gravitational wave theory is necessary to ensure that signals can be efficiently extracted from LIGO's data stream, and be used to address the key questions.

Some theorists will design the software that will enable the large quantities of data to be efficiently archived, while others are helping to develop the algorithms for optimally extracting signals from the detectors' data streams. Others will continue to work to understand as well as possible the predictions of the theory of general relativity, so that we will be ready to recognize the signals created by the astronomical sources of gravitational waves. Some of this theoretical work has been of the traditional analytic style. Another part of this program involves the numerical solution of the Einstein equation; this is probably

the only way to deal with the highly relativistic final stages of binary coalescence, for example. Still other parts of the theoretical effort have the flavor of astrophysics, trying to predict the abundance of objects that will serve as strong gravitational wave sources.

The Connection Between Gravitational Waves and Other Astronomical Observations. In a nascent branch of astronomy like gravitational wave detection, success will be greatly aided by making the strongest possible link to other branches of astronomy. For example, new observations from the Compton Gamma Ray Observatory (CGRO) satellite (and more recently from BeppoSAX) have sparked a strong suspicion that the enigmatic gamma-ray bursts may result from the merger of two neutron stars. Since that is precisely the kind of event expected to give detectable gravitational wave signals, there exists the possibility that gravitational wave observations can help to resolve one of the outstanding puzzles in astronomy. By the same token, ongoing observations of gamma-ray bursts may allow a more sensitive search for the gravitational wave signals, by establishing the arrival times of the signals at Earth.

Another kind of violent astrophysical event that may create strong gravitational waves is the stellar core collapse that initiates a Type II supernova. As in the case of gamma-ray bursts, knowledge of the times of nearby supernova events will help both in possible gravitational wave signal detections and in testing models for the strength of the emitted waves. Neutrino observations from current and planned detectors may provide an observation of core collapse from a nearby supernova. Coincident detection of a gravitational wave signal would yield new insights into the nature of the collapse.

In these and other possible cases, it is expected that gravitational wave detection and electromagnetic astronomy will complement each other in the study of some of the most dramatic events in the universe. For this scientific cross-fertilization to lead to the richest possible results, it is important to support astronomical research likely to overlap with gravitational wave observations. Astronomical study of compact objects, especially the carrying out of all-sky surveys for transient events in the optical, x-ray, and gamma-ray bands, will continue to play an important part in gravitational wave research.

Searches for Gravitational Waves of Cosmological Origin. Although our understanding of the physics of the universe in its very early stages is still primitive, we do know that if the universe began violently, then gravitational waves arriving today will carry the imprints of those first few moments. If a phase transition after the first moments produced a background of gravitational waves, then interferometric observation (possibly ground-based but more likely space-based) is one means to search for it, as discussed in the sections on ground-based and space-based observations above. Pulsar timing experiments, which examine longer wavelengths, are also sensitive probes.

On the other hand, if a background of gravitational waves was produced by cosmological inflation, then it could have a spectrum that makes the study of the cosmic microwave background's polarization the most sensitive probe.

In the absence of a clear understanding of the early universe, it is reasonable to pursue the search for these relics of the earliest instants of our history on as diverse a spectrum of wavelengths as possible. Even upper limits in any range of wavelengths can provide important information on early universe physics. For example, if the radiation is less than predicted from the merger of bubbles arising in first-order phase transitions or from the evaporation of cosmic strings, that would yield a much clearer picture of the nature of inflation and the origin of fluctuations.

Addendum: Gravitational Wave Event Rates

The Committee on Gravitational Physics reviewed and reevaluated estimates for two important sources of gravitational radiation in the light of current astrophysical knowledge—binary neutron star mergers, and supermassive black hole mergers. There are many ideas for brighter sources. But the case of binary neutron star mergers represents the minimal, conservative estimate of what might be seen by LIGO. For a space-based detector, the signal from supermassive black hole mergers would be arguably the most exciting scientific discovery. These results inevitably are more technical than those described in the rest of this report.

Binary Neutron Star Mergers and LIGO

An upgraded version of LIGO is expected to become operational around 2005 (see Table 1.1 in Chapter 1). This facility will be able to detect binary mergers out to a significant distance in the universe: an average of 200 megaparsecs for NS-NS binaries, 350 megaparsecs for NS-BH binaries, and 1000 megaparsecs for BH-BH binaries, where NS and BH stand for neutron star and black hole, respectively. (Here it is assumed that the neutron stars have masses of 1.4 solar masses and the black holes 10 solar masses.)

How many binary mergers can LIGO be expected to detect? There are two ways of estimating this.

In one approach, theoretical models of stellar evolution are used to estimate the rate at which binaries of the three kinds form in our Galaxy and other galaxies. NS-NS, NS-BH, and BH-BH binaries form as the end products of the evolution of double stars with masses greater than about 10 solar masses. The binaries of interest to LIGO are those in which the two compact stars are quite close together, such that they merge within a time less than the age of the universe. The evolutionary pathway by which such close binaries form involves a "common envelope" phase in which the cores of the two progenitor stars orbit each other inside a single gaseous envelope. Hydrodynamic interactions during

this stage are complex and poorly understood. Further, the formation of an NS or a BH involves core collapse and (usually) a supernova explosion, with additional uncertainties. It is, therefore, difficult to estimate what fraction of massive stars ends up in the kind of close compact binaries of interest to LIGO. Estimates range over several orders of magnitude, depending on the assumptions made.

The other approach is to make a census of immediate progenitors of merging binaries, and to estimate as directly as possible, from observations, the rate at which these progenitors form. No NS-BH or BH-BH binaries have been discovered so far, so this method does not apply to these systems. However, among the dozens of binary radio pulsars discovered in our Galaxy, four NS-NS binaries are known. Three of these, namely the systems B1534+12, B1913+16, and B2127+11C, will merge in a time less than the current age of the universe. If these three systems represent typical examples of a steady population of merging NS-NS binaries, then the rate of mergers in our Galaxy can be estimated.

The calculation proceeds as follows. Let there be N potentially merging NS-NS binaries in the Galaxy, each with a lifetime given by t_i, which is the sum of the present age of the binary and the time left to merger. By the steady state assumption, the rate of mergers in the Galaxy is

$$R = \sum_{i=1}^{N} (1/t_i).$$

There is little uncertainty in t_i for the three NS-NS binaries listed above, as the merger times can be calculated accurately ($t_i = 2.7 \times 10^9$, 3.0×10^8, 2.2×10^8 years, respectively), and the present ages of the systems (which are typically smaller than the merger times) can be estimated reasonably well. The quantity N is more of a problem since the three NS-NS binaries represent only a fraction of all NS-NS binaries in our Galaxy. Radio pulsars are faint objects, and even the best pulsar searches discover only nearby or relatively bright objects. Therefore, to calculate R, the CGP first estimated, from the observed sample of three systems, the total number of merging NS-NS binaries in our Galaxy.

With a good understanding of the sensitivity limits of various pulsar searches, the regions of the sky covered by each survey, and estimates of the distances to the three NS-NS binaries, it is possible to estimate N. There are several complications, however. First, distances are not usually very accurate. Second, with a sample of only three objects, statistical fluctuations are likely to be large. Third, two of the binaries (B1534+12, B1913+16) are found in the main disk of the Galaxy, while the third (B2127+11C) is in a globular cluster, a somewhat special environment. The disk and the globular clusters need to be treated separately in estimating N. Fourth, it is not known how high above the plane of the galactic disk the NS-NS population extends. The estimates given just below are for a scale height of 1 kiloparsec (kpc).

A number of groups have published more or less independent estimates of R:

- Narayan, Piran, and Shemi (1991): $R \sim 10^{-6}$ per year,
- Phinney (1991): $R \sim 10^{-6}$ per year,
- Curran and Lorimer (1995): $R \sim 10^{-7}$ per year,
- van den Heuvel and Lorimer (1996): $R \sim 3 \times 10^{-7}$ per year; and
- Stairs et al. (1998): $R \sim 10^{-7}$ per year.

These estimates are not necessarily in conflict with one another. Between 1991 and 1995, a number of additional pulsar surveys were carried out, but no new NS-NS binaries were discovered. Also, the distances to pulsars were revised. Both effects led to a reduction in R. Then, in 1998, the distance to B1534+12 was measured more accurately by Stairs et al. (1998), using a new method based on general relativity. The revised distance is larger than the previous estimate, and this again led to a reduction in R. The final rate, $R \sim 10^{-7}$ per year, seems somewhat robust, subject of course to uncertainties due to the small sample.

This estimate must be corrected upward for various reasons. First, pulsars radiate into relatively narrow beams (this is why they pulse in the first place), and we need to allow for pulsars that are inaccessible to our surveys because their beams do not intersect Earth. The correction factor is estimated to be about 6, for pulsars of this general class; a conservative minimum value would be 3. Second, the large tangential velocities observed for the two pulsars in the disk suggest that they belong to a population with a scale height of roughly 3 kpc, rather than the 1 kpc assumed above. Thus, the CGP estimates that the rates should be revised upward by a factor of 18 to include the combination of beaming effects and scale height correction.

Allowing for these effects, the CGP's best estimate of R is $R \sim 2 \times 10^{-6}$ per year. If the beaming correction is smaller or if the observed pulsars have anomalously high velocities, then the correction factor is smaller and the rate can be an order of magnitude lower; the CGP's conservative estimate is then $R \sim 3 \times 10^{-7}$. It is also possible to argue, on statistical grounds, that the estimate of R obtained above is an underestimate of the real rate. Curran and Lorimer (1995) suggest that the rate should be corrected upward by a factor of 10, which seems a little optimistic.

Bailes (1996) has suggested an alternative calculation that places an *upper limit* on the formation rate of NS-NS and NS-BH binaries. There are two classes of pulsars: young pulsars and recycled pulsars. The young pulsars are thought to be a fair sample of young neutron stars. Since none of the roughly 1000 known young pulsars in the Galaxy is in a close binary (NS-NS or NS-BH), the binary fraction of neutron stars is likely to be $<10^{-3}$. Neutron stars are estimated to form at a rate of about 1 per 100 years. Thus, Bailes' argument implies that close NS-NS and NS-BH binaries form/merge at a rate $R < 10^{-5}$ per year.

Once the rate of mergers R in our Galaxy has been estimated, it is simple to estimate how many events will be seen by LIGO. Since at mid-decade LIGO will detect NS-NS mergers out to a distance of 200 megaparsecs, the simplest method

is to count how many galaxies similar to ours are present out to this distance; the answer is ~2×10^5. A more sophisticated calculation (Phinney, 1991) makes use of the fact that NSs and BHs form from massive stars, which produce most of the blue light in the universe. The CGP compared the luminosity of the blue light of the universe out to a distance of 200 Mpc to its best estimate of the blue luminosity of our own Galaxy (which is a factor of 2 lower than the value used in Phinney, 1991); this gives a somewhat larger scaling factor of ~6×10^5. Taking the latter value, the CGP estimates the rate at which LIGO will detect NS-NS mergers to 200 megaparsecs to be R_{LIGO} ~1 every year. Unfortunately, the rate is rather uncertain. The CGP's conservative estimate, including the uncertainty due to small-number statistics, is 1 event every 10 years, while more optimistic assumptions give rates of up to several events per year.

The above estimate is a lower limit for the total rate of detection of binary mergers. There could potentially be additional events due to mergers of NS-NS and NS-BH binaries in which the neutron stars do not produce radio radiation. However, from present data there is no way of assessing the number of such objects. LIGO could also detect mergers of BH-BH binaries. Some recent models of stellar evolution suggest significant rates of these mergers, but there are many uncertainties associated with the estimates.

Supermassive Black Hole Mergers and Space-based Detectors

A supermassive black hole with 2 million times the mass of our Sun lurks in the center of our Galaxy. Recent work with the Hubble Space Telescope has found that most nearby large galaxies contain supermassive black holes in their centers with masses ranging up to a few billion solar masses. These supermassive black holes are thought to be the engines that are powering active galactic nuclei and quasars.

Galaxies are constantly merging with their neighbors. Within the next few billion years, our own Galaxy will merge with the Andromeda galaxy. Andromeda also has a massive black hole in its center. After the two galaxies merge, their central black holes will spiral inward and eventually collide.

The collision of two supermassive black holes releases in 10 seconds 100 million times more energy than the Sun will release in its lifetime. Almost all of this energy will be released in gravitational waves. Understanding the physics of the collisions of supermassive black holes is one of the outstanding problems in classical general relativity.

Because supermassive black hole mergers release such enormous amounts of energy, more than 1 billion times the energy released in a neutron star binary merger, these events could be seen by space-based detectors throughout the visible universe. A space-based observatory that could detect these events would enable researchers to probe the physics of general relativity in the strong regime and deepen our understanding of quasars and the formation of black holes.

The predicted supermassive black hole event rate is uncertain. There are roughly 7 billion elliptical galaxies in the visible universe. If each of these galaxies experienced a single major merger event in its lifetime, then we would expect roughly 0.5 events per year. If most galaxies have experienced multiple mergers, then this rate will be higher. Smaller-mass black holes (~10^6 solar masses) lurk in bulges and dwarf spheroidals. If these systems undergo mergers, the event rate will be another order of magnitude higher, ~5 to 20 events per year. These merger events will generate gravitational wave bursts with amplitude 2 to 3 orders of magnitude above the contemplated sensitivity of a space-based detector. For the million-solar-mass black holes, such a detector will be able to follow their inspiral for roughly 1 month before their final death throes.

References

Bailes, M. 1996, in *Compact Stars in Binaries*, eds. J. Van Paradijs et al., International Astronomical Union, pp. 213-223.

Curran, S.J., and Lorimer, D. R. 1995, *Monthly Notices of the Royal Astronomical Society*, 276, 347-352.

Narayan, R., Piran, T., and Shemi, A. 1991, *Astrophysical Journal*, 379, L17-L20.

Phinney, E.S. 1991, *Astrophysical Journal*, 380, L17-L21.

Stairs, I.H., et al. 1998, *Astrophysical Journal*, 505, 352-357.

van den Heuvel, E.P.J., and Lorimer, D.R. 1996, *Monthly Notices of the Royal Astronomical Society*, 283, L37-L39.

II. BLACK HOLES

Key Questions

Perhaps no other object from physics has had as much impact on public consciousness in recent times as the black hole has. General relativity predicts that black holes will be formed whenever sufficient mass is compressed into a small enough volume. The gravitational force at the surface becomes so large that nothing can escape, no matter how fast it accelerates. Not even a beam of light can escape, hence the name black hole. Black holes are quintessential strong-gravity phenomena.

It is not possible to compress matter on Earth enough to make a black hole. But on astronomical scales, gravity itself can do the job. When a very massive star reaches the endpoint of its thermonuclear burning phase, nuclear reactions no longer supply thermal pressure, and gravitational collapse can proceed all the way to a black hole. By contrast, the collapse of a somewhat less massive star halts at high density when the core is transformed entirely into nuclear matter. The envelope of the star is blown off in a gigantic supernova explosion, leaving the core behind as a nascent neutron star. Gravitational collapse of very massive stars is expected to produce black holes with masses of a few or a few tens of solar masses. Several candidate black holes have been discovered. In addition,

there is evidence for supermassive black holes, with masses ranging from a million to a few billion solar masses, in the centers of galaxies. These large black holes may be produced by the gravitational collapse of a supermassive gas cloud or via the growth of a seed black hole that captures stars and gas from a dense star cluster.

Three key questions confront us about black holes:

• Do black holes actually exist in nature? Are the black hole candidates of astrophysics the black holes of general relativity?
• What are the detailed properties of black holes as predicted by Einstein's theory?
• If black holes do exist, what observations can we carry out to confirm that they have the properties predicted by general relativity?

The discovery of real examples of black holes has been a central goal of gravitational physics for many years. Several excellent black hole *candidates* have been identified by astrophysicists, and we speak loosely of black holes having been discovered. What has actually been discovered are objects with a large amount of mass in a small region. By a process of elimination, we conclude that they must be black holes. But so far there is no convincing evidence that any of the candidates has the distinguishing features of a black hole, such as an event horizon, the one-way membrane that prevents anything from escaping.

A black hole is the most compact configuration of matter possible for a given mass; for a mass M, the size of a black hole is given by the Schwarzschild radius, $R_s = 2GM/c^2$. One way of verifying the compactness of a candidate black hole is by measuring the speed of matter in orbit around it; the speed is expected to approach c in the vicinity of the horizon. This test is feasible since orbiting gas flows, called accretion flows, are common around gravitating objects in the universe.

In a few objects, direct evidence for high orbital speeds is obtained by measuring the Doppler broadening of spectral lines from the accreting gas. In addition, many black hole candidates exhibit gas outflows, or jets, with relativistic speeds. Such motions require an object with a relativistic potential, hence suggesting a black hole. A more indirect indication of compactness comes from observations of strong x-ray emission from the accreting gas. The radiation requires temperatures in excess of 10^9 K, which is most easily achieved with a relativistic object.

When the radiation (typically x-rays) from a compact object varies on a characteristic time scale t, the size of the object must be less than the distance light can travel in this time, ct. If the size limit thus computed is comparable to R_S, the object can be identified as a candidate black hole. For solar-mass black holes, this implies looking for variability on a time scale less than a millisecond, whereas for supermassive black holes the relevant time scale is minutes to hours.

Unfortunately, the demonstration of compactness alone is not sufficient to identify a black hole; a neutron star, with a radius of about 3 times R_S, is only slightly larger than a black hole of the same mass.

Neutron stars can exist happily in equilibrium for small enough masses. But beyond a certain critical mass, the inward pull of gravity overwhelms the balancing pressure force and the star will collapse to a black hole. This then provides one of the key astronomical signatures of a black hole: Look for a system containing a dark massive object. If the object is compact enough and if its mass is greater than the maximum allowed mass of a neutron star, then we infer that it must be a black hole.

The value of the maximum neutron star mass is uncertain theoretically because we do not understand nuclear physics well enough to calculate it reliably. Current conventional nuclear equations of state predict a maximum mass around 2 solar masses, but our confidence in this value is not high. Because of this uncertainty, astrophysicists generally rely on a calculation that assumes we understand nuclear physics up to some density, and then varies the pressure-density relation over all possibilities beyond this point to maximize the resulting mass. This procedure yields an upper limit to the maximum mass of around 3.2 solar masses. Rotation increases the amount of matter that can be supported against collapse, but only by about 25 percent even for stars rotating near breakup speed. Circumventing these limits would require us to accept some unconventional physics—much more unconventional than black holes!

Thus, any compact relativistic object with a mass above about 3 solar masses is considered an excellent black hole candidate. By this criterion, a number of very good black hole candidates have been discovered in the nuclei of galaxies, including our own, and in x-ray binaries. However, the evidence that these candidates are black holes is still somewhat indirect, since what can be demonstrated is only that the objects are not neutron stars.

Direct proof that a candidate is a black hole requires a demonstration that the object has an event horizon, the one feature that is unique to a black hole. Such proof does not currently exist for any astronomical object (but see Box 3.1 on energy advection). The detection of gravitational waves from the final merger of a pair of black holes in a binary will probably provide even firmer proof. Finding absolutely incontrovertible evidence for a black hole in nature would be the capstone of one of the most remarkable discoveries in the history of science.

Achievements

Astrophysical Black Hole Candidates

Black hole candidates have been discovered in x-ray binaries distributed throughout our Galaxy. Each of these x-ray-emitting double stars consists of a compact star that strips gas gravitationally from the outer layers of its more

normal companion. The x-rays are produced when the stripped gas swirls down to the compact star through an accretion flow.

Spectroscopic observations, coupled with a simple application of Newton's laws of mechanics and gravity, allow astrophysicists to set a firm lower limit on the mass of the x-ray-emitting star. In half a dozen x-ray binaries, the lower limit on the mass thus obtained is greater than the maximum mass of a neutron star. These objects are among the best black hole candidates in astrophysics. Several x-ray binaries exhibit rapid variability in their x-ray emission, including very interesting quasi-periodic oscillations that are yet to be understood, and a few exhibit relativistic jets. These observations confirm that the objects are very compact.

The discovery in the 1960s of luminous quasars prompted astrophysicists to consider the possibility that supermassive black holes may have existed at the centers of galaxies when the universe was young. It appeared likely that present-day galaxies would contain dead quasars as massive black holes in their nuclei. Evidence for these black holes has accumulated rapidly in recent years.

The new evidence has emerged primarily from improved observational capabilities in the optical, infrared, and radio bands, both on Earth and in space. Detailed spectroscopic studies of the central regions of galaxies allow us to measure the gravitational effect of a candidate black hole on the line-of-sight velocities of gas and stars. These measurements show that most nearby galaxies contain dark supermassive objects with masses in the range from 10^6 to 10^{10} solar masses.

The nuclei of our own Milky Way Galaxy and the galaxy NGC 4258 (see Figure 3.5) have been studied particularly well since astronomers are able to

---

FIGURE 3.5 *pages 56 and 57* Evidence for the presence of a supermassive black hole in the galaxy NGC 4258. The top panel is an artist's sketch of the molecular accretion disk at the center of this galaxy which was detected by means of its water vapor maser emission. Below this is the spectrum of the emission. The middle picture shows a radio image of the very center of the disk. The small clumps are the images of the maser-emitting water clouds obtained with radio interferometry superposed on a grid representing the unseen portions of the disk. The plot in the lower left shows the line-of-sight velocities of the clouds as a function of position along the major axis of the image. The velocities trace a Keplerian profile, corresponding to a central mass of 3.5×10^7 solar masses. The deviations from Keplerian behavior are so small (less than 0.3 percent) that we can be sure the central mass lies almost entirely inside the inner edge of the disk at 0.13 parsec. The mass density of the object is thus extremely large, greater than 4×10^9 solar masses/parsec^3. Given the large mass and high density, it is hard to imagine the object being anything other than a black hole. The image in the lower right panel shows radio emission from the jets that emerge along the spin axis of the molecular accretion disk (the jets are indicated by the blue cones in the artist's sketch). (Courtesy of James Moran and Lincoln Greenhill, Harvard-Smithsonian Center for Astrophysics.)

56
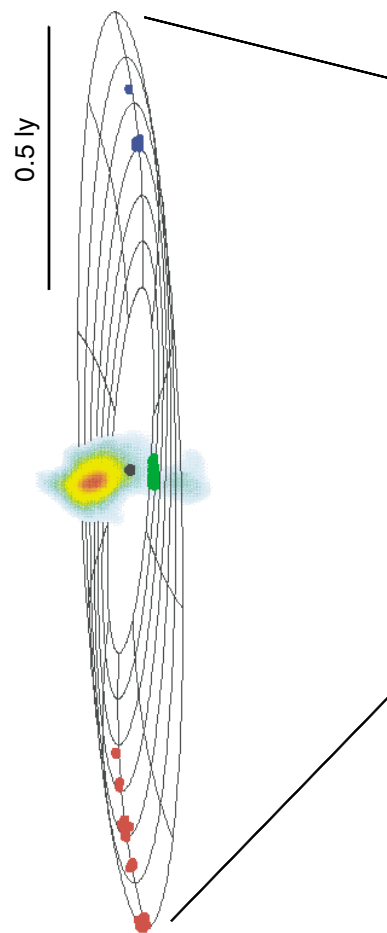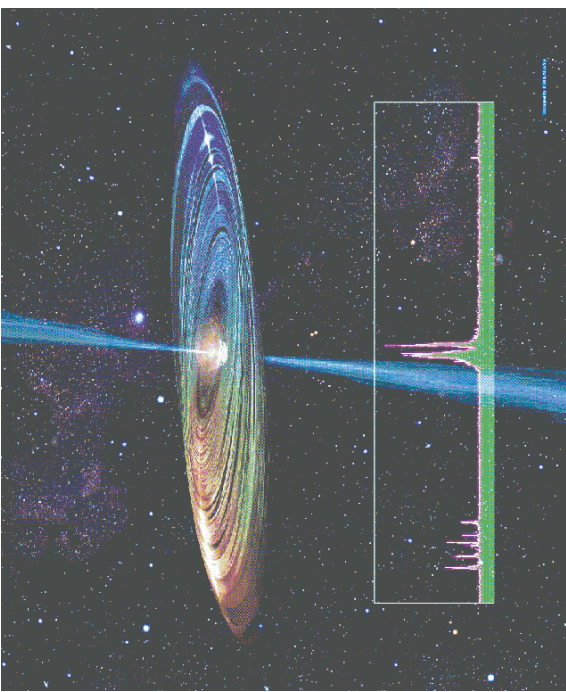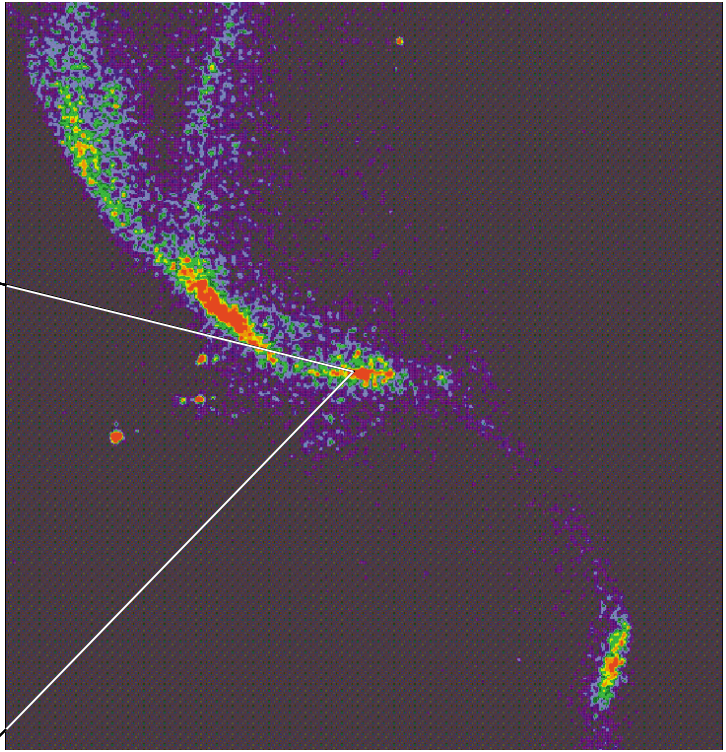
0.5 ly

FIGURE 3.5   *caption page 55*

BOX 3.1   Energy Advection and the Black Hole Event Horizon

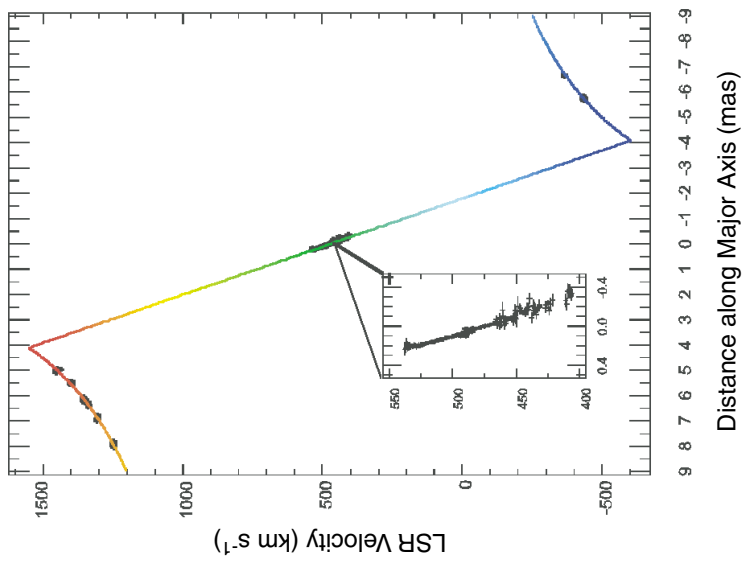Gas accreting onto a compact object such as a black hole invariably has angular momentum and first goes into orbit around the object. Friction in the disk of accreting gas slows its rotation, allowing the gas to spiral into the center. The viscous forces also produce heat, just as ordinary friction both slows down a moving object and generates heat. In the case of accretion onto a black hole or a neutron star, the amount of heat released is very large, about 10^{20} ergs for every gram accreted.

In a thin accretion disk, the gas promptly radiates away the heat released by viscosity. In contrast, in an advection-dominated accretion flow (ADAF), only a small fraction of the heat energy is radiated; the bulk of the energy is transported, or "advected," to the center.

What happens to the energy in an ADAF when it reaches the center? If the central object is a black hole, the energy simply disappears through the event horizon, never to be seen again. On the other hand, if the object has a surface, the energy will be re-radiated; this radiation will dominate the observed radiation from the system and will have a recognizable signature in its spectrum. It is thus possible, in principle, to distinguish black holes from objects with surfaces.

So far, the evidence is that several candidate black holes do have event horizons. This is the most direct indication yet that black holes are present in nature.

The evidence is, however, not conclusive. ADAF models assume without proof a two-temperature gas, where protons and other atomic nuclei are much hotter than electrons. A larger uncertainty has to do with outflows. The evidence for the event horizon is principally one of missing energy. But if the gas were simply to flow out of the system and not accrete, there would be no missing energy. These complex issues need more study.

---

measure both the line-of-sight velocities and transverse velocities of stars and gas. The center of the Milky Way is close, and the galaxy NGC 4258 has strong maser emission with which the velocity of the matter in its central regions can be mapped with precision. The results of these studies are that the center of our Galaxy has a dark object of a few times 10^6 solar masses, and NGC 4258 has an object of a few times 10^7 solar masses. The dark objects in these and other galactic nuclei are much too massive to be neutron stars. Furthermore, several of them emit x-rays and many have jets. For these reasons, they are considered strong black hole candidates.

While there is no doubt that both these classes of objects described above are compact enough to be black holes and are too massive to be neutron stars, scientists cannot yet claim victory. The black hole is one of the most extraordinary objects in all of physics. We cannot accept its reality merely by showing that a candidate object is not a neutron star. More compelling proof is required. Ideally, we would like to see direct evidence that the object has an event horizon. Failing this, we would at least like to see strong evidence that space in the vicinity

of the object is curved in the manner described by Einstein's theory. Some progress has been made on both fronts via astrophysical observations and modeling. Most of the information we receive about black hole candidates is through radiation emitted by accreting gas. Substantial theoretical effort has gone into understanding the physics of accretion flows and modeling the observations. Two kinds of accretion are now recognized to be important. One form, called an advection-dominated accretion flow, has provided a tantalizing hint that some black hole candidates may actually possess event horizons as predicted by general relativity (see Box 3.1). The second form of accretion, called a thin accretion disk, has also proved useful. From Doppler studies of fluorescent iron emission in x-rays, we have preliminary evidence for spacetime curvature near black hole candidates. These studies, which are still in their infancy, are very promising as a probe of the environment around black holes. Spectral models of thin disks in x-ray binaries have also allowed us to estimate the rate of spin of a few black hole candidates. Finally, models of the oscillation modes of thin disks are being compared to observed variability in x-ray binaries. Such studies may provide powerful diagnostics of relativistic effects in the vicinity of black hole candidates.

Advances in Computational General Relativity

Black Holes. While many remarkable properties of black holes have been uncovered by analytic techniques, the astrophysically significant problems of black hole formation and black hole collisions require numerical solution of Einstein's equations. Such simulations can also yield an understanding of other features of Einstein's theory (see the section "Insights into the Structure of Spacetime" below). Already in the 1960s there were attempts to numerically simulate gravitational collapse to black holes and the head-on collision of two black holes. Progress was slow at first, in part because of inadequate computer power but also due to difficulties in understanding many technical issues. The past decade, by contrast, has seen a flowering of computational general relativity. The next decade could see a transformation of gravitational physics as large-scale computing allows the treatment of ever more complicated spacetimes.

A major undertaking of the 1990s was the Binary Black Hole Alliance supported by the National Science Foundation as a Grand Challenge effort in computational science. The alliance was a multi-institution collaboration of computational physicists and computer scientists developing and applying the cutting-edge computational infrastructure necessary to tackle simulations of black hole collisions. The alliance focused on the calculation of the coalescence of two black holes in binary orbit as the system loses energy by gravitational wave emission. This is the simplest problem that includes the full panoply of general relativistic difficulties—gravitational waves, black holes and their associated singularities, and dynamics in three spatial dimensions with no symmetry. It also happens to

be a very important problem for gravitational wave detection (see Section I of this chapter). Moreover, developing the ability to solve this problem would provide a basis for solving other astrophysically interesting problems involving Einstein's equations.

Among the accomplishments of the Binary Black Hole Alliance were the following: (1) development of the "apparent horizon boundary condition" technique to excise a black hole from the computational domain, thereby avoiding the singularity inside; (2) development of new "hyperbolic" formulations of Einstein's field equations; these new formulations have better mathematical properties than the standard formulation and are likely to lead to more accurate simulations; (3) a highly successful collaboration with computer scientists, culminating in the production of the Distributed Adaptive Grid Hierarchy (DAGH), a software package that supports the solution on large parallel supercomputers of the equations encountered in numerical simulations of general relativity; and (4) detailed results of black hole physics, such as the geometry of the black hole surface as black holes merge (see Figure 3.6).

The Binary Black Hole Alliance has taken computational general relativity from simulations in two spatial dimensions to full three-dimensional simulations, and its results suggest that the ultimate goal of simulating something as complicated as two black holes in binary orbit is within reach.

Neutron Stars. The astrophysics of stellar-mass black holes is inextricably bound up with that of neutron stars, whose coalescence to form a black hole may provide an important source of detectable gravitational radiation. The methods of computational general relativity can play a significant role in several important astrophysical problems involving neutron stars.

On the one hand models of binary neutron stars are simpler to construct because they do not have to handle event horizons and singularities, but on the other hand they must deal with the complications of neutron star matter. The NASA Neutron Star Grand Challenge effort to model coalescing binary neutron stars was recently concluded, in part building on the techniques developed by the Binary Black Hole Alliance. The binary neutron star problem is, among other things, of great interest for gravitational wave detectors such as LIGO (see Section I of this chapter).

The structure and stability of rapidly rotating neutron stars represent an astrophysical problem of great importance. Millisecond pulsars spin fast enough that rotational effects are important in their structure, and there is great potential for learning about properties of nuclear matter from observations of rapidly rotating pulsars. During the past decade fully relativistic simulations of spinning neutron stars have become possible with arbitrary nuclear equations of state. There has also been a significant breakthrough in understanding the stability of these models. Even in Newtonian gravity, diagnosing the stability of a rapidly
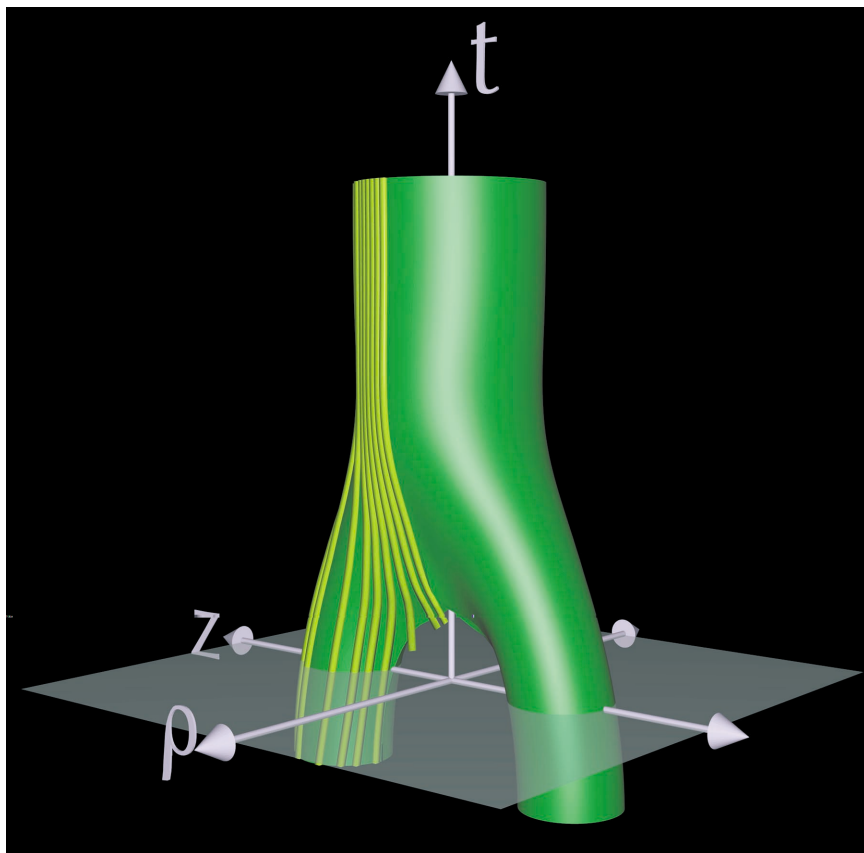
FIGURE 3.6 Head-on collision of two black holes to form a single black hole that becomes spherical at late times. The location of the black hole horizons in the numerically generated spacetime is found by tracing the path of light rays (yellow) that just fail to escape to infinity. Time t is plotted on the vertical axis, z is the symmetry axis, and ρ is the other cylindrical coordinate axis. The coordinate angle ϕ is suppressed in the figure. (Courtesy of Joe Libson, Joan Masso, Edward Seidel, Wai-Mo Suen, and Paul Walker, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.)

rotating star against nonaxisymmetric perturbations is a difficult problem. It was reduced to a tractable form only in 1989. The analogous reduction in general relativity was accomplished in 1996 and is even more complicated. Nevertheless, reliable statements about stability can now begin to be made for the first time.

Black Hole Entropy

The past decade has continued to see exciting progress in our theoretical understanding of the properties of black holes.

It has been known since the early 1970s that black holes satisfy several rules that are very similar to the familiar laws of thermodynamics. In particular, black holes have an analog of the usual temperature and entropy. Over the last 5 years, these thermodynamic considerations were extended from general relativity to a large class of relativistic theories of gravitation.

In the early 1970s, it was believed that these similarities were only mathematical. How could the black hole have a physical temperature since a hot object emits radiation while a black hole is an object from which nothing can escape? This view changed completely with the discovery of Hawking radiation, which demonstrated that, because of quantum effects, black holes do radiate like a hot object.

In all contexts in physics, thermodynamic behavior is the result of averaging over a large number of different microscopic configurations with the same macroscopic properties. In particular, the entropy is related to the number of microstates with the same macroscopic parameters. For a black hole, the entropy turns out to be proportional to its area in units of the Planck length squared—10^{-66} cm^2. This is an enormous number, much larger than the entropy of a corresponding amount of ordinary matter. Physicists have tried for more than 20 years to understand the origin of this enormous entropy. Recently, an explanation has been found in quantum gravity. Quantum states of some black holes can be counted, and the result is in complete agreement with the thermodynamic prediction. Some aspects of Hawking radiation have also been reproduced starting with the microscopic theory. See Section V of this chapter for further discussion.

Insights into the Structure of Spacetime

One of the great open questions in general relativity is the outcome of the gravitational collapse of a general configuration of matter. Powerful theorems require the formation of a singularity from well-behaved initial conditions, if the gravitational field becomes strong enough to trap light. At a singularity, the predictive capabilities of general relativity break down. Indeed, the occurrence of singularities is one of the motivations for the search for a *quantum* theory of gravity, in which such behavior would not happen.

In the simplest, best-known, and analytically calculable example—the spherically symmetric collapse of a pressureless fluid—the eventual singularity is hidden deep within the event horizon of the black hole that is formed. Information from the singularity inside the black hole cannot reach any observer outside the event horizon, so physics can proceed normally outside. What happens in spacetimes without the special symmetries of this example? What happens with more realistic models of the collapsing matter?

An astrophysically interesting possibility is a "naked" singularity—a region of infinite tidal force and breakdown of physical laws that *is* visible (thus "naked") to observers far away from it. Even though quantum effects can be expected to prevent true infinities from actually occurring, the possibility of naked singularities is an issue that must be resolved within classical (not quantum) gravitation to see when the classical theory breaks down. The idea that naked singularities are never formed from realistic initial conditions is known as the cosmic censorship conjecture.

An important development during the past decade has been a synthesis of analytic and numerical techniques to investigate the validity of the cosmic censorship conjecture. Proofs of cosmic censorship were found in certain simplified (but not trivial) cases that assume special symmetries. It was also shown, using sophisticated analytic methods, that an initial state of weak but otherwise arbitrary nonlinear gravitational waves does not develop singularities. The waves disperse, leaving flat spacetime.

Numerical studies of highly nonlinear gravitational collapse of a matter field showed that naked singularities could form if the initial amplitude of the field is exactly right. If the amplitude is larger a black hole forms. To study the transition between the weak- and strong-field results, sophisticated numerical simulations were required. This program led to the 1993 discovery that black hole formation is accompanied by critical behavior, analogous to the critical behavior seen in many physical systems at a phase transition. (In this case, the "transition" is between not forming a black hole and forming a black hole.) Subsequent work has focused on analyzing the generality of this critical behavior in Einstein's equation. This is an example of how the combination of analytic methods and numerical simulation can uncover new fundamental qualitative features of the theory.

Opportunities

Detection and Study of Gravitational Waves from Merging Black Holes

The discovery of signals from the merger of a binary black hole system, and the demonstration that the signals are indeed from black holes, would be a spectacular achievement. All the astrophysical probes described elsewhere in this section provide only indirect information on physics near the event horizon. Gravitational waves from a binary merger, on the other hand, come directly from the sloshing event horizon of the merged object. This is the most direct probe we are likely to have of the very essence of a black hole—its event horizon. The building of gravitational wave detectors sensitive to such signals is, therefore, of the utmost importance for the field (see Section I of this chapter). Confronting theoretical predictions with the observationally determined properties of black holes would provide a scientific bonanza.

Binary black hole-black hole and black hole-neutron star systems, with black hole masses of around 10 solar masses, are very likely to form as a natural consequence of stellar evolution. Some of these systems are expected to merge (as a result of the loss of energy and angular momentum in gravitational waves) within a time less than the age of the universe. The mergers of such binaries in moderately distant galaxies could be detected with gravitational wave receivers such as LIGO. Theoretical predictions of the event rates are uncertain, but it is possible that these might be the first signals that LIGO sees. (See "Binary Neutron Star Mergers and LIGO" in the addendum to Section I of this chapter.)

Binary supermassive black holes, with masses of 10^6 to 10^{10} solar masses, are very likely to form and merge in the centers of many galaxies (see "Supermassive Black Hole Mergers and Space-based Detectors" in the addendum to Section I of this chapter). Such mergers occur naturally in the current paradigm of hierarchical galaxy formation. Mergers of supermassive black holes in even the most distant galaxies in the universe could be easily detected with space-based interferometers. This is one of the strongest reasons for building such detectors.

Computational General Relativity

Computer simulations of Einstein's equations are necessary to calculate the detailed shape (the waveform) of the gravitational waves emitted by the inspiral and merger of binary black holes or neutron stars in the strong-field regime. Such waveforms can be used to improve the chances of detecting signals with LIGO and space-based detectors. In addition, a comparison between the predicted waveforms and those observed is likely to provide the best evidence that black holes have the properties that Einstein's theory predicts.

The Binary Black Hole Alliance moved computational general relativity into fully three-dimensional problems, and the Neutron Star Grand Challenge collaboration built upon that foundation. They proved the value of multi-institution collaborative efforts on problems too large to be solved with the efforts of a single investigator. To realize the opportunity that has thereby been created will require careful attention from funding agencies so that the expertise that has been assembled does not dissipate.

In addition to the human resources, significant access to computing resources will be required. To calculate the waveform from the last orbit and coalescence of a binary black hole system with a modest accuracy of around 10 percent, researchers would need computer runs taking 12 hours on a machine running at 10 teraflops and requiring more than 10 terabytes of memory. Fully three-dimensional black hole calculations will continue to push the envelope of what is possible on the largest machines in the coming decade.

Pulsar searches in the coming decade are likely to become sensitive to periods of 1 ms and less. If nature provides us with such rapidly spinning pulsars,

there is great potential for learning new aspects of nuclear matter at high densities. This will require further development of codes to study the stability of rapidly rotating neutron stars.

In astrophysics, general relativity is routinely invoked to explain observations but is seldom used in calculations because it is too complicated. This will change as algorithms from "pure" general relativity codes are used in conjunction with the microphysics of astrophysics codes. Prime examples involve nonspherical supernova calculations and simulations of neutron star mergers, but it can also be expected that the modeling of phenomena involving quasars, x-ray binaries, and gamma-ray bursts will begin to incorporate general relativity.

Detection and Study of Black Holes by Astrophysical Means

As described in the previous parts of this section, astrophysicists have demonstrated beyond reasonable doubt that there exist objects in the universe that are extremely compact and that cannot be neutron stars. The key goal now is to show that these objects are genuine black holes as understood within general relativity. We must show that the objects warp spacetime in their vicinity in the manner predicted by Einstein's theory. We must show that the objects have that most remarkable of features—the event horizon. To do this, we need to study regions close to the black hole.

Radiation from gas accreting onto a compact object provides many clues about this critical region. It is in the observation and modeling of the radiation, especially in the x-ray band, that the best opportunities are likely to be found.

The spectral shape of x-ray line profiles from orbiting gas flows can be used to measure the warped geometry in the vicinity of a black hole. The study of fluorescent emission from iron atoms, for instance, has already provided preliminary information on the properties of spacetime near a black hole. The expanded capabilities of future x-ray missions should enable researchers to use this and related probes with much greater precision. We should be able to measure the rate of spin of black holes or even detect the dragging of inertial frames by a spinning hole.

Time variability studies of the x-ray emission may also have large payoffs. Noise spectra and the discovery of quasi-periodic oscillations have already led to new insights on accreting neutron stars. The extension of these studies to black holes is clearly the next frontier in this rapidly growing field.

While the above studies will enable us to examine the environment near a black hole, investigations of advection-dominated flows (see Box 3.1) may lead to confirmation of the event horizon itself. To achieve this goal, observations must be done over a wide range of the electromagnetic spectrum, from radio to gamma-rays. The observations are difficult since the systems are very dim, but they are worthwhile, especially if theoretical models can be improved. In particular, it is important to investigate whether the signatures that have been associ-

ated with the presence of an event horizon could be attributed to other effects such as mass outflows.

These studies will of necessity concern a few chosen systems. It is important, however, that we continue to search for more black hole candidates, both in the nuclei of galaxies and in x-ray binaries. Each new system that we discover is in some sense unique. Somewhere out there, doubtless, is a system that could become the "Rosetta stone" of black hole astrophysics. The only way we will find it is by continuing to search with the best telescopes that can be built on Earth and in space. Increasing the population of known black hole systems will also lead to a better understanding of many statistical questions. In particular, the question of event rates for gravitational wave detectors requires more observational input for better estimates.

All this requires that we update and maintain the infrastructure for astronomical observations. X-ray and gamma-ray observatories in space, high-resolution optical and infrared telescopes on Earth and in space, and radio interferometers are all important. Adequate support should also be provided for theoretical work and interpretive modeling.

Analytical Studies of the Structure of Spacetime

Substantial further progress on elucidating compact binary inspirals and collisions, cosmic censorship, critical behavior, and singularities may be expected from combining analytic calculations with simulations from computational general relativity. Analytic work on new formulations of Einstein's equations may benefit numerical relativity and quantum gravity. Insights from pure mathematics will be helpful here, for example in understanding such issues as the long-time behavior of solutions of Einstein's equations.

III. ORIGIN, EVOLUTION, AND FATE OF THE UNIVERSE

Key Questions

Gravity governs the structure of the universe on the largest scales of space and time. Gravity is the weakest of the four fundamental forces, but it is universal, long range, and unscreened. Cosmology and gravitational physics are thus inextricably linked. An understanding of gravity is necessary to understand the universe, and cosmology provides an arena for testing theories of relativistic gravitation.

Cosmology is the subject of a separate document, *Cosmology: A Research Briefing* (National Academy Press, Washington, D.C., 1995), which is part of this decadal survey of physics. It is therefore not necessary for this report to aim at a

description of all of cosmology, nor would it be possible in this brief compass. These are the key questions for gravitational physics in this area:

- Is Einstein's theory right on the largest scales? Is the big bang model the correct description of our universe?
  - What is the origin of the fluctuations that grew to form galaxies?
  - What is the shape and fate of the universe?
  - What is the universe made of?
  - What is the value of the cosmological constant?
  - How did the universe begin?

Cosmology: The Basic Facts

An increasingly detailed, reliable, and consistent web of observations has given scientists a broad-brush picture of the structure of the universe on the largest scales that can be observed and to the longest times that can be extrapolated. These include the observations of the distribution of the galaxies in space (called the *large-scale structure*), the temperature of the cosmic background radiation, and the primordial abundances of the elements. The basic facts about the universe that emerge from these observations are as follows (see Box 3.2):

- *Galaxies are the basic building blocks of the universe.* The distribution of these gravitationally bound collections of gas and billions of stars is homogeneous (the same in one place as in any other) and isotropic (the same in one direction as in any other) when averaged over distance scales above several hundred million parsecs. On smaller scales, galaxies are clustered into groups, filaments, clusters, and superclusters. But in the large, the present universe is much the same in one place as in any other. (One parsec (pc) is roughly 3 light-years or 3×10^{16} meters. The nearest stars are roughly 1 parsec from the Sun. One megaparsec (Mpc) is 1 million parsecs. The nearest large galaxy is about half a megaparsec away. The size of the visible universe is about 3000 megaparsecs.)
- *The universe is expanding.* The galaxies are receding from one another with a speed v which is related to the distance d that separates them by $v = H_0 d$ where H_0 is the Hubble constant, 70 ± 10 km/s per Mpc. Because of this expansion, light from distant galaxies is redshifted to a longer wavelength. Since light travels at a finite speed, observations of distant (large-redshift) objects reveal their appearance in the distant past.
- *The early universe was simpler than the universe today.* It was more homogeneous, more isotropic, and more nearly in thermal equilibrium. The variations in the temperature of the cosmic background radiation in different directions—the earliest structure that we can see—are only a few parts in a million.

68

EVIDENCE FOR THE BIG BANG
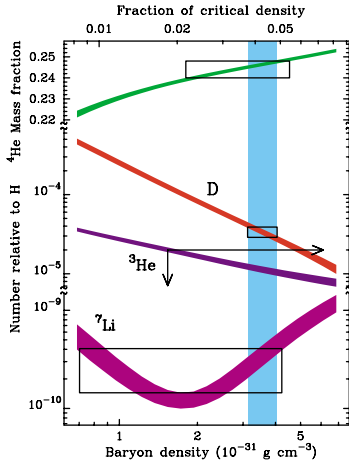
Light Element Abundances
As the expanding universe cooled, chemical elements condensed out of the primeval plasma with predictable relative abundances. The observed abundances of the light elements agree well with these predictions, providing one of the most compelling pieces of evidence for the big bang.

Nucleosynthesis

Quantum gravity

1-10 minutes

? Inflation? 10^{-35} s

10^{-43} s
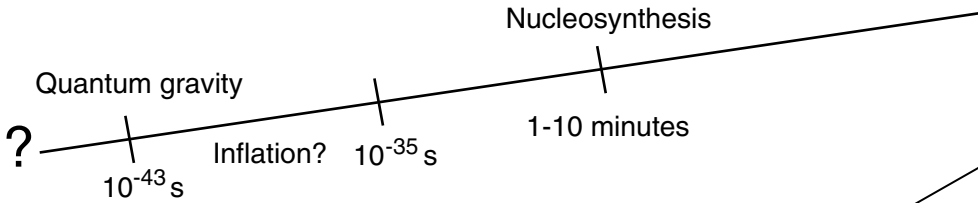
CMB Spectrum
If matter in the universe was compressed in the big bang, it would have been hot, and would radiate like a hot blackbody. Today we see that radiation in the cosmic microwave background (CMB) radiation, much cooled by the expansion of the universe, with the predicted blackbody spectrum. The CMB is the light from the big bang. (The actual data measurements are shown in this figure; the error bars are smaller than the width of the blue line, which indicates the predicted blackbody spectrum.)
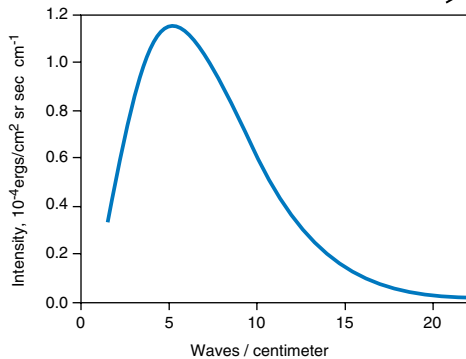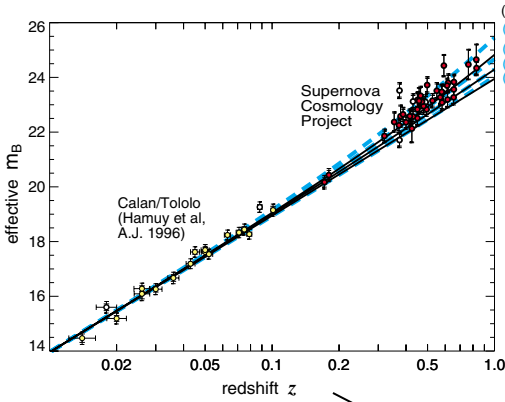
Expanding Universe
The redshift of a galaxy is a measure of the speed at which it is receding from us. A plot of effective apparent magnitude (m_B; a measure of relative distance) versus redshift (z) shows that the universe was expanding from a compact state approximately 13 billion years ago, i.e., the big bang. Deviation from a straight line at large redshift measures whether the expansion is slowing down or speeding up.

CMB produced

Today

10^6 years
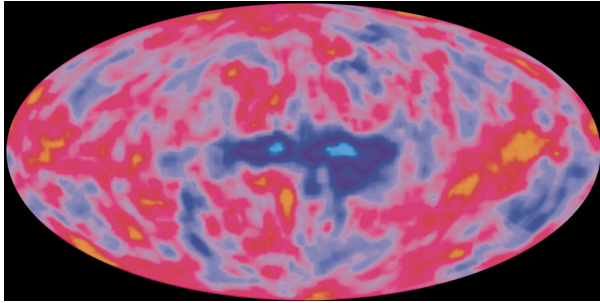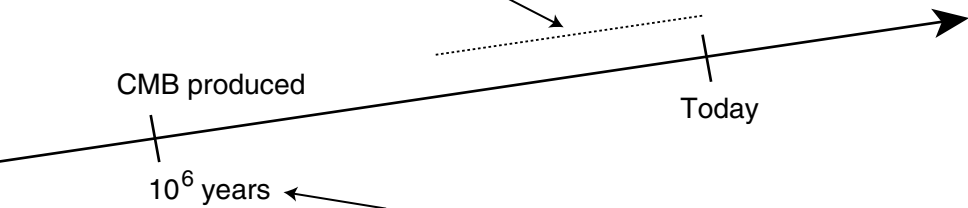
CMB Fluctuations
The universe at the big bang was almost perfectly smooth, but it must have contained the seeds that condensed by gravity to evolve into today's galaxies. These seeds are seen in tiny 30-millionths-of-a-degree fluctuations in the temperature of the cosmic microwave background shown here as different colored regions in a map of the sky.

BOX 3.2 Evidence for the Big Bang. SOURCES FOR FIGURES: Light element abundances—Scott Burles, University of Chicago (Burles, S., et al., 1999, *Physical Review Letters*, 82, 4176-4179); expanding universe—Saul Perlmutter, Lawrence Berkeley National Laboratory (Perlmutter, S., et al., 1999, *Astrophysical Journal*, 517, 565-585); CMB spectrum, CMB fluctuations—NASA's Goddard Space Flight Center and the COBE Science Working Group.

The Intersection of Gravitational Physics and Cosmology

These basic observations, when combined with general relativity, give a powerful theoretical framework for understanding both how the universe is structured and how it evolved to be that way. Modern cosmology lies clearly within the realm of strong gravitational physics. The strongest curvatures in the universe occur in the initial big bang. Some of the places where gravitational physics and cosmology intersect most strongly are described below.

The Evolution of the Universe. When the observed approximate symmetries of homogeneity and isotropy are enforced exactly, Einstein's theory predicts the Friedmann-Robertson-Walker (FRW) cosmological models (see Chapter 2). These models are characterized by only a few parameters. The first is the scale of the expansion rate, which we can take to be the present value of the Hubble constant H_0. The other important parameters are the present-day energy densities in matter and radiation, usually expressed as the ratios Ω_m and Ω_r, respectively, of their values to the critical density required to close the universe. In addition, there is the value of the cosmological constant Λ. A cosmological constant could arise from many sources, including the energy of empty space (the vacuum). The value of the cosmological constant is important not only for classical gravity and cosmology, but also for quantum gravity. The smallness of the cosmological constant (less than 10^{-128} in the Planck units that characterize quantum gravity; see Chapter 2) remains a mystery whose resolution likely awaits the development of the "final theory" uniting quantum theory with gravity.

As mentioned above, our observations of the distant universe reveal its past. Detailed observations of the properties of galaxies at different distances should thus determine the parameters of the best-fit FRW model that describes the evolution of the universe in time and also should measure the value of the cosmological constant. With sufficiently detailed observations the FRW parameters would be *overdetermined* so that it would then be possible to check whether the FRW models are correct, that is, to test Einstein's theory on this largest of possible scales.

The Growth of Inhomogeneities. Galaxies, clusters of galaxies, and the voids and filaments in their observed distributions are examples of inhomogeneities in the universe. The anisotropies in the background radiation are evidence of the progenitors of these present inhomogeneities seen at a much earlier time. The prevailing opinion is that all of today's inhomogeneities grew by gravitational attraction from tiny quantum deviations from the smoothness of the early universe. Different theories for the origin of the structure and the composition of the universe make different predictions for the spatial distribution of galaxies and clusters of galaxies as well as for the statistical properties of the microwave background. By comparing observations of these inhomogeneities and anisotro-

pies to theory, we can test different ideas for the origin of density fluctuations and for the composition and evolution of the universe.

Gravitational Lensing. The gravitational attraction of a mass will deflect the course of a passing light ray. This prediction of general relativity was one of the first to be tested observationally in 1919 when Eddington observed the gentle deflection of starlight by the Sun. Much stronger examples of the same phenomenon are now known. A large concentration of mass can act as a lens producing one or more images of a source behind it. (See Figure 3.7.) For example, more than a dozen quasars lensed by intervening galaxies have been observed.

    The lensing of quasars by intervening galaxies not only provides a dramatic

FIGURE 3.7   An image of multiple arcs in galaxy cluster 0024+1654, taken by the Hubble Space Telescope. The mass in the foreground cluster (the bright clustered objects) acts as a gravitational lens and distorts the shape of background galaxies into arcs. The cluster lens is so powerful that it produces five images of the same galaxy. (Courtesy of J. Anthony Tyson, Lucent Technologies, Wesley N. Colley, Harvard-Smithsonian Center for Astrophysics, Edwin L. Turner, Princeton University, and NASA.)

confirmation of one of the predictions of Einstein's theory but also in principle allows the verification of its predictions for the propagation of light in cosmological spacetime geometries. There are three major roles that lenses play in cosmology:

• *Determining the distribution of matter in galaxies and clusters.* A gravitational lens distorts and amplifies background images by amounts that depend on the quantity of imaging mass and the distances to the source and lens. Weak lensing observations of galaxies and quasars can therefore be used to map the distribution of matter in galaxies and clusters.

• *Measurement of cosmological parameters.* If the foreground cluster or galaxy is a powerful enough lens, it can produce multiple images of the same source. The light's travel time for the different images can differ by several hundred days and depend upon the mass of the lens and the distances to the source and lens. Such timing observations allow astronomers to determine distances to lensing systems and thereby to determine the Hubble constant and constrain other cosmological parameters, especially the cosmological constant.

• *Determining the nature of the dark matter.* By observing stars in nearby galaxies and looking for "microlensing events" produced when a massive object passes between Earth and a star, astronomers can detect such massive compact objects in the neighborhood of our Galaxy.

Gravitational Waves from the Early Universe. The early universe is a source of gravitational radiation. Many of the theoretically proposed processes that might have shaped the evolution of the early universe produce a gravitational wave background that is potentially detectable today. Inflation parametrically amplifies quantum fluctuations in the geometry of spacetime to produce a gravitational wave background. Gravitational radiation is an important mode of decay of cosmic strings—topological defects formed in phase transitions in the early universe. Collisions of bubbles produced in any phase transition at the end of an inflationary epoch may have been a copious source of gravitational radiation. Each of these processes leads to a gravitational wave background with a characteristic amplitude and non-thermal spectrum today. The observation of any of these gravitational wave backgrounds would be of enormous importance for cosmology because it would give us a picture of the universe at a much earlier stage than any available in electromagnetic radiation. The electromagnetic background radiation originated about 300,000 years after the big bang when the universe became transparent. By contrast, gravitational radiation interacting much more weakly may have originated as early as 10^{-22} seconds after the big bang.

Quantum Cosmology. Cosmology presents a problem that is fundamentally different from those encountered elsewhere in physics. This problem is the need

for a theory of the initial condition of the universe. The familiar laws of physics, such as Newton's second law, Maxwell's equations, or the Einstein equation, describe evolution in time. Such dynamical laws require boundary conditions to yield predictions, and the Einstein equation governing the evolution of the universe is no exception. Usually boundary conditions summarize observations of the part of the universe outside a subsystem whose evolution is being studied. But in the study of the whole universe there is no "rest of the universe." The cosmological boundary conditions must be part of the laws of physics themselves. A theory of the cosmological boundary condition is as necessary a part of any "final theory" as a fundamental theory of the laws governing evolution in time.

In a quantum theory, a theory of the initial condition of the universe is a theory of its initial quantum state. That is why the area of astrophysics concerned with this initial condition is called *quantum cosmology*. What is the quantum state of the universe? Does it arise from a fundamental principle that explains the simplicity of the early universe revealed by observations? Is it connected with the fundamental dynamical theory? These are the kinds of questions quantum cosmology seeks to answer by working at the intersection of the disciplines of cosmology, quantum mechanics, and quantum gravity. In quantum cosmology the physics of the universe's largest scales are united with the physics of its very smallest.

Achievements

Tests of the Big Bang Model

NASA's Cosmic Background Explorer (COBE) satellite measured the energy distribution of the microwave background radiation. COBE showed that this distribution was thermal to an accuracy of 1 part in 10,000. A thermal spectrum is one of the fundamental predictions of big bang theory. Since alternative cosmological theories have not been able to explain the existence of this nearly uniform thermal energy, this observation has become one of the pillars of the big bang picture.

Astronomers have measured the redshifts to more than 100,000 galaxies, a more than 10-fold increase in the past decade. They found that the galaxy distribution is clustered on scales as large as 100 million parsecs. On even larger scales, the distribution of galaxies appears to be uniform. These observations finally confirmed that the universe is nearly homogeneous on large scales—much the same in one place as in any other.

With a successfully repaired Hubble Space Telescope and the new Keck 10-meter telescopes, astronomers have been able to observe galaxies at high redshift. These have lower abundances of carbon, oxygen, and iron, elements that are produced through stellar evolution, and also are more irregular in shape

and show much more evidence of star formation. Thus, as predicted in the big bang model, the distant galaxies appear younger.

Observations of Primordial Fluctuations

The COBE satellite also detected very small directional variations in the temperature of the microwave background of only 30 millionths of a degree. (See Figure 3.8.) These observations directly probe the distribution of matter only 300,000 years after the big bang, the moment when electrons and protons first combined to form hydrogen, and the universe became transparent to most electromagnetic radiation. Ground- and balloon-based microwave background experiments have confirmed the COBE result and detected additional fluctuations on smaller angular scales.

Estimates of the Size of the Universe

Astronomers have significantly improved determinations of the size, density, and geometry of the universe. The Hubble Space Telescope has monitored

$$\Delta T = 18 \mu K$$
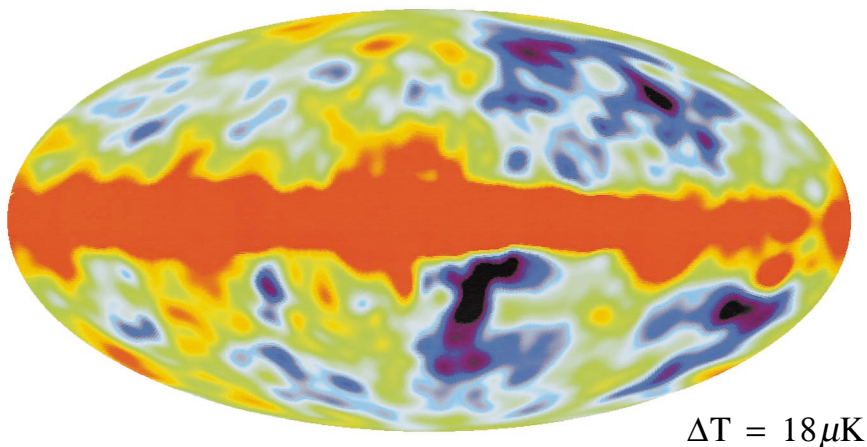
FIGURE 3.8 COBE's map of the microwave sky. The entire sky is shown in this false-color map in galactic coordinates. Radio emission from gas and dust in the disk of our own Galaxy appears as a strip across the middle of the map. Away from this strip, the fluctuations are due to tiny variations (1 part in 100,000) in the temperature of the cosmic background radiation. If the inflationary model is correct, these fluctuations have their origins in processes that occurred 10^{-36} seconds after the beginning of the big bang. (Courtesy of NASA's Goddard Space Flight Center and the COBE Science Working Group.)

variable stars in nearby galaxies whose well-established relation between their period of variability and their intrinsic brightness allows them to be used to measure the distance to nearby galaxies and to infer the Hubble constant, H_0 (the expansion rate of the universe). Measurements of the motions of galaxies enabled astronomers to infer the mean density of the universe. While the data is not yet definitive, these measurements suggest that the density of the universe in matter is not enough to stop the expansion of the universe. Astronomers are also using Type Ia supernova explosions in distant galaxies to probe the geometry of spacetime. These observations also suggest that the universe will expand forever and even accelerate its present expansion.

Dark Matter

Astronomers can measure the mass of galaxy and galaxy clusters through a variety of techniques. All of these techniques find that the gravitational mass of a galaxy exceeds the mass in luminous stars by at least a factor of 10. Astronomical searches had ruled out cold gas, warm gas, hot gas, and dust as possible candidates for this mysterious dark matter. In the past few years, gravitational microlensing searches have ruled out two other candidates—planets and brown dwarfs (bound objects with masses greater than Earth's but less than 8 percent of the mass of the Sun) as possible dark matter candidates.

Origin of Elements and the Number of Light Neutrinos

The big bang model successfully explains the origin of primordial deuterium, helium, and lithium. These light elements were produced in the first minutes of the universe. Heavier elements such as carbon, oxygen, and iron were produced subsequently by nuclear burning in stars. The big bang explanation for the abundances of deuterium and helium fits the data only if there are no more than three light neutrinos. During the past decade, measurements of the properties of the Z boson at CERN (the European Laboratory for Particle Physics) successfully confirmed this prediction.

The Inflationary Paradigm

Over the past 15 years, cosmologists developed the inflationary universe paradigm, amalgamating ideas from cosmology and particle physics. This extension of the big bang theory posits an early epoch in which vacuum energy drove an ultrarapid expansion of the universe. This rapid expansion inflated a tiny region of space into the entire part of the universe visible today, smoothing any primordial variations in the geometry of space, and driving the density of the universe toward the critical density required for closure. The inflationary paradigm goes far toward explaining the homogeneity, large size, and age of the

universe. Because of initial quantum fluctuations, inflation proceeds at slightly different rates in different regions of the universe, generating spatial variations in its density, and in turn, these weak density fluctuations produce temperature variations in the microwave background that are seen by COBE (see Figure 3.8). These fluctuations eventually grow through their own self-gravity to form galaxies, clusters, and other large-scale structures that we see today.

The specific predictions of the inflationary scenario depend on the composition of the universe and the details of particle physics. In studying inflation, cosmologists tie together and test ideas from grand unified theories and from quantum gravity. Over the past 20 years, theorists have developed the conceptual and numerical tools needed to make detailed quantitative comparisons of the variants of the inflationary model with the data, primarily the distribution of galaxies. By the mid-1980s, cosmologists realized that models in which most of the matter in the universe was in the form of neutrinos (generically called "hot" dark matter) were not consistent with the large-scale structure observations. The "cold" dark matter model, which posited that galaxies grew from inflationary fluctuations in a critical density universe filled with weakly interacting massive particles, emerged as the standard theory for the origin of structure. While this model can fit the basic qualitative features of the microwave background data, it cannot consistently fit the detailed quantitative features of the large-scale structure and the microwave background data. Nor can the current data be fit by models of structure formation in which phase transitions in the early universe generate topological defects that seed the formation of galaxies and large-scale structure.

The currently most successful models of structure formation are inflationary models with a subcritical density of matter. These low-density-universe models either posit the existence of a cosmological constant to make the universe flat or posit a modified form of inflation that produced a negatively curved universe. Either way, they suggest the existence of new physics.

Opportunities

In the coming decade astronomers and physicists will acquire a wealth of new data that will enable testing and refinement of current ideas and, perhaps, lead to new models for the origin of the universe.

Precision Cosmological Measurements

The U.S.-led Sloan Digital Sky Survey and the Anglo-Australian 2DF project will measure the redshifts of more than 1 million galaxies and quasars, a 10-fold increase in the number of measured redshifts over that available today. These observations will enable astronomers to quantify to new accuracy the level of inhomogeneities in the galaxy distribution. Observations of gravitational lensing

of distant galaxies and clusters will also directly quantify the level of inhomogeneities in the mass distribution.

NASA's Microwave Anisotropy Probe (MAP), which is scheduled for launch in late 2000, and the European Space Agency's Planck Surveyor, which is scheduled for launch in 2007, will map the microwave background radiation across the whole sky with angular resolution 30 to 60 times better than that obtained with the COBE satellite.

These microwave background observations will measure the level of inhomogeneities in the universe 300,000 years after the big bang. The combination of these data sets will enable astronomers to describe the statistical properties of density fluctuations and their evolution with very high precision.

Inflationary theories make very specific predictions for the statistical properties of the microwave background fluctuations and for the statistical properties of the density fluctuations traced by galaxies. This confluence of experimental accuracy and theoretical calculability will enable cosmologists to finally answer some of the fundamental questions that span astrophysics, gravitational physics, and particle physics:

- *What is the origin of the fluctuations that grew to form galaxies?* Inflation makes very specific predictions for the microwave background. If they are confirmed, then this will be a major success for the inflationary paradigm. If the upcoming observations are not consistent with the inflationary model, then we will need to find an alternative theory for the origin of the universe.

- *What is the shape and fate of the universe?* By measuring the characteristic angular size of a typical hot spot in the microwave background, we can measure the geometry of the universe. The physical diameter of the spot is fixed by the speed of light and the age of the universe when electrons and protons combined to make hydrogen, and so is related to the geometry. In turn, general relativity directly relates the geometry of the universe to the density of matter and the value of the cosmological constant.

- *What is the universe made of?* By measuring the characteristic intensity profile of the typical hot spot, we can measure the sound speed in the early universe. The sound speed depends on the ratio of density in electrons and protons to density in radiation, a central number in cosmology. By measuring the characteristic physical scale of galaxy clustering, we can determine the ratio of energy density in matter to the energy density in radiation, a central cosmological parameter.

- *What is the value of the cosmological constant?* By measuring the relative amplitude of temperature fluctuations and matter fluctuations as a function of size, we can measure the rate of growth of gravitational fluctuations. This growth depends sensitively on a combination of cosmological parameters, particularly the cosmological constant and the nature of the dark matter.

There are several ongoing complementary approaches to answering these cosmological questions. Large-scale structure observations measure the distribution of galaxies and their motions. Lensing observations measure the statistical properties of dark matter. Microwave background observations measure the distribution of matter and radiation in the early universe. Type Ia supernovae, which have already been detected out to large redshifts, can be used to measure the cosmological constant through the luminosity-redshift relation. Gravitational lenses, which probe the angular distance-redshift relation, will provide another opportunity to measure this important parameter. If all the observations agree, then astronomers may have measured a number that might be predicted by the correct quantum gravity theory.

• *Is Einstein's theory right on the largest scales?* This wealth of data can also be used to test Einstein's equations on the largest of possible scales. If the FRW model is indeed the correct cosmological model, then different techniques will yield consistent values for cosmological parameters and thereby will verify general relativity on the largest possible scales.

Probing the Composition of the Universe

In the coming decade, laboratory techniques should achieve the sensitivity needed to detect supersymmetric dark matter (a kind of cold dark matter) in underground direct detection experiments. Neutrino telescopes will search for high-energy neutrinos produced through the annihilation of supersymmetric particles trapped in Earth and the Sun. A detection will have profound implications for our understanding of the universe. Particle physicists will test supersymmetry directly at the Large Hadron Collider (LHC). If supersymmetry is not detected at the LHC, the lightest supersymmetric particle will no longer be the most attractive dark matter candidate. If supersymmetry is detected, then measurements of supersymmetric parameters will lead to a prediction of its cosmological abundance.

With the use of larger telescopes and larger cameras, gravitational lens measurements will also be able to characterize the distribution of mass in our own Galaxy. Microlensing experiments have already detected tens of lensing events, whose nature is a mystery. In the coming decade, astronomers will likely detect hundreds more. With this much larger data set, astronomers should be able to uncover the nature of the lensing objects and determine if they make a significant contribution to the total mass of our Galaxy.

Windows onto the Beginnings of the Universe

Observation. The development of LIGO and other ground-based detectors, along with a future space-based interferometer, will open a new window into the earli-

est moments of the big bang. As described more fully in the "Key Questions" part of this section, there are a number of potential sources of gravitational radiation in the early universe—rapidly varying fields during inflation or another phase transition, for example. Such sources could produce a gravitational wave background potentially detectable by the methods described in Section I of this chapter. Our understanding of the physics of the universe when these waves were formed is primitive, so we cannot as yet predict with any confidence the expected amplitude of this signal. However, if the universe did begin with violent physics, then the gravitational waves will bear the imprint of these first moments. That makes their detection interesting even if it is not clear whether this is an opportunity for the next decade or future ones. It is reasonable, therefore, to search for these probes of the earliest instants of our history over as wide a spectrum of wavelengths as possible with existing instruments. Even upper limits in a range of wavelengths can provide important constraints on early-universe physics.

Theory. Several promising theoretical directions may lead to a deeper understanding of the origin of the universe and its basic physical laws. The continued development of non-perturbative formulations of string theory and canonical quantum gravity (see Section V of this chapter) will give, for the first time, workable quantum theories of gravity that can be applied with confidence to the quantum era of the universe's evolution and will provide a framework for formulating and exploring theories of the initial condition. Quantum cosmology offers the prospect that, in the next decade, we may not only understand how the universe is structured on the largest scales of space and time, but also begin to understand, as a basic law of physics, why it is the way we see it today.

IV. GENERAL RELATIVITY AND BEYOND: EXPERIMENTAL EXPLORATION

Key Questions

Einstein's gravitational theory is more than 80 years old, yet until the 1960s, there was only modest progress in testing it. This was due to the fact that, for the majority of physical systems so far accessible to careful measurement, its predictions differ from those of standard Newtonian gravity by only minute amounts. Consequently, testing the theory has always been a challenging task. Only in the past few decades have truly precise tests been made possible by the rapidly evolving technology of high-precision measuring tools, such as atomic clocks, lasers, radio telescopes, and torsion balances. At the same time, theoretical advances have made it possible to understand clearly the observable predictions of the theory, and to compare and contrast them with possible alternative theories.

80   *GRAVITATIONAL PHYSICS: EXPLORING THE STRUCTURE OF SPACE AND TIME*

Despite great progress on both the experimental and theoretical fronts, key questions remain open:

• How "right" is general relativity? Will further experimental tests continue to agree with the theory, or will a deviation appear at some level?
• If a deviation appears, will that signal a new theory of gravity, new physical interactions, or new elementary particles?

The possibility of new physics is not idle speculation; it is motivated by today's searches for theories of the fundamental interactions. For example, string theory suggests that general relativity will fail at some level, although the level is not yet predictable. One result of this failure would be a violation of the principle of equivalence (see Box 3.3). Another connection is through the cosmological constant. Observations explained by a cosmological constant might also be explained by an evolving, low-mass scalar field (although the "constant" would then be time dependent). Such a scalar field might very well show up in tests of the equivalence principle. Thus, although many actual experiments will take place under conditions of weak gravity, they will be searching for the imprints there of physics at strong gravity scales. Detection of such imprints would be a profound discovery.

Achievements

Prior to 1960, the empirical basis of general relativity consisted of the body of evidence supporting special relativity, experiments that verified the principle of equivalence underpinning general relativity, and two experiments that checked the theory itself: the deflection of starlight and Mercury's perihelion advance. The latter two experiments were regarded as being good to accuracies only between 10 and 50 percent. Since 1960, dramatic progress was made, both in improving the precision of existing tests and in performing new high-precision tests. Here the CGP focuses on developments that have occurred since the release of the last decadal survey (*Gravitation, Cosmology, and Cosmic-Ray Physics*, National Academy Press, Washington, D.C., 1986).

Tests of the Universality of Free Fall

Improved tests were performed of the equality of acceleration of different bodies, or the universality of free fall (UFF), which served as the inspiration for the equivalence principle. This principle is the foundation for the geometric theories of gravity like general relativity. From a quantum viewpoint, it is a very special consequence of theories in which the force is produced by a spin-two graviton. New spin-zero and spin-one particles and their couplings that are

---

BOX 3.3 Einstein's Principle of Equivalence

Einstein generalized the feature of Newtonian mechanics that all bodies in a gravitational field fall with identical accelerations, into a guiding principle for his relativistic theory of gravity. If all objects have the same acceleration, gravitational effects are equivalent to an acceleration of the frame of reference. Thus a freely falling observer will find that gravitational effects disappear. In Newtonian terms, this is a consequence of the exact equality of the inertial mass (the mass in $F = ma$) and the gravitational mass (the mass in $F = GMm/r^2$). This equivalence principle, or universality of free fall (UFF), leads directly to many of the celebrated predictions of general relativity such as the existence of spacetime curvature, and the gravitational redshift of clocks. The UFF is a special property of general relativity and is expected to be violated by quantum mechanical forces that typically arise in unified field theories.

Because of the deep importance of the equivalence principle, and the possibility that UFF violation could be used to discover new tiny forces, the UFF has been tested in a number of ways. Precise tests compare the gravitational accelerations of different test bodies in the field of attractors such as the Sun, Earth, local topography, laboratory-scale movable masses, and the dark matter in our Galaxy. The most sensitive experiments use torsion balances (see Figure 3.9) that produce no signal when two different types of test bodies have identical accelerations. Currently, torsion balances and lunar laser ranging are able to compare accelerations with a precision of 5×10^{-13} cm/s^2. An object starting 2000 years ago from rest with a constant acceleration of 5×10^{-13} cm/s^2 would now be moving at the same speed as the end of the minute hand on a wall clock.

Other measurements of extraordinary precision have been exploited to test that gravitational self-energy itself obeys the UFF. Gravitational self-energy is completely negligible for laboratory-sized objects; even for objects the size of Earth and the Moon it is very small—only 46 parts and 2 parts in 10^{11}, respectively. Earth and the Moon also differ in that Earth has an iron-nickel core, whereas the Moon does not. The relative accelerations of Earth and the Moon toward the Sun have been extracted from the lunar laser-ranging data, while the relative accelerations of test bodies with "Earth-like" and "Moon-like" compositions have been checked using torsion balances, permitting possible violations of UFF by gravitational self-energy to be isolated. None was found down to the level of about 2 parts in 1000.

---

typical of unified theories (such as string theory) produce violations of the equivalence principle at some level and length scale.

Tests of the UFF reached the level of a part in 10^{12} and attained sufficient sensitivity to demonstrate that the interaction of ordinary matter with galactic dark matter obeys the UFF to a part in 1000. Many laboratory experiments of this type performed since 1986 were done specifically to search for new forces predicted by theories of the elementary particles or by string theories. Ongoing laser ranging to corner reflectors on the Moon provided ever improving tests of the
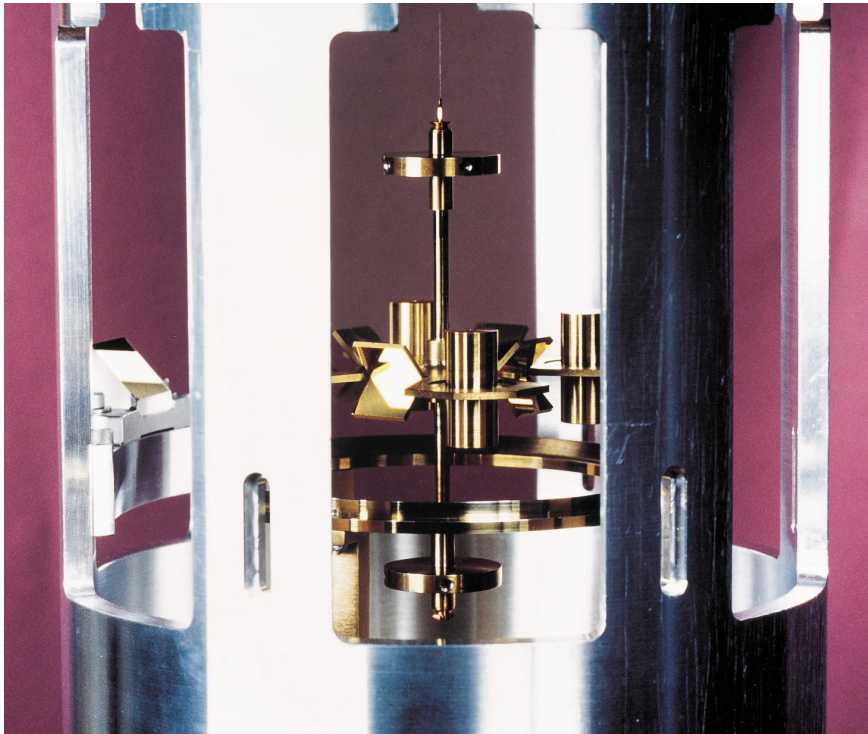
FIGURE 3.9   Torsion pendulum used to test Einstein's equivalence principle for test bodies attracted by Earth, the Sun, and the dark matter in our Galaxy. The pendulum is small (its overall diameter is about 3 inches) to minimize disturbing effects from local variations in the gravitational force. It hangs from a tungsten fiber that is so thin that it cannot be seen in the photograph above. The circular plate holds four cylindrical test bodies (two of copper and two of beryllium) along with four right-angle mirrors (see inset) that are part of a sensitive optical system for detecting pendulum twists. The annular ring underneath the pendulum is a "safety net" to catch the pendulum should a small earthquake shake the apparatus and break the suspension fiber. The pendulum is suspended in a vacuum and the entire instrument is rotated continuously at about one revolution per hour. A violation of the equivalence principle would show up as a pendulum twist that varied at this rotation frequency. (Courtesy of Eric Adelberger and the Eöt-Wash Group, University of Washington.)

equality of acceleration of the Moon and Earth toward the Sun, to a level today that is actually slightly better than that from laboratory experiments. The lunar results also verified to a part in 1000 that bound gravitational energy within Earth and the Moon falls with the same acceleration as mass and other forms of energy, providing a direct test of general relativity.

Another test of the equivalence principle was made possible by techniques developed within atomic physics, such as atom and ion traps, and laser cooling, which were used to put exquisitely stringent constraints on any anisotropy, or preferred direction, in local physics that might be generated by new cosmic interactions.

Pulsar Tests of Relativistic Gravity

The Hulse-Taylor binary pulsar provided a definitive test of the existence of gravitational waves, in agreement with the prediction of general relativity to a third of a percent. Because the stars in the binary pulsar system are neutron stars, with strongly relativistic, nonlinear internal gravitational fields, the observations also provided indirect support for the theory in the strong-gravity regime. Tests of the UFF, of momentum conservation, and of anomalies related to a hypothetical preferred cosmic reference frame were also performed using binary pulsars and millisecond pulsars.

Search for Frame Dragging

The last decade also witnessed the development of a number of approaches to measuring the twisting up of spacetime in the vicinity of rotating bodies, known as the dragging of inertial frames, or the Lense-Thirring effect. This effect has not been measured directly to date, although aspects of it are implicit in other observed relativistic effects, such as lunar and planetary motion, and the motion of the Hulse-Taylor binary pulsar. Not only is this effect important as a prediction of general relativity, but it also has deep conceptual implications for the meaning of absolute rotation. Furthermore, the forces associated with frame dragging are thought to play an important role for jets of matter that are seen being ejected from quasars and active galactic nuclei, coming from the rotating, supermassive black holes believed to reside there.

The main ongoing project to measure frame dragging accurately is the NASA Relativity Mission, informally known as Gravity Probe B, in which gyroscopes are to be placed in orbit around Earth and their precessions relative to distant stars measured. (See Box 3.4.) This experiment is expected to measure the Lense-Thirring frame-dragging effect to a precision of about 1 percent. It will also measure the larger geodetic precession caused by space curvature around Earth to a precision that could reach a thousandth of a percent. The project was conceived and begun in the 1960s. Since the release of the last decadal survey of

BOX 3.4 Gravity Probe B

Gravity Probe B (GP-B) is a space project to measure the tiny precession of gyroscopes relative to distant stars. The precession is induced by the twisting of spacetime caused by the rotating Earth (dragging of inertial frames) and by the curvature of space around it. The satellite is scheduled for launch in 2000.

The GP-B gyroscopes are spheres of fused quartz, 1.5 inches in diameter. Each gyroscope is electrically suspended by applying voltages to saucer-shaped electrodes in the two halves of the housing. It is spun up to 150 Hz by gas running through a channel in the right-hand hemisphere, after which the gas is pumped out and the ball runs freely in a vacuum. The direction is read out by a superconducting quantum-interference device (SQUID) connected to the superconducting circuit on the face of the left-hand hemisphere. (Photograph courtesy of Ms. Denise Freeman.)
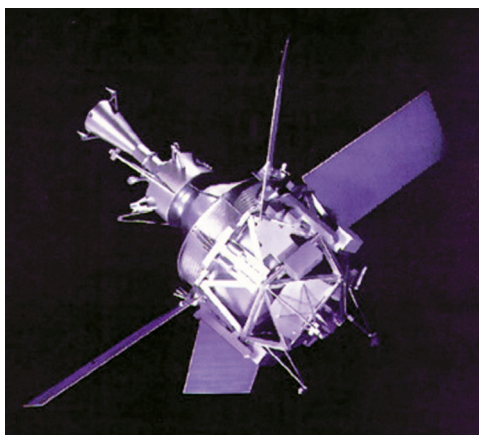
The GP-B spacecraft will hold a dewar for liquid helium (the main body of the spacecraft shown), with the probe containing the gyroscopes inside. At the upper left is the telescope assembly that will establish a reference direction to a distant star against which to measure the gyroscope precessions. Also visible are solar panels for power. (Courtesy of Lockheed Martin Missiles & Space.)

physics in 1986, the project has made substantial technological progress and secured the support of NASA through launch and data analysis. At present, the spacecraft is built, construction and installation of the flight experimental hardware is complete, and the project is proceeding toward a proposed launch in 2000. For the purpose of this report, Gravity Probe B is treated as part of the committed, ongoing program in gravitational physics.

Re-emergence of Scalar-Tensor Theory

An important theoretical development of the last decade was the re-emergence of scalar-tensor gravitational theory as an alternative to general relativity. Scalar-tensor theories augment the standard tensor gravitational interaction of general relativity with a scalar field. Their exemplar, the Brans-Dicke theory, fell into disfavor during the 1970s, as experiments strongly pointed toward general relativity. But a new class of more general scalar-tensor theories was a product of developments in particle theory, specifically string theory (see Section V of this chapter), where the presence of scalar fields (spin-zero particles) in addition to the tensor (spin-two) metric is required. In many string-inspired models, the "low-energy" limit (appropriate for everyday gravity, astrophysics, and most of cosmology) is necessarily a scalar-tensor theory. In many such models, standard cosmological evolution drives the parameters of the scalar-tensor theory toward, but not quite to, those of general relativity. This permits such theories to agree closely with general relativity in weak-field situations, such as the solar system, while diverging strongly from it in early-universe and strong-field situations. The differences for weak fields, while perhaps no more than a part in 10^5, may be sought with experiments, such as Gravity Probe B's proposed measurement of the geodetic precession, or improved tests of the equivalence principle. This conjunction between particle theory and gravity in scalar-tensor gravity is an opportunity for further theoretical study and future experimental tests.

Applications of Relativistic Gravity

A surprising development since the decadal survey in 1986 is the extent to which a number of general relativistic effects have taken on practical significance. In the Global Positioning System (GPS), a high-precision navigational and time-transfer system that is the basis for a multibillion-dollar commercial industry, the relativity of time plays a crucial role (see Box 2.1 in Chapter 2). The effect of gravity on the rates of clocks (the gravitational redshift), together with the special relativistic time dilation due to the clocks' motion, results in a difference between the rates of the atomic clocks on the 24 orbiting satellites of the GPS and those of clocks on the ground. If the accumulated difference in time, almost 40 milliseconds per day, were not taken into account, the GPS could not

meet its stated navigational accuracy of 15 meters, or its time transfer accuracy of 100 nanoseconds.

Relativistic corrections are routinely incorporated into computer modeling of the orbits of interplanetary spacecraft, and into analyses of Earth's gravity field from precisely tracked terrestrial orbiters.

Another relativistic effect that has been put to use is the deflection of light, in this case as an astronomical tool. When light from a distant object passes by or through a galaxy or cluster of galaxies, the gravitational light deflection can produce multiple images of the distant object, arc-like distorted images, or complete ring-like images. Careful study of these images can provide a map of the mass distribution in the lensing system, and it plays a role in the search for dark matter and the measurement of cosmological parameters. An analogous phenomenon is microlensing, in which a foreground mass passes in front of a distant star, augmenting its intensity momentarily through gravitational focusing. This effect has also been used extensively in searches for dark objects in and around our own Galaxy. (See Section III of this chapter.)

Opportunities

The opportunities described here are necessarily of an open-ended, exploratory character. Scientists seek evidence that current theories might be breaking down, revealing clues to new physics beyond their limits of applicability. It is impossible to predict with any confidence what sensitivity will be required in these experiments. But it is not unreasonable to expect that one or more of these lines of research will reveal evidence that new theories must replace or augment the current ones. If this happens, the impact of such a discovery will likely be revolutionary.

Tests of the Equivalence Principle

A key opportunity for the coming decade in experimental gravity will be the testing of fundamental aspects of the gravitational interaction in search of, or to constrain, new physical interactions. The interactions inspired by the particle physics-gravity connection generally produce violations of the UFF, violations of the gravitational inverse square law, and spin-dependent gravitational interactions. Existing laboratory tests of the UFF are still far from reaching the ultimate limitations of the technique and can reasonably be expected to improve by a factor of at least 50. A space test of the UFF could achieve a millionfold improvement over current levels, reaching a part in 10^{18}, by monitoring the relative motion inside a satellite of two freely moving bodies of different materials. During a year-long mission, this amounts to thousands of repetitions of Galileo's famous experiment. The principal advantage of the space environment is the

substantial reduction of the effects of vibrations that plague ground-based experiments.

Any additional "gravitational" quanta that are spinless (such as the dilaton and moduli of string theory) would violate the gravitational inverse square law at some length scale. String theory and other models that endeavor to unify the interactions are not yet at a stage where they can make a definite prediction for the size of the new effects. But this situation is likely to improve in coming years, and violations at some scale seem inevitable. For example, there is an intriguing recent suggestion that large violations of the inverse square law could occur at the micron scale. Laboratory and astronomical tests made prior to the last decade explored length scales down to 1 millimeter. Attention is now focusing on even shorter length scales. In this regime, the gravitational forces are necessarily weak compared to electromagnetic interactions between neutral objects, so that new techniques are required.

It is useful to point out that many of the experimental improvements needed to perform these delicate measurements, such as vibration isolation and sensitive gravity gradiometers, may have payoffs in other areas, such as gravitational wave detection and geophysics. At the same time, this effort will benefit from theoretical progress, which could lead, for example, to more precise predictions from string theory for the strength and range of violations of the UFF or the inverse square law.

Search for Frame Dragging

Another opportunity that can be realized in the next decade is the detection and study of the effects of frame dragging. The ongoing Gravity Probe B project (see above) is the only experiment capable of a high-precision test of the effect. Current plans call for launch in 2000 and completion of the mission by 2002. Another way to detect frame dragging is to measure the rotation of the orbital plane of a satellite in Earth orbit. This could be accomplished using laser-ranged geodynamic satellites called LAGEOS, albeit at a substantially lower level of accuracy than the goal of Gravity Probe B. Data from the two LAGEOS satellites currently in orbit are being analyzed by various groups, with the goal of understanding and reducing sources of systematic error. Ideally, the launch of a third LAGEOS satellite with orbital tilt having a prescribed relationship to that of one of the other LAGEOS satellites could provide a "cleaner" measurement, free of some of the error sources.

Solar System Tests of Gravitation

General relativity is a fundamentally nonlinear theory (gravity begets gravity), yet the only experiments in the solar system that test that aspect of the theory are the perihelion advance of Mercury and lunar laser ranging, now known to

agree with general relativity to parts in a thousand. Another test of nonlinearities could be provided by a space experiment in which atomic clocks travel close to the Sun. In such an experiment, the shift in frequency between satellite and Earth clocks, or between satellite clocks of different physical structure, contains not only the first-order, or linear contribution, which could be tested to a part in a billion, but also a nonlinear term, which could be tested to a part in a thousand.

Lunar laser ranging has been one of the most cost-effective and scientifically productive projects arising from the space program. It has yielded new information on the orbit and rotational motion of the Moon and on crustal motions on Earth. It has also verified the equivalence principle for Earth and the Moon and has helped set bounds on a temporal variation of the Newtonian gravitational constant. Improved lasers, better modeling of the lunar motion, and the continued accumulation of data will provide further improvements in all of these areas. For example, the measurement of the equality of acceleration could be improved by an order of magnitude.

Binary and Millisecond Pulsars

To date, only about 1000 out of a predicted 10^5 pulsars in our Galaxy have been discovered; of these, 100 could be in binary systems. While continued searches for and observations of millisecond and binary pulsars will be driven by independent astrophysical considerations, they will provide opportunities for further tests of relativity in the radiative and strong-field regimes. For example, the fortuitous discovery of binary pulsars of the right characteristics, such as systems containing both a pulsar and a black hole, could result in a 10-fold improvement in accuracy in the test of gravitational radiation damping, provide a high-precision measurement of a companion black hole mass, detect the precession of the spin of a neutron star, contribute to a determination of the distribution of neutron star masses, and help sharpen the event rate of inspiraling and coalescing neutron star binaries.

The Newtonian Gravitational Constant

The Newtonian constant G, which governs the strength of all gravitational interactions, was historically the first "fundamental constant" in physics. Yet today, with an official uncertainty of about a part in 10^4, it is the least precisely determined of any of these constants. Recently, measurements of G at several laboratories have cast doubt on the accepted value of G and especially on its uncertainty. The next decade will see the completion of G measurements using a wide variety of techniques and devices, such as torsion balances, fountains of ultracold atoms, or gravimeters that see a modulated field, possibly reaching a level of a part in 10^6. Consistent results from several groups will be needed to give confidence that the systematic errors in the measurements are understood.

Improved bounds on any temporal variation of *G* of cosmic origin could help constrain alternative gravitational and cosmological models, such as those arising from string theory.

V. UNIFYING GRAVITY AND QUANTUM THEORY

Key Questions

There have been three major crises in theoretical physics in this century. In each case, two well-established theories were found to be incompatible, either because they were based on contradictory assumptions about the workings of the physical world, or because they led to physically untenable conclusions. The first was the conflict between Newtonian gravity and special relativity, which was resolved by Einstein's theory of general relativity. The second arose from tension between thermodynamics and electromagnetism, which led to the development of quantum theory. In both of these cases, the crisis was resolved not by a small modification of one of the theories, but rather by an entirely new fundamental theory that introduced a new framework and made some old concepts obsolete. Over time, both of these theories passed stringent experimental tests and now form the cornerstones of modern physics. Unfortunately, however, general relativity and quantum theory are themselves mutually incompatible, presenting physicists with a third crisis. It is likely that the resolution of this crises will be as profound as the previous two, yielding a major change in our view of nature.

The resulting theory, quantum gravity, is expected to be crucial for understanding the strongest gravitational fields in the universe. It is possible to make a rough estimate of when the effects of quantum gravity should become important. This is because, as mentioned in Chapter 2, there is a unique combination of the fundamental constants of general relativity and quantum theory which has dimensions of length, $\ell_p = (\hbar G/c^3)^{1/2} \approx 10^{-33}$ cm, called the Planck length. When the gravitational field is so strong that space is curved on this scale, then quantum gravity is indispensable. Such extreme conditions were present in the early universe. Our current cosmological models can in principle be extrapolated back to $t = \ell_p / c \approx 10^{-43}$ seconds after the big bang, but then they completely break down. Quantum gravity should provide a description of the first moments after the big bang, and perhaps of the big bang itself.

In certain situations, effects of quantum gravity can be important even when the gravitational fields are significantly weaker than the above estimate. In particular this is the case around black holes. It turns out that, because of the unusual properties of space and time near a black hole, when quantum effects are included black holes are not really black; they radiate via a quantum tunneling process, losing their mass and becoming hotter. This radiation can be quite significant. For example, if black holes with mass of the order of 10^{15} grams

were formed in the early universe, they would appear today to be white hot and would explode.

Some of the key questions in quantum gravity are the following:

• How can the principles of general relativity and quantum theory be unified into one consistent framework?
• What are the quantum properties of black holes?
• What is the nature of space and time at the smallest possible scales?
• Is this smallest-scale structure responsible for curing the short-distance infinities of quantum theories of other fundamental forces?
• What was the universe like 10^{-43} seconds after the big bang?

As the CGP describes below, there has been enormous progress over the past decade in trying to answer these questions, especially the first three. This is due partly to a new formulation of the problem, and partly to the influx of ideas from high-energy physics. As the search for a unified theory of all forces expanded to encompass gravity, particle physicists were naturally led to seek a quantum theory of gravity. The next decade promises to be a very exciting one, in which ideas from different approaches may fuse together, providing deeper insights and perhaps complete answers to these fundamental questions about nature.

Achievements

The usual formulations of quantum theory require a fixed notion of time since, e.g., quantum states are specified at an instant of time. But in quantum gravity, physicists expect the spacetime geometry to fluctuate. This raises deep conceptual problems, especially in the cosmological context where there are no external observers with clocks. To address such problems, a new route was developed in the 1990s; quantum theory was *generalized* to accommodate quantum spacetime geometry. Usual quantum mechanics is recovered in those epochs and situations in which the spacetime geometry is approximately classical. This conceptual advance resulted from a fruitful interchange of ideas between experts working on quantum gravity, quantum theory of closed systems, and foundations of quantum mechanics.

In spite of this progress in constructing a framework, however, the task of actually constructing a quantum theory of gravity remains formidable. Straightforward quantized general relativity is "perturbatively non-renormalizable"—it yields infinite answers to physical questions. Therefore, it is natural to start with approximation schemes. Perhaps the simplest among them is quantum field theory in curved spacetimes, where the gravitational field is treated as a classical, passive entity and analyzes the effects of the curved spacetime geometry on quantized matter. At first, this appears to be an oversimplification of the problem. However, this approach led to some key insights in the mid-1970s. The

most striking among them is the Hawking effect: black holes radiate as though they are blackbodies. This discovery brought together general relativity, quantum field theory, and thermodynamics and provided powerful hints for quantum gravity. In particular, it became physically meaningful to assign thermodynamic parameters such as temperature and entropy to black holes in terms of their geometric properties. This posed a concrete challenge to any candidate quantum theory of gravity: Explain the origin of these thermodynamic properties in terms of microscopic degrees of freedom as is done for ideal gases and other everyday systems. As described below, there has been considerable recent progress on this challenge.

Over the past decade, quantum field theory in curved spacetimes has evolved considerably, especially through the development of powerful algebraic methods, and has now become a mature branch of mathematical physics. On the physical side, to gain insight into the role of very high frequencies in the Hawking effect, quantum physics around "dumb holes"—the acoustic analogs of black holes—has been studied. These model systems also exhibit Hawking radiation (see Box 3.5) and the radiation spectrum turns out to be quite robust, surprisingly insensitive to changes in the assumed properties of the extreme high frequency modes. Substantial progress occurred also in another approximation method, that of effective field theories. Here, theorists do treat the gravitational field quantum mechanically, but they focus only on the implications of these quantum effects on length scales large compared to ℓ_p. Thus, although quantum general relativity is "perturbatively non-renormalizable," theorists can nonetheless extract from it meaningful information to describe physics in the situation where the curvature of spacetime is small compared to that at the Planck scale. This is an interesting development, especially because in the 1970s, such non-renormalizable quantum theories were widely regarded as being devoid of physical content.

A number of approaches have been developed to go beyond such approximations. Notable among them are Euclidean quantum gravity, the dynamical triangulation method, Regge calculus, asymptotic quantization, reformulation of general relativity as a dynamical theory of null hyper-surfaces, twistor theory, and non-commutative geometry. However, in terms of providing answers to the key questions, two directions stand out: string theory and quantum theory of geometry. In both of these approaches, considerable progress has already been made on the first three key questions listed above.

String Theory

During the past decade, string theory has emerged as a leading candidate for a quantum theory of gravity. In addition, this theory appears to achieve another long-standing goal of theoretical physics: It may provide a unified theory of all known forces and particles. The starting point is remarkably simple. One assumes that elementary particles are not point-like, but rather actually different

---

BOX 3.5 Black Holes and Quantum Physics

In the early 1970s, the properties of black holes were studied and formulated in terms of the laws of black hole mechanics. It was noticed that there was a striking similarity between these laws and the ordinary laws of thermodynamics. The analog of the temperature T was a multiple of the surface gravity of the black hole κ, which is similar to g on Earth's surface. The analog of the entropy S was a multiple of the area of the event horizon A. More precisely, there were the following correspondences:

Black Hole Mechanics	Thermodynamics
κ is constant.	T is constant.
$\partial M = \kappa \partial A/8\pi G.$	$\partial E = T\partial S.$
$\partial A \geq 0.$	$\partial S \geq 0.$

It was believed at the time that the above relationships were only an analogy, since the defining property of black holes was that nothing can escape their gravitational pull. In particular, they did not emit the radiation characteristic of an object with non-zero temperature. However, a few years later it was shown that when quantum effects are included, this analogy becomes exact. Because of an analog of the tunneling process in elementary quantum mechanics called the Hawking effect, particles *can* escape from black holes. Thus, black holes are not really black. They radiate just like hot objects with temperature $T_{bh} = (\hbar\kappa/2\pi c)$. This implies that black hole entropy is given by $S_{bh} = A/4\ell_p^2$, where ℓ_p is the Planck length. Therefore, in a fundamental statistical mechanical description, a macroscopic black hole of area A should have $e^{S_{bh}}$ microstates—a huge number.

It has taken 25 years to obtain a fundamental description of these microstates. In the past few years this fundamental description has finally been achieved in both of the main approaches to quantum gravity. However, important issues still remain, and there should be further exciting developments in this area over the next decade.

---

excitations of a one-dimensional extended object—the string. When an ordinary violin string is plucked, it vibrates at certain characteristic frequencies producing the usual notes. Fundamental strings are much smaller (of order ℓ_p in size), but when they are excited they also vibrate at certain frequencies. Different modes of vibration are seen as electrons, quarks, photons, and so on. This provides a strikingly simple unified picture.

The basic interaction between strings is through a simple splitting and joining process. Remarkably, the description of this process automatically incorporates the known interactions between the elementary particles. The relation between string theory and general relativity can be seen in two ways. First, one mode of the string describes a graviton (a small fluctuation of the gravitational field), and the classical scattering of strings in this mode reproduces the

perturbative expansion of general relativity. More importantly, when strings in curved spacetime are studied, a consistency condition on the spacetime geometry emerges which is Einstein's equation of general relativity, augmented by corrections that are only important at the Planck scale. In fact, if general relativity were not already known, string theory would predict it as the description of gravity for distances much larger than the size of the string.

Until recently, quantum effects in string theory were mostly discussed in the context of perturbation theory. A background spacetime geometry is assumed (satisfying the above consistency condition) and small fluctuations about it are quantized. This is not sufficient to answer the important questions about the big bang or what happens deep inside a black hole. These issues appear to require a complete non-perturbative formulation of string theory which is not yet available. However, in the past few years, a number of non-perturbative facts about the theory have nonetheless been found. This is possible largely because string theory incorporates supersymmetry, a powerful symmetry first discussed in the 1970s. Supersymmetry implies that certain results computed perturbatively are, in fact, exact.

String theory has had a number of achievements over the past decade. Here the CGP focuses on those achievements that relate directly to gravity and the structure of space and time. The apparently obvious three-dimensional nature of space need not be correct. It is possible that we experience three dimensions because the extra dimensions are curled up into a very small ball. Our measurements of space so far might simply be too crude to detect such extra dimensions. In string theory, space and time are no longer fundamental, but instead are derived concepts. One early result is that space must have more than three dimensions. (The perturbative formulation of string theory predicts nine dimensions, but recent non-perturbative arguments suggest that in the full theory, the number is ten.)

T-duality. Suppose one direction in space is curled up into a circle of radius R. In addition to the usual string states, there are now extra states corresponding to strings winding around this circle. The net effect of these extra states is that the spectrum of the string is exactly the same as if the circle had radius $1/R$. Furthermore, the interactions between string states are also invariant under changing the radius of a circle from R to $1/R$. This means that very small circles are indistinguishable from large circles in string theory.

Singularities. As discussed elsewhere in this report, general relativity predicts the existence of places in the universe where the spacetime curvature is infinite. General relativity breaks down at these "singularities." Since string theory modifies general relativity even classically, it is important to know whether singularities exist in string theory as well. It has been shown that several spacetimes that are singular in general relativity are completely nonsingular when embedded

in string theory. However, it has also been shown that there are some singular solutions even in string theory. These examples are not very physical, and so it is not yet known whether these singularities are likely to arise in nature.

Topology Change. The topology of space is a measure of how the space is connected—e.g., whether it has holes like the surface of a donut. A long-standing question is whether the topology of space in our universe can change over time. According to general relativity, the answer is no: Topology change always results in singularities. In string theory the situation is different. It has been shown that the topology of space can change in a way that is non-singular in string theory.

Quantum Properties of Black Holes. The most important achievement has undoubtedly been the successful description of black hole entropy in terms of quantum string states. As discussed above, this has been a challenge for more than 20 years. For certain large black holes with electric charge near the maximum allowed value, the number of string states with the same charge and mass turns out to be precisely the number predicted from black hole thermodynamics. Even more importantly, the interactions between these states turn out to precisely reproduce the Hawking spectrum of radiation. This is a remarkable achievement. For more general black holes, theorists can identify a class of string states associated with a black hole which scale in the expected way with the mass and charge, but numerical coefficients in the entropy formulas have not yet been checked.

Quantum Theory of Geometry

In general relativity, spacetime geometry is a dynamical entity that interacts with matter and has degrees of freedom of its own. Therefore, to unify the principles of general relativity and quantum theory, it is natural to take this physical role of geometry to be fundamental and probe its quantum nature from first principles. Over the past decade, a detailed theory has been developed starting from this viewpoint which in turn has provided some key insights on the nature of quantum gravity effects.

This approach is "non-perturbative" in the sense that a classical spacetime is not the starting point to which quantum fluctuations to its geometry are *then* added. There is *no* background spacetime; everything, including geometry, is dynamical and quantum mechanical. Indeed, the strategy is just the opposite of that followed in perturbative treatments: Rather than starting with quantum matter on classical spacetimes, one first quantizes geometry and then incorporates matter. This procedure is motivated by two considerations. The first comes from general relativity in which some of the simplest and yet most interesting physical systems—black holes and gravitational waves—consist of "pure geometry." The second comes from quantum theory where the occurrence of infinities

at short distances suggests that it may be physically incorrect to quantize matter assuming that spacetime can be regarded as a smooth continuum at arbitrarily small scales.

The detailed implementation of these ideas requires a new mathematical and conceptual framework, since the standard methods used in quantum theories of non-gravitational forces are tied to the availability of a background spacetime that is now absent. An important recent advance was the systematic construction of the appropriate substitutes. The resulting mathematical framework is now sufficiently rigorous to ensure that there are no hidden infinities or other internal inconsistencies.

The key results obtained from this framework can be summarized as follows.

Fundamental Discreteness. Over the past 5 years, this approach has led to a detailed quantum theory of geometry. The framework shares several basic concepts with gauge theories, thereby bringing gravity closer to other fundamental forces. In particular, the fundamental excitations of geometry are coded in the gravitational Wilson loops. They are one-dimensional so that quantum geometry resembles a polymer. However, when densely packed in appropriate configurations, these excitations can approximate the three-dimensional spatial continuum. Quantum analogs of observable geometrical quantities such as areas of surfaces and volumes of regions—called geometric operators—are well-defined. They have the striking property that their values are quantized, that is, can change only in discrete steps. Thus, at the Planck scale, the continuum picture breaks down and geometry becomes "polymer-like."

Several properties of the geometric operators have been worked out. For instance, the allowed discrete values of area crowd rapidly as area increases; the difference between them, called the level spacing, goes to zero exponentially quickly, making the continuum picture an excellent approximation in laboratory physics. However, since the Planck length ℓ_p is so small, can such details ever bear on the macroscopic world? What if, for example, the level spacing were uniform, like that in a harmonic oscillator, with steps of the order of ℓ_p? Could such alternatives be physically distinguished? Surprisingly, using quantum field theory in black hole spacetimes, one can. While the actual level spacing of area is consistent with the blackbody spectrum of the Hawking effect, the uniform level spacing is not. Thus, there are checks on predictions.

Black Hole Thermodynamics. Since a black hole in general relativity is "pure geometry," it is natural to use quantum geometry to unravel its microscopic degrees of freedom. Recently, this task was carried out for nonrotating black holes, possibly with charges. For large black holes, the number of microstates grows exponentially with area, showing that the entropy is proportional to area. From this perspective, the mechanism underlying black hole evaporation is strikingly simple: Quanta of area are converted to quanta of matter. This ongoing

work provides a quantum mechanical manifestation of the idea that geometry is a physical entity.

Quantum Dynamics. Quantum geometry provides a mathematical language to formulate a wide variety of quantum gravity theories, just as differential geometry does in the case of classical gravity. However, so far quantum dynamics has been explored only in general relativity, possibly coupled to matter, and supergravity, the extension of general relativity which incorporates supersymmetry. The central question is, Can these quantum theories admit an exact, mathematically consistent formulation even though they are "perturbatively nonrenormalizable?" In two (rather than three) space dimensions, the answer is in the affirmative even with certain types of matter. Although this lower-dimensional theory is not of direct physical interest, it faces most of the conceptual difficulties of the three-dimensional theory, and the application of perturbative methods had led to a general belief that a consistent quantum theory would not exist. Not only does a satisfactory theory exist but the non-perturbative formulation also provides useful hints for higher-dimensional theories.

Over the last 3 years, there has also been considerable work in four dimensions which has provided an example of how quantum Einstein equations can be formulated rigorously. This is an interesting development in mathematical physics. However, it is not yet clear whether this formulation can successfully answer those physical questions that are motivated by semi-classical considerations. Another approach to quantum dynamics leads to an unexpected interplay with a branch of topology, namely, the theory of knots. In particular, some of the well-known "knot invariants" automatically solve quantum Einstein equations, and there are indications of deeper relations between quantum general relativity and knot theory.

Opportunities

The coming decade is likely to see substantial further progress in quantum gravity. With an eye toward the key questions listed above, the CGP briefly describes some of the opportunities that await us.

The pace of progress in string theory has been extremely rapid during the past few years. Indeed, within the past year, a proposal for a non-perturbative formulation of the theory has been made, which is applicable when the cosmological constant is negative. While this is not believed to be the case in nature, this formulation can still be used as a model to study quantum gravitational processes such as the evaporation of black holes. This proposal incorporates a novel "holographic" view of space and time, in which our usual notions of locality and causality hold only approximately. Much effort will be devoted in coming years to establish in detail that this proposal is correct and to extend it in such

a way that the cosmological constant need not be specified a priori. If these steps can be completed, we may finally have a workable quantum theory of gravity.

Significant advances are also expected in the approach based on quantum geometry. In this framework, quantum dynamics was initially discussed using Hamiltonian methods. Spacetime formulations of the required theory are now being pursued vigorously in which quantum geometry serves to unify results from apparently three distinct areas of research, pursued independently by relativists, quantum field theorists, and mathematicians. These methods provide a new avenue to formulate and discuss quantum Einstein equations and are better suited for semi-classical considerations. Physical ramifications of the quantum nature of geometry will also be explored further. In particular, it is likely that quantum field theories on quantum geometries will be studied. Since the fundamental geometric excitations are one-dimensional, the effective spacetime dimension is now reduced. The key question then is whether this feature will free quantum theories of non-gravitational interactions from the usual short-distance infinities. If so, the old and cherished hope that quantum gravity may cure quantum field theories will be realized.

So far, string theory has been developed largely by high-energy theorists and quantum geometry by relativists. This is reflected in the choice of issues that are emphasized in the two approaches. However, there are tantalizing similarities such as the importance of one-dimensional objects. Furthermore, as the results on black hole thermodynamics indicate, both approaches are now addressing the same physical problems. Their strengths are complementary. One enables quantum physics to be done without a background spacetime but provides no guidance on how various physical fields couple to gravity and to one another. The other has an in-built powerful principle that dictates all couplings but has yet to completely free itself from reliance on a background geometry. Much progress would occur if there were more interaction between the two communities.

It is clear that recent results on the quantum properties of black holes will be extended using both string theory and quantum geometry. In this area, there are several exciting challenges. The two approaches have led to rather different physical pictures of a quantum black hole, one based on extended objects in higher-dimensional spacetimes, and the other, on the polymer-like excitations of geometry of ordinary space. Are these two pictures "complementary" in a suitable sense? More generally, what is the relation between them? Another challenge is to derive the laws of black hole thermodynamics from quantum gravity, in full generality, allowing for departures from thermal equilibrium. An even more important open question is whether information thrown into a black hole is lost forever, or is ultimately recovered in the evaporation process. If it is indeed lost forever, as suggested by the original semi-classical calculations in the 1970s, then some of the basic principles of quantum theory would have to be modified. However, the recent string calculations indicate that information is not lost. It is

possible that this long-standing "black hole information puzzle" will be resolved in the near future.

Considerable progress can also be expected in applying these developing theories of quantum gravity to cosmology. Indeed it could be argued that while most effects of quantum gravity are not observable, the effects of the quantum fluctuations in geometry and matter near the big bang are all around us. We see them in anisotropies in the cosmic background radiation and in the large-scale distributions of galaxies. The objective of quantum cosmology is to understand the earliest moments after the big bang. Quantum gravity is central to this task, and cosmology is one of quantum gravity's most important applications. (These issues are further discussed in Section III of this chapter.)

It is quite possible that qualitatively new and unexpected effects will be discovered in the coming years, the quantum gravity analogs of $E = mc^2$ of special relativity. For example, it was recently found that in the two- and three-dimensional exactly soluble models, unexpectedly large quantum fluctuations can arise in the spacetime geometry because the coupling between general relativistic gravity and matter can magnify small quantum uncertainties in matter sources of gravitation into huge uncertainties in the gravitational field. In the coming years, these results are likely to be extended to four dimensions and may then have experimental consequences. Similarly, there are directions in which the current ideas in string theory could be confronted with experiments. For example, discovery of supersymmetry in particle accelerators would lend support to an important ingredient of string theory. A second example arises from the fact that one mode of the string, the dilaton, has interactions like gravity but couples to matter in a different way. (See Section IV of this chapter.) This might produce violations of the equivalence principle at a detectable level.

More direct tests may also be possible in spite of the fact that the energy scale of quantum gravity is very high, 10^{19} GeV. In the 1980s, for example, experiments were performed to directly test the predictions of grand unified theories on proton decay. The processes responsible for this phenomenon are only a few orders of magnitude below the quantum gravity scale. Yet, it was possible to test these theories without having to accelerate particles to such high energies; the experiments involved confining a very large number of protons and waiting sufficiently long to see if any of them decayed. In the same spirit, attempts have recently been made to put limits on certain quantum gravity effects using observations of TeV gamma-ray flares. The idea is that the tiny effects on the propagation of gamma-rays due to the Planck-scale fluctuations in the spacetime geometry can accumulate during their long flight over cosmological distances and lead to an observable dispersion. It is likely that these ideas will be refined over the next decade and enable researchers to experimentally distinguish between possible quantum gravity scenarios.

Appendixes

*A*ppendix A

Activities of the Committee on Gravitational Physics

The Committee on Gravitational Physics (CGP) held its first meeting on October 7-9, 1997, at the National Research Council's facility in Washington, D.C. The first part of the meeting was devoted to gathering information on gravitational physics projects and programs. The committee discussed letters and e-mail that it had received in response to a call for comments. The committee then heard presentations on the following topics:

• NASA gravitational physics activities. Alan Bunner, Science Program Director, Structure and Evolution of the Universe (SEU), and Hashima Hasan, Discipline Scientist, Ultraviolet, Visible, and Gravitational Astrophysics.

• Astrophysics and gravitational physics and the recommendations of the newly released Structure and Evolution of the Universe Science Roadmap prepared for NASA's Office of Space Science by an advisory committee chaired by Roger Blandford, Caltech.

• The National Science Foundation's (NSF's) gravitational physics program. David Berley, NSF Program Manager for the Laser Interferometer Gravitational-Wave Observatory (LIGO).

• The LIGO project. Barry C. Barish, Principal Investigator.

• The Gravity Probe B relativity mission. C.W. Francis Everitt, Stanford University, Principal Investigator.

• The Satellite Test of the Equivalence Principle (STEP) and Mini-STEP missions. C.W. Francis Everitt, Stanford University, Principal Investigator, and Paul Worden, Stanford University, Co-Investigator.

101

• The Laser Interferometer Space Antenna (LISA). Peter Bender, JILA, Principal Investigator.

• Resonant mass gravitational wave detectors. William Hamilton, Louisiana State University.

• The Microwave Anisotropy Probe (MAP). Gary Hinshaw, Astrophysicist, Goddard Space Flight Center.

In the second part of the meeting, held in closed session, the committee identified key questions in the field. The committee concluded by drafting an outline for its report.

The CGP held its second meeting at the National Research Council's facility in Washington, D.C., on February 20-22, 1998. The first day of the meeting was conducted entirely in open session and began with greetings and brief remarks from the committee chair, Prof. James Hartle. This was followed by a brief explanation by Board on Physics and Astronomy (BPA) Director Donald C. Shapero regarding the National Research Council's response to the new law concerning amendments to the Federal Advisory Committee Act. The CGP heard the following presentations:

• Richard Isaacson, Program Manager for gravitational physics at NSF, discussed the opportunities in gravitational physics experiment, computation, theory, international collaboration, university training, and the LIGO project.

• P.K. Williams, Senior Program Officer, Office of Energy Research, Department of Energy (DOE), explained the workings of the joint DOE-NASA-NSF Scientific Advisory Group for Non-Accelerator Physics (SAGENAP) and the DOE activities in non-accelerator physics connected to cosmology and gravitation.

• Committee member Eric Adelberger presented a summary of the recent laboratory experiments to measure to high accuracy Newton's gravitational constant, G, and Earth's gravitational acceleration, g.

• Committee members Clifford Will and Peter Michelson presented current and proposed gravitational experiments based in space, such as the lunar laser ranging experiment, the Gravity Probe B mission, OMEGA, the "Galileo Galilei" equivalence principle mission, the Satellite Test of the Equivalence Principle (STEP), the Laser Interferometer Space Antenna (LISA), and others.

• Committee members Ramesh Narayan, David Spergel, and Joseph Taylor led an astrophysical discussion of the estimated number of sources of gravitational waves that would be detectable by LIGO.

The next 2 days' sessions were closed. They began with a review of the previous day's items, including a continuation of the estimate for LIGO source counts. The discussion of draft chapters of the report and revisions to drafts was followed by a consideration of the goals and opportunities of the field and a

preliminary formulation of recommendations. The committee discussed a draft section on gravitational physics that was sent to the Board on Physics and Astronomy for inclusion in the forthcoming Overview report of the physics survey *Physics in a New Era* (to be published by the National Academy Press in 2000). The CGP also discussed themes for a research briefing on gravitational physics and the schedule for completing the draft report.

The CGP requested input from the community of gravitational physicists in a number of ways:

- A description of the CGP's charge and activities was published in the newsletter of the American Physical Society's (APS's) Topical Group on Gravitation. This newsletter is available to gravitational physicists worldwide, both in print form and on the Los Alamos e-print server.
- A similar notice was posted on the e-mail service maintained by Queen Mary College in London which reaches hundreds of gravitational physicists around the world.
- Requests for input were made through standard announcement services of the Division of Particles and Fields of the APS, the Precision Measurements Topical Group of the APS, and the American Astronomical Society.
- The committee chair, J. Hartle, made presentations and solicited input at two meetings of gravitational physicists: the 1998 Pacific Coast Gravity Meeting in Eugene, Oregon, and the April 1998 meeting of the American Physical Society in Columbus, Ohio.

In total the CGP received written responses from approximately 20 scientists. A great many of these were thoughtful and helpful. All electronic and written input was distributed to the members of the CGP and duly considered in its deliberations.

$\mathcal{A}$ppendix B

Glossary

accretion: the process by which gas flows around and onto a compact gravitating object.

advection-dominated accretion flow (ADAF): an accretion flow in which most of the energy released by viscous action is carried with the gas to the center rather than being radiated.

anisotropic: not isotropic (q.v.), that is, not the same in all directions. An anisotropy is a measure of the difference between directions. The cosmic background radiation (q.v.) is approximately isotropic in its temperature in different directions on the sky, but has minute anisotropies in temperature.

bending of light: deflection of light (or other electromagnetic radiation) from a straight-line path as it falls in a gravitational field.

BeppoSAX satellite: BeppoSAX is an Italian/Dutch mission that can accurately determine the location of x-ray and gamma-ray bursts (for information online, see <http://www.sdc.asi.it>).

big bang: the initial instant of the universe characterized in general relativity by arbitrarily high density, temperature, and curvature.

Binary Black Hole Alliance: a multidisciplinary collaboration among relativity theorists and computer scientists at eight institutions to develop algorithms and software for solving Einstein's equations on supercomputers. The alliance focused on the coalescence of two black holes in binary orbit about each other. It was funded by the NSF from 1993 to 1998 as one of its Grand Challenge projects in computational science. (Cf. Neutron Star Grand Challenge.)

104

binary pulsar: a radio pulsar (q.v.) that is gravitationally bound to a companion star and orbits it. The signals from such a system can be used to test some aspects of general relativity to great precision.

blackbody spectrum: Every body in thermal equilibrium emits the same kind of light characterized only by the body's temperature. The distribution of intensity with wavelength is called the blackbody spectrum, and an object emitting this characteristic spectrum is called a blackbody.

black hole: a region of space where the gravitational pull is so strong that classically, nothing can escape. The boundary of this region is called the black hole's event horizon (q.v.). Black holes can form when a massive star undergoes gravitational collapse (q.v.).

black hole evaporation: the quantum mechanical process whereby a black hole losses mass and becomes smaller due to Hawking radiation (q.v.).

Chandra X-ray satellite (formerly Advanced X-ray Astrophysics Facility): a NASA satellite observatory that was launched on the space shuttle on July 23, 1999. It will image the x-ray sky over the energy range from 0.1 to 10 keV with resolution similar to that of the Hubble Space Telescope (for information online, see <http://chandra.harvard.edu/index.html>).

classical: a general term meaning non-quantum mechanical. For example, general relativity is a classical theory with deterministic equations of motion for the geometry of spacetime. In a non-classical quantum theory of gravity only probabilities for spacetime geometries would be predicted.

closed universe: a finite-volume universe resulting from the gravitational pull of a high density of matter. It may be visualized as the three-dimensional analog of the surface of a sphere; if one travels straight in any direction, one eventually returns to the same place.

COBE: the Cosmic Background Explorer satellite for seeing the details of the light from the big bang (for information online, see <http://www.gsfc.nasa.gov/astro/cobe/cobe_home.html>).

cold dark matter: another name for subatomic particles that interact weakly with ordinary matter and radiation. These subatomic particles can cluster gravitationally and form galaxies. Weakly interacting massive particles are a form of cold dark matter.

Compton Gamma Ray Observatory (CGRO): Named after American Nobel laureate Arthur Holly Compton, the Gamma Ray Observatory was launched by NASA in 1991 to study the spectrum, location, and nature of gamma-rays and gamma-ray bursts from astronomical sources. For information online, see <http://antwrp.gsfc.nasa.gov/cossc/cossc.html>.

cosmic background radiation (CMB): the residual light from the big bang. While nearly uniform, there are tiny variations in its temperature due to fluctuations in the density of the early universe. These tiny density fluctuations grew to form today's galaxies.

cosmic censorship conjecture: the conjecture that the inevitable singularities (q.v) formed in a physically realistic gravitational collapse are formed inside black holes, hidden from the view of a distant observer.

cosmic string: a string-like defect in matter fields that is the relic of a phase transition in the early universe. Cosmic strings are one possible seed for the formation of galaxies.

cosmological constant: the energy density associated with the vacuum (empty space). Recent astronomical observations suggest that there is a net energy associated with the vacuum. If there is a positive vacuum energy, then the expansion of the universe will eventually accelerate and our descendants will find themselves in a nearly empty universe.

cosmology: the study of the origin, evolution, fate, and physical properties of the universe as a whole.

critical density to close the universe: the dividing value between a universe with density high enough for its gravitational pull to stop the cosmological expansion and one with density low enough that the expansion continues forever.

curvature: the bending or warping of space and time, predicted by general relativity and theories like it.

dark matter: Astronomers can determine the mass of galaxies using a variety of techniques. All of these methods find a mass that exceeds the mass in stars by more than a factor of 10. These observations imply that most of the mass of our Galaxy, and most of the mass of other galaxies, is made up of some kind of non-luminous matter or dark matter. Possible candidates for the dark matter range from subatomic particles to supermassive black holes. (See also "hot dark matter" and "cold dark matter.")

density fluctuations: The universe is not uniform. The density of matter varies from place to place. These variations are called density fluctuations.

distribution of galaxies: see large-scale structure.

dragging of inertial frames: a general relativistic phenomenon predicted to occur near rotating masses, in which freely falling laboratories would be dragged slightly around the body. One consequence is that a gyroscope in such a laboratory would precess with respect to the direction it would point in empty space.

Einstein's equation: a mathematical equation written down by Einstein in 1915 to describe how matter and energy curve space and time. This curvature accounts for gravity, superseding Newton's theory of a gravitational force, which remains a good approximation only when gravity is weak.

equation of state of nuclear matter: An equation of state describes how the density of a substance increases as the pressure on it is increased. Stars remain in equilibrium by balancing the inward pull of gravity against the outward pressure force, so the equation of state must be known to construct

theoretical models of stars. Neutron stars are at such high densities and pressures that their atoms have been completely crushed until the nuclei merge, producing "nuclear matter."

equivalence principle: a fundamental principle of general relativity one of whose consequences is that all objects (and light) fall in a gravitational field in the same way independent of their internal structure or other properties. This universality of free fall is one of the most accurately verified principles in physics.

event horizon: the surface of a black hole. It is a one-way membrane, allowing matter or signals to flow in but not out.

Friedmann-Robertson-Walker (FRW) cosmological models: the mathematical description of the simplest possible universe—with matter density and expansion rates the same everywhere in space and in all directions (homogeneous and isotropic). It approximately describes the behavior of our universe on the largest spatial scales.

galaxy: a large assemblage of stars and gas with a total mass in the range from 10 million to 100 billion solar masses. The Sun and the solar system are part of the Milky Way Galaxy.

gamma-ray: electromagnetic radiation more energetic than x-rays.

gamma-ray bursts: bursts of gamma-rays from cosmic sources observed by detectors on satellites. Several hundred are detected per year and range in duration from fractions of a second to several tens of seconds. They are seen distributed uniformly across the sky, suggesting that the sources are at cosmological distances.

general relativity: Einstein's theory of gravity in which the gravity is the curved geometry of space and time.

Global Positioning System (GPS): a U.S. navigation system in which 24 Earth-orbiting satellites carrying atomic clocks broadcast precise time and location information. A receiver intercepting signals from four or more GPS satellites uses the information to determine its precise absolute location, in some circumstances to better than 15 meters.

gravitational collapse: A star remains in equilibrium by balancing the inward pull of gravity against an outward pressure force. If gravity overwhelms the pressure, nothing can hold the star up, and it undergoes gravitational collapse to a black hole.

gravitational inverse square law: the prediction, originally due to Newton, that gravitational forces become weaker as the inverse square of the distance between objects. Any quantum force produced by the exchange of massless objects must also satisfy an inverse square law.

gravitational lens: an object that deflects the rays of light from a distant astronomical source by the gravitational pull of an intermediate mass that may be

a galaxy or a cluster of galaxies. The deflection causes a distortion in the image of the distant source, and sometimes also leads to multiple images.

gravitational wave: a ripple in the geometry of spacetime propagating as a wave according to general relativity.

gravitational wave background: (1) gravitational waves arriving from so many sources that the individual signals are indistinguishable; (2) gravitational wave "static," akin to the (electromagnetic) cosmic background radiation.

gravitational Wilson loops: one-dimensional integrals of the quantum geometric variables around closed paths that are analogous to the similar quantities occurring in gauge theories.

Gravity Probe B: also known as the NASA Relativity Mission. It is expected to make the first measurement of the "frame-dragging" effect predicted by general relativity, i.e., the modification of spacetime produced by Earth's rotation. The satellite contains ultraprecise gyroscopes whose precession with respect to the fixed stars is monitored (for information online, see <http://einstein.stanford.edu>).

Hawking radiation: the approximately thermal radiation emitted by a black hole as a result of quantum effects.

hertz: a standard unit of frequency equal to one cycle per second.

higher dimensions: String theory predicts that space has more than three dimensions, but the extra dimensions can be seen only at very high energies.

"holographic" view of space and time: the possibility that physical processes in a region of space can be completely described by quantities defined only on the boundary of that region.

homogeneous: a situation in which the basic properties of a system are the same from place to place, at a selected moment of time. When observed at scales so large that the fine details of galaxies and clusters can be ignored, the universe appears to be homogeneous.

horizon: see event horizon.

hot dark matter: dark matter (q.v.) that is moving today with velocities comparable to or equal to the velocity of light. Neutrinos are an example.

Hubble constant (H_0): a measure of the expansion rate of the universe (usually in units of kilometers per second of increasing galaxy recession speed per megaparsec [10^6 parsecs] of galaxy distance). $1/H_0$ is a measure of the age of the universe.

inflationary universe: The inflationary theory proposes an extremely rapid period of expansion shortly after the big bang. During this rapid expansion, the energy density of the universe was dominated by vacuum energy. This vacuum energy later was converted into the matter and radiation that fills the universe today.

infrared: a region of the electromagnetic spectrum with wavelengths longer than those of visible light. Hot objects typically are very bright at infrared wavelengths.

isotropic: a situation in which the basic properties of a system observed from one location are the same in every direction. The cosmic background radiation (q.v.) is isotropic to a few parts in a hundred thousand.

LAGEOS (laser-ranged geodynamics satellite): a massive, spherical satellite studded with laser reflectors, in orbit around Earth. Accurate tracking of the two existing LAGEOS satellites yields information about Earth's structure (through its gravity field) and about the motion of the ground on which the tracking stations sit, and can study relativistic effects in gravity (for information online, see <http://ilrs.gsfc.nasa.gov/ilrs/lageos.html>).

large-scale structure: the arrangements of the positions of the galaxies in the universe. On the largest scales this shows clusters, walls, and voids.

laser interferometer: a device that uses laser light to make accurate comparison of the lengths of two perpendicular paths.

Lense-Thirring effect: synonymous with "dragging of inertial frames" (q.v.). The effect is named after Josef Lense and Hans Thirring, Austrian physicists who first calculated the general relativistic predictions for dragging in 1918.

LIGO (Laser Interferometer Gravitational-Wave Observatory): an NSF-sponsored project to build and operate two 4-kilometer laser interferometers to detect gravitational waves (for information online, see <http://www.ligo.caltech.edu/>).

lunar laser ranging: a technique for precise determination of the lunar orbit, in which laser beams are bounced off special reflectors deposited on the Moon by Apollo astronauts and Soviet unmanned landers. The Earth-Moon distance can be effectively monitored to centimeter accuracies.

MAP satellite: NASA's Microwave Anisotropy Probe, which is scheduled for launch in fall 2000, will accurately map the microwave sky with an angular resolution of 0.2 degrees. At MAP's frequencies (22 to 96 GHz), most of the fluctuations in the microwave sky are due to variations in the cosmic microwave background (for information online, see <http://map.gsfc.nasa.gov>).

microlensing: the phenomenon in which the deflecting mass of a gravitational lens (q.v.) is a star rather than a galaxy or cluster, with a correspondingly smaller angle of deflection.

neutron star: a star at such a high density and pressure that its atoms have been completely crushed until the nuclei merge and most of the electrons have been squeezed onto the protons, forming neutron-rich material.

Neutron Star Grand Challenge: a collaboration funded by NASA to calculate the properties of a binary system of two neutron stars coalescing as they emit gravitational radiation (cf. Binary Black Hole Alliance).

Newtonian gravity: Newton's theory of gravity, which states that falling and orbiting of a mass in the vicinity of another mass are caused by an attractive force along a line joining them. This theory is the limit of general relativity when speeds are much less than the speed of light and gravitational fields are weak.

Newton's gravitational constant G: the fundamental constant that determines the strength of all gravitational phenomena. In the usual SI system, it is numerically equal to the gravitational force (in Newtons) between two 1-kilogram objects separated by 1 meter.

"no hair" theorem for black holes: the surprising result that no matter how complicated the initial physical situation that produces a black hole, the final black hole is described by only a few parameters. The simplest black holes are completely specified by only their mass, charge, and spin.

parsec: a unit of distance of roughly 3 light-years or 3×10^{16} meters.

Planck length: the unique length that can be constructed from Newton's gravitational constant, the velocity of light, and the quantum of action and that characterizes quantum gravity phenomena. Its value is 10^{-33} cm. There is a corresponding Planck energy (10^{19} GeV) and Planck time (10^{-43} s).

Planck satellite: the European Space Agency's Planck surveyor satellite, scheduled for launch in 2007, will measure the microwave sky over a wide range of wavelengths (22 to 900 GHz) with an angular resolution of 0.1 degree (for information online, see <http://astro.estec.esa.nl/SA-general/Projects/Planck/>).

polarization: the directional pattern of a wave's effects on test bodies.

polymer-like geometry: an intuitive term used to describe geometry of the physical space at the smallest scales in a specific approach to quantum gravity. At a fundamental level, space resembles a one-dimensional polymer, and the three-dimensional continuum arises only as an approximation at scales much larger than 10^{-33} cm.

post-Newtonian: General relativity and theories like it reproduce Newton's laws of gravity as a first approximation. The first general relativistic corrections beyond Newtonian theory (called post-Newtonian) are responsible for such phenomena as the bending of light or the advance of Mercury's perihelion. This approximation is not valid in the vicinity of neutron stars or black holes or at the big bang.

precession of an orbit: An orbit precesses when the long axis of its elliptical shape rotates slowly rather than remaining in a fixed direction.

principle of equivalence: see equivalence principle.

principle of relativity: the statement that there is no absolute rest; given two objects in uniform relative motion, each one can be treated as being at rest and the other as moving with respect to it.

pulsar: a spinning neutron star that emits beamed radiation. The radiation is received as series of pulses as the beam sweeps over the observer—just like a lighthouse.

quadrupole moment: a mathematical quantity that measures non-sphericity of a mass distribution. If this quantity changes in time, the body emits gravitational waves.

quantum cosmology: the area of physics and astrophysics concerned with a theory of the quantum initial state of the universe (q.v.) and its consequences for observations today.

quantum initial state of the universe: sometimes called the wave function of the universe. If quantum mechanics applies to the universe as a whole, then the universe must have been in some particular initial quantum state. That state provides a boundary condition for the big bang and all subsequent epochs of the universe.

quantum of area: an elementary unit of area whose existence is predicted by the quantum theory of geometry (q.v.).

quantum theory of geometry: a theory of the microscopic structure of space, derived from the application of the principles of quantum mechanics to the geometry of physical space.

quasar: a very compact and extraordinarily luminous source of radiation in the nucleus of a distant galaxy. Quasars are believed to be powered by accretion of gas onto massive black holes.

quasi-normal modes: the characteristic vibrations of a disturbed black hole, analogous to the normal modes of vibration of a bell. The black hole's quasi-normal modes are, however, strongly damped by the emission of gravitational waves.

quasi-periodic oscillations: rapid not-quite-regular variations in the brightness of the x-rays emitted by matter accreting (q.v.) onto a neutron star or black hole. The almost periodic variations (whose period varies in time) are believed to reflect the dynamics of the inner part of the disk of accreting matter.

radio waves: electromagnetic waves with wavelengths very long compared to those of visible light. The radio band is usually considered to include all electromagnetic waves with wavelengths greater than about 1 millimeter.

redshift: the shift of a spectral line to longer wavelengths. The radiation emitted from a receding body on Earth, or a receding galaxy in the universe, or a body deep in the gravitational potential of a black hole is all redshifted.

Regge calculus: an approximation to general relativity in which geometry is represented by flat units joined together to approximate a four-dimensional

curved spacetime, much in the same way that the two-dimensional curved surface of a geodesic dome is made from flat triangles.

ring-down: loss of energy by gravitational waves as a newly formed black hole settles down to equilibrium (cf. quasi-normal modes).

scalar-tensor gravitational theory: a theory that modifies general relativity by adding a field known as a "scalar" to the equations for spacetime curvature. This scalar, which ascribes a single number to each point in space and time, can play the role of the gravitational coupling strength (usually denoted *G*), which thus can vary in space and time. In general relativity, *G* is strictly constant.

Schwarzschild radius: the location of the "surface" of a black hole, from whose interior it is impossible to escape.

singularity: a region of the universe where a classical description breaks down because it predicts infinite spacetime curvature or density of matter. General relativity predicts that a singularity is the ultimate result of gravitational collapse.

solar mass: the mass of the Sun.

spacetime: the four-dimensional continuum in which we live, consisting of the three dimensions of space and one dimension of time. General relativity (q.v.) is concerned with the curvature (q.v.) of spacetime.

special relativity: Einstein's theory of spacetime structure, in which Newton's notion of absolute time is abandoned to account for the experimental fact that the speed of light is a universal constant and does not depend on the relative motion between the observer and the light source.

string theory: a new physical theory that appears to be both a consistent quantum theory of gravity and a unified theory of all particles and forces.

strong gravitational fields: gravitational fields that are so strong that Newton's theory of gravity is inadequate, because the new effects predicted by general relativity are important.

structure formation: Galaxies are clustered into filaments, sheets, and clusters. Structure formation describes how gravity forms these structures out of tiny initial density fluctuations.

supermassive black holes: The cores of most galaxies appear to contain black holes with masses 1 million to 1 billion times the mass of our Sun. These supermassive black holes are thought to be the engines that power quasars. Our own Galaxy has a 2-million-solar-mass black hole in its center.

supernova: a gigantic explosion that signals the death of a massive star. Often, the explosion leaves behind a neutron star; in other cases it may produce a black hole. (Cf. Type IA supernova, Type II supernova.)

Supernova 1987A: a supernova that was observed in 1987 in the nearby galaxy called the Large Magellanic Cloud. This is the closest supernova to have been observed since the invention of the telescope.

supersymmetry: a postulated symmetry relating particles with integer and half-integer spin. This symmetry would relate all of the elementary particles and forces in a grand unified theory. The validity of supersymmetry at high energy is required for string theory (q.v.) to describe our world.

thermal radiation: radiation emitted with a blackbody spectrum (q.v.).

topology change: a possible physical phenomenon in which the topology of space (q.v.) changes. Classical general relativity forbids these changes in all physically reasonable circumstances. Whether quantum effects of gravity will enable them is an important open question.

topology of space: A two-dimensional surface could be an infinite plane, but might instead have the form of the surface of a sphere, of a donut, or a more complicated shape. The topology of space refers to the analogous attributes of our physical three-dimensional space.

Type IA supernova: the sudden burning of the carbon and oxygen in a white dwarf star, producing a powerful explosion. Because they appear to have uniform peak luminosities, supernovae of Type IA are used as standard candles to measure the geometry of the universe.

Type II supernova: When a massive star has exhausted all of its nuclear fuel, its core collapses and forms a neutron star. In the process it blows off the outer envelope in a gigantic explosion, releasing a thousand times more energy in a millisecond than our Sun will release in its entire lifetime. Most of this energy is released in the form of neutrinos. While photons carry off less than 1 percent of the total energy of a supernova, a single supernova will outshine an entire galaxy for several weeks. If the collapse is asymmetric, then the supernova explosion will also radiate gravitational waves.

unified theory, unification of all forces: a theory that in one conceptual framework provides a fundamental theory of the electromagnetic, weak, strong, and (often) the gravitational interaction. String theory is an example. The characteristic features of such a theory are expected to emerge only at very high energies at or near the Planck scale.

universality of free fall (UFF): a central prediction of general relativity that the gravitational acceleration of a small object depends only on its location in space, and not on any properties of the object itself.

velocity of light c: a very high speed (about 3×10^8 m/s) that is "nature's speed limit." This is the speed at which any massless object (such as light) travels. Objects with mass must always travel more slowly than c.

waveform: a wave's strength (amplitude) as a function of time.

weakly interacting massive particles: hypothetical candidates for the dark matter. These particles are potentially detectable in underground dark matter searches. Alternatively, they can be produced in particle accelerators. They are a form of cold dark matter.

X-ray binary: a double star in which one of the stars accretes matter from its companion and emits a copious amount of x-rays. The x-ray-emitting star is either a black hole or a neutron star.