

GRAVITATION AND GENERAL RELATIVITY

BERNARD F. SCHUTZ, *Department of Physics and Astronomy, University of Wales College of Cardiff, Cardiff, Wales, United Kingdom*

CLIFFORD M. WILL, *Department of Physics, Washington University, St. Louis, Missouri, U.S.A.*

| | | | | |
|--|-----|---|---|-----|
| Introduction | 304 | 1.2.3 | General Relativity: Gravity as Geometry | 312 |
| 1. General Relativity and Relativistic Gravity | 305 | 1.2.3.1 | Why Gravity Is Geometrical... | 312 |
| 1.1 Gravitation Theory: An Overview | 305 | 1.2.3.2 | The Metric of Gravity..... | 312 |
| 1.1.1 Relativistic Gravity in Physics | 305 | 1.2.3.3 | Local Flatness of Spacetime and the Local Freely Falling Frames | 313 |
| 1.1.1.1 Roots in Newtonian Gravity... | 305 | 1.2.3.4 | Curvature and Tidal Forces ... | 313 |
| 1.1.1.2 The Impact of Special Relativity: Relativistic Gravity Is Required | 306 | 1.2.4 | Sources of Gravity: How Matter Creates the Geometry.. | 314 |
| 1.1.1.3 The Impact of Quantum Mechanics: Quantum Gravity Is Required | 306 | 1.2.4.1 | The Stress-Energy Tensor..... | 314 |
| 1.1.2 Where Relativistic Gravity Is Important..... | 307 | 1.2.4.2 | Einstein's Field Equations..... | 314 |
| 1.1.2.1 Compact Objects of High Density | 307 | 1.2.5 | Other Theories of Gravity | 315 |
| 1.1.2.2 Cosmology: Relativistic Gravity at Low Densities..... | 307 | 1.2.6 | A Cosmological Term in Einstein's Equations..... | 315 |
| 1.1.2.3 Small Relativistic Effects: Gravitational Waves and Equations of Motion..... | 307 | 2. Some Consequences of Einstein's Field Equations ... | | 316 |
| 1.2 Fundamental Ideas and New Concepts of General Relativity | 308 | 2.1 | Momentum and Stress Also Make Gravity | 316 |
| 1.2.1 The Incorporation of Newtonian Gravity | 308 | 2.1.1 | Gravitomagnetism..... | 316 |
| 1.2.1.1 The Equivalence Principle: How Gravity Affects Matter ... | 308 | 2.1.2 | Gravitational Collapse | 317 |
| 1.2.1.2 Newton's Field Equation: How Matter Creates Gravity..... | 308 | 2.2 | Black-Hole Theory | 317 |
| 1.2.1.3 Galileo's Principle of Relativity | 308 | 2.2.1 | Black Holes in the Eighteenth Century..... | 317 |
| 1.2.2 Basic Ingredients of General Relativity | 308 | 2.2.2 | Black Holes in General Relativity | 317 |
| 1.2.2.1 Special Relativity..... | 308 | 2.2.3 | Singularities Inside the Hole .. | 318 |
| 1.2.2.2 The Einstein Equivalence Principle..... | 310 | 2.2.4 | Black Holes Have No Hair.... | 318 |
| 1.2.2.3 Tidal Forces Are the Real Gravity | 310 | 2.2.5 | Black-Hole Thermodynamics.. | 318 |
| 1.2.2.4 Gravitational Redshift | 310 | 2.2.6 | Wormholes | 319 |
| 1.2.2.5 Gravitational Deflection of Light | 311 | 2.3 | Gravitational Waves | 319 |
| | | 2.3.1 | The Necessity of Gravitational Waves | 319 |
| | | 2.3.2 | The Interaction of Gravitational Waves with Matter | 319 |
| | | 2.3.3 | Wave Emission: The Quadrupole Formula | 321 |
| | | 2.4 | Applications of General Relativity | 322 |
| | | 2.4.1 | Relativistic Stars (Pulsars) and Gravitational Collapse | 322 |
| | | 2.4.1.1 | Stellar Evolution | 322 |
| | | 2.4.1.2 | White Dwarfs | 322 |

304 Gravitation and General Relativity

| | | | | | |
|---------|--|-----|---------|---|-----|
| 2.4.1.3 | Neutron Stars | 322 | 3.2.1 | The Deflection and Retardation of Light | 330 |
| 2.4.1.4 | Black Holes | 323 | 3.2.1.1 | Light Deflection | 330 |
| 2.4.1.5 | Pulsars | 323 | 3.2.1.2 | Shapiro Time Delay | 331 |
| 2.4.2 | Black Holes in X-Ray Binaries and in Galactic Centers | 323 | 3.2.2 | Mercury's Perihelion Advance | 331 |
| 2.4.2.1 | X-Ray Binaries | 323 | 3.2.3 | Test of the Strong Equivalence Principle | 332 |
| 2.4.2.2 | Galactic Black Holes | 324 | 3.3 | The Binary Pulsar: An Astronomical Relativity Laboratory | 332 |
| 2.4.3 | Gravitational Lensing | 324 | 3.4 | Future Work in Experimental Gravitation | 333 |
| 2.4.4 | Solar System and Stellar Orbits | 325 | 3.4.1 | Search for Gravitomagnetic Effects | 333 |
| 2.4.5 | Cosmology, Inflation, and the Origins of the Universe | 325 | 3.4.2 | Tests of the Einstein Equivalence Principle | 334 |
| 2.4.5.1 | Cosmological Models | 326 | 3.4.3 | Further Fifth-Force Searches .. | 334 |
| 2.4.5.2 | Hubble's Law: The Big Bang .. | 326 | 4. | Gravitational-Wave Detection: A Technological Frontier | 334 |
| 2.4.5.3 | Cosmic Microwave Background Radiation | 326 | 4.1 | Likely Sources of Detectable Waves | 334 |
| 2.4.5.4 | The Age of the Universe | 327 | 4.1.1 | Supernovae | 334 |
| 2.4.5.5 | Mass and Missing Mass | 327 | 4.1.2 | Coalescing Binaries | 335 |
| 2.4.5.6 | Microwave Anisotropies | 327 | 4.1.3 | Pulsars | 336 |
| 2.4.5.7 | Inflation and the Homogeneity Problem | 328 | 4.1.4 | Ordinary Binaries | 336 |
| 3. | Tests of Gravitational Theories and Their Technological Demands | 328 | 4.1.5 | Cosmological Background | 336 |
| 3.1 | Tests of the Einstein Equivalence Principle | 328 | 4.1.6 | Unexpected Sources | 336 |
| 3.1.1 | Tests of Special Relativity | 328 | 4.2 | Detectors | 336 |
| 3.1.1.1 | Michelson-Morley Experiments | 329 | 4.2.1 | Bar-Type Detectors | 337 |
| 3.1.1.2 | Hughes-Drever Experiment ... | 329 | 4.2.2 | Laser-Interferometric Detectors | 337 |
| 3.1.2 | The Eötvös Experiment, the Weak Equivalence Principle, and the Fifth Force | 329 | 4.2.3 | Space-Based Detectors | 338 |
| 3.1.2.1 | Eötvös and the Weak Equivalence Principle | 329 | | Glossary | 338 |
| 3.1.2.2 | Fifth-Force Experiments | 329 | | Works Cited | 338 |
| 3.1.3 | Gravitational Redshift | 330 | | Further Reading | 340 |
| 3.2 | Solar-System Tests of General Relativity | 330 | | | |

INTRODUCTION

Albert Einstein's development of general relativity in 1915 was one of the landmarks of twentieth-century theoretical physics. Einstein (1879–1955) had few experimental facts to guide him in replacing Newtonian gravity with a theory compatible with his earlier theory of special relativity (see RELATIVITY, SPECIAL). Using arguments of consistency, simplicity, and aesthetics, Einstein arrived at

a theory that has subsequently passed stringent experimental tests, and that has predicted and explained a wide range of astronomical phenomena that were completely unknown at the time Einstein did his work.

Einstein's methodology had a great influence on the style of later attempts to simplify and unify theories of high-energy physics. Nevertheless, it is only since about 1970 that developments in astronomy and in experimental physics have placed general relativity into

the mainstream of physics, part of the expected competence of any theoretical physicist.

General relativity is now the accepted theory of gravity. Although compatible with special relativity, it is incompatible with quantum theory. Presumably it will eventually be replaced by a quantum theory of gravity, and for this reason there is considerable interest in alternative theories of gravity and in testing gravity to find small departures from the classical theory. So far, all experiments and observations have been consistent with general relativity, many at a high level of precision.

This article discusses relativistic gravitation from a nonmathematical point of view, highlighting the underlying connections of general relativity with the rest of physics (including Newton's theory of gravity), examining the foundations of the theory (equivalence principle, geometrical approach), and emphasizing the observable and experimental consequences of the theory. These include astronomical phenomena, such as black holes, gravitational lenses, and the Big Bang; high-precision tests of the equivalence principle, of solar-system effects, and of gravitational radiation; and present efforts to make direct observations of gravitational waves from supernova explosions, colliding black holes and neutron stars, and the Big Bang itself.

Most aspects of experimental gravity push modern technology to its limits, and they are therefore fruitful areas for the application and development of new technology. Examples include the development of low-temperature superconducting quantum-interference device (SQUID) technology for space-based tests of magnetogravity and for gravitational-wave detectors; the routine employment of corrections for the gravitational redshift in the U.S. Air Force's Global Positioning System (GPS) navigation system; the application of dilution refrigeration on a large scale to cool gravitational-wave bar detectors weighing several tons to temperatures approaching 10 mK; and the use of high-power cw lasers and ultralow-loss optical components in interferometric gravitational-wave detectors. Current research into squeezed light for telecommunications (see OPTICAL COMMUNICATIONS) was stimulated in large part by theoretical work on ways of making gravitational-wave detectors more sensitive.

The field of experimental relativity is an active one; this article presents a snapshot of current developments as of the time of its completion, early in 1993.

1. GENERAL RELATIVITY AND RELATIVISTIC GRAVITY

Einstein's theory of general relativity (GR) is the accepted theory of gravity in modern physics. In the Solar System and other regions of weak gravitational fields, it differs little from Newton's theory of gravity; but when gravity is strong enough to accelerate particles to near the speed of light, its predictions are very different. General relativity describes gravity in geometrical terms; the curved geometry of space and time is an arena in which all the rest of physics takes place. The theory has become an everyday tool for astrophysicists who model the most unusual phenomena in astronomy: black holes, gravitational waves, and the theory of cosmology are all features of the theory that find application in astronomy. For textbook introductions to general relativity, see Misner *et al.* (1973) or Schutz (1985).

1.1 Gravitation Theory: An Overview

1.1.1 Relativistic Gravity in Physics

1.1.1.1 Roots in Newtonian Gravity.

Newton's laws of motion and his theory of gravity had a revolutionary effect on science. They explained the motions of the planets and taught scientists that the heavenly bodies were subject to the same laws as bodies on Earth. Newton showed that masses exert forces on each other that diminish with the square of the distance between them. The success of Newtonian gravity in the Solar System, as well as in other stellar systems, is described in the article ASTRONOMY.

In Newton's theory, the gravitational forces between bodies act instantaneously across any distance. Space is flat and time is universal: once synchronized, identical clocks keep time with one another regardless of their state of motion. It was not until the nineteenth century that physicists had any experimental evidence that challenged this idea about time (the Michelson–Morley experiment). The first evidence for the *curvature* of space came from the 1919 eclipse expedition, described in Sec. 3.2.1 below.

1.1.1.2 The Impact of Special Relativity: Relativistic Gravity is Required. Einstein's development of special relativity (SR) changed notions of space and time (see RELATIVITY, SPECIAL). Space, while still flat, merged with time to form *spacetime*. Clocks in relative motion would not remain synchronized, and moving lengths would contract. Most importantly, no influences could travel faster than light.

This limiting speed undermines Newtonian gravity: gravity cannot "act at a distance" without a delay. In the solar system, where light crosses an orbit in a small fraction of the orbital period, the delay is small, and Newton's theory works well. But the theory had to be replaced.

It took Einstein another decade to find a suitable replacement: *general relativity*. It incorporates SR as the limit of vanishing gravity. It approaches Newton's theory as a limit when gravity is the dominant force but is still weak enough that the resulting motions are much slower than light. But in other regimes GR makes startlingly new predictions. Black holes, gravitational waves, the Big Bang: Einstein had no inkling of these when he devised the theory, and it is remarkable that these unforeseen consequences of the theory have been verified by astronomical observations in recent years.

It is interesting that Einstein was "beaten" to the correct form of the equations of GR by the great mathematician David Hilbert, whose paper (Hilbert, 1915) was submitted five days before Einstein presented his theory to the Prussian Academy (Einstein, 1915). Einstein is universally credited with the theory because his physical insight into its foundations, such as the equivalence principle and the geometric form of the theory, guided Hilbert as much as they guided Einstein; and it was Einstein who worked out the first physical consequences of the equations.

Despite GR's success, it is not the only possible relativistic generalization of Newtonian gravity. In many ways it is the simplest, but many others exist. Most, like the Brans-Dicke-Jordan class of theories (Sec. 1.2.5), introduce extra fields that have further gravitational effects. Others depart from GR by violating the equivalence principle (Sec. 1.2.2.2). The only way to decide which theory is right in detail is to test them experimentally. Testing relativistic gravity is one of the areas

in which modern technology has made possible great advances (Sec. 3).

1.1.1.3 The Impact of Quantum Mechanics: Quantum Gravity is Required. Although GR reconciled the contradiction between SR and Newtonian gravity, it soon fell afoul of a different inconsistency. The theory of quantum mechanics matured only two decades after the birth of GR, and now almost all of physics is described by quantum theories; GR is the only major exception.

Much current research attempts to find an appropriate quantum theory of gravity. So far that effort, while producing many interesting new ideas, has not reached its goal. Because GR is a highly nonlinear theory, conventional methods of quantizing theories have so far failed to produce a consistent quantum theory of gravity. The search is made more difficult because of the paucity of experimental guidance. If there are properties of the known universe that give us decisive clues to quantum gravity, we have not yet recognized them.

Dimensional analysis is interesting in this respect. If quantum gravitational effects dominate in some situation, then the laws must involve the three fundamental constants of nature, G (Newton's constant of gravitation), h (Planck's constant), and c (the speed of light). From these constants one can construct numbers with any of the fundamental dimensions, e.g.,

$$(hc^5/G)^{1/2} = 4.9 \times 10^9 \text{ J} = 3.1 \times 10^{28} \text{ eV}$$

(Planck energy); (1)

$$(hG/c^3)^{1/2} = 4.1 \times 10^{-35} \text{ s}$$

(Planck time); (2)

$$(hG/c^5)^{1/2} = 1.4 \times 10^{-43} \text{ m}$$

(Planck length). (3)

These are so far from accessible experimental domains that it is not surprising that we have little experimental guidance for quantum gravity.

It may well be, however, that effects related to quantum gravity are important on other scales, particularly if gravity is part of a *grand unified theory* of physical interactions. (See UNIFIED FIELD THEORIES.) There may be small corrections to Einstein's equations that are

appropriate in a semiclassical limit, and these might take the form of extra gravitational fields. For this reason, there has been renewed interest of late in alternative theories of gravity and in the experimental limits that can be placed on them.

The regime of quantum gravity may be far from the experimental domain, but the theoretical motivation for finding the right theory is very strong. Besides the need simply for consistency with other theories, there is also the striking fact that GR seems to predict its own failure: many solutions, including apparently all black-hole solutions, contain *singularities*, which are places where GR fails to describe the Universe in a consistent way. (See Sec. 2.2.3.) The Big Bang begins with such a singularity. There is a strong feeling that quantum gravity will tell us what the singularities really mean.

1.1.2 Where Relativistic Gravity Is Important. Relativistic gravity is required where the speeds produced by gravitational attractions approach the speed of light. A system with mass M and overall size R will produce velocities of order

$$v_{\text{self-gravity}} = (GM/R)^{1/2}. \quad (4)$$

For example, the escape velocity from a spherical body is $\sqrt{2}v_{\text{self-gravity}}$, and the velocity of a planet in a circular orbit of radius R about a star of mass M is exactly $v_{\text{self-gravity}}$. A test for the importance of relativistic effects is

$$\Phi := GM/Rc^2 \ll 1 \Rightarrow \text{gravity is nonrelativistic.} \quad (5)$$

At the Sun's surface, $\Phi = 2 \times 10^{-6}$; everywhere else in the Solar System it is smaller. This accounts for the accuracy of Newtonian gravity. One needs relativistic gravity when Φ is not small.

1.1.2.1 Compact Objects of High Density.

To make Φ larger one can increase M and/or decrease R . If a body of fixed mass shrinks in size, relativistic effects become important when the radius nears the *gravitational radius* R_g :

$$R_g := 2GM/c^2. \quad (6)$$

The factor of 2 in this is conventional.

The gravitational radius of the Sun is about 3 km. Remarkably, stars do collapse to nearly this size: perhaps 1% of all stars are *neutron stars* (Sec. 2.4.1), with radii about 10 km. If a

star actually reaches the gravitational radius, then it collapses further to a *black hole* (Sec. 2.2).

The density ρ of a neutron star is enormous. If we take $M = 1M_\odot$ (astronomer's notation for the mass of the Sun) and $R = 10$ km, we have $\rho_{\text{NS}} = 5 \times 10^{17} \text{ kg m}^{-3}$. This is comparable to the density of a typical atomic nucleus.

The critical density ρ_c at which the object reaches its gravitational radius depends only on M :

$$\rho_c := 3c^6/32\pi G^3 M^2. \quad (7)$$

For the Sun it is about $2 \times 10^{19} \text{ kg m}^{-3}$, but such enormous densities are not required for bodies of larger mass. A cluster of some 10^8 stars will collapse to a black hole if it reaches the density of water. This seems to have happened in the centers of galaxies (Sec. 2.4.2.2).

1.1.2.2 Cosmology: Relativistic Gravity at Low Densities. A second interesting way of reaching a situation where relativistic gravity is required is by increasing the size of a system of fixed density. As the mass increases, the critical density drops. Eventually the critical density approaches the system's fixed density and the body becomes relativistic. For most systems, this size would be unrealistically large. But there is one system that we can make as large as we like: the Universe itself.

The average mass density of the Universe (or at least of the piece we can see) is highly uncertain, but an observational lower bound is $10^{-28} \text{ kg m}^{-3}$. A sphere of radius $1.3 \times 10^{27} \text{ m}$, or about 10^{11} light years, would enclose enough matter to give a relativistic gravitational field. Such distances are the regime of *cosmology*, the study of the Universe as a whole (Sec. 2.4.5).

Being able to describe cosmology consistently was in fact one of the great triumphs of GR: Newtonian theory is highly ambiguous when applied to infinite systems, whereas GR provides well-defined models.

1.1.2.3 Small Relativistic Effects: Gravitational Waves and Equations of Motion. Relativistic gravity is essential for the description of compact objects and of cosmology, but it can also be important for situations where its effects make small but measurable changes to Newtonian gravity. These are easiest to measure when the new relativistic effect is

absent from Newtonian gravity, so that its signature is easy to see. Some effects we will look at below include the following:

1. Gravitational radiation (Sec. 2.3) is a direct consequence of the limitation on the speed of propagation of gravitational influences: oscillating motions of stars (say, in a binary system) produce an oscillating field that moves through space, a wave in the gravitational field. These can be directly detected essentially by looking for time-dependent gravitational fields. The weakness of gravity makes this one of the greatest technical challenges in physics today (Sec. 4). Even when direct detection of the waves is impossible, the back-reaction effects of their emission can sometimes be observed. An example of great importance is the binary pulsar PSR1913+16 (Sec. 3.3).
2. Gravity deflects light and produces observable time delays, effects that can be directly observed (Sec. 3.2.1). Light deflection also leads to gravitational lensing of astronomical images (Sec. 2.4.3).
3. Relativistic gravity produces small anomalies in the motions of planets and satellites of the Earth. We will discuss these in Secs. 3.2.2 and 3.4.1.

1.2 Fundamental Ideas and New Concepts of General Relativity

1.2.1 The Incorporation of Newtonian Gravity. Any theory of gravity must address two issues: how gravity affects matter, and how matter creates gravity. We review these in the context of Newtonian gravity, and then look at what new ideas Einstein brought to the subject.

1.2.1.1 The Equivalence Principle: How Gravity Affects Matter. Gravity affects everything: unlike any other force in nature, it cannot be turned off or screened out. Galileo showed that it affects all things in the same way: all bodies fall with the same acceleration in a gravitational field. For this reason, one cannot experimentally distinguish a uniform gravitational field from a uniform acceleration of the experimenter. This equivalence between gravity and acceleration is called the *weak equivalence principle* (WEP).

1.2.1.2 Newton's Field Equation: How Matter Creates Gravity. Newton's law for the gravitational force between two bodies of masses m_1 and m_2 , separated by a distance r ,

$$F_{\text{grav}} = Gm_1m_2/r^2, \quad (8)$$

incorporated the equivalence principle: the acceleration of body 1, F_{grav}/m_1 , is independent of m_1 . Newton arrived at the $1/r^2$ law by studying the motion of the Moon, but the resulting law turned out to describe all the known properties of the solar system. One of its great triumphs was that it predicted that planetary orbits should be ellipses with the Sun at a focus.

Later development of Newton's theory found that the simple law given by Eq. (8) was in an inconvenient form to deal with complicated bodies. The modern form introduces the *gravitational potential* created by particle 1 as the following integral over its density ρ_1 :

$$\phi_1(\mathbf{r}) = -G \int \frac{\rho_1(\mathbf{r}') d^3x'}{|\mathbf{r} - \mathbf{r}'|}. \quad (9)$$

The force on particle 2 due to particle 1 is

$$\mathbf{F}_{12} = -m_2 \nabla \phi_1(\mathbf{r}). \quad (10)$$

The field ϕ solves the *Poisson equation*:

$$\nabla^2 \phi = 4\pi G \rho. \quad (11)$$

This very compact form for Newton's potential is usually known as *Newton's field equation*. It is the equation that must be generalized in any new theory of gravity.

1.2.1.3 Galileo's Principle of Relativity. Galileo's other seminal contribution to modern gravitational theory is what is now called the *principle of relativity*. A simple statement of the principle is that the (uniform) speed of an experimenter does not affect the outcome of an experiment. Galileo used the example of a ship moving on a calm sea: if you stand on the deck and drop a ball, it falls straight down relative to you, not along a vertical line as judged by an observer on the shore. For modern travelers, the experience of flying in

an airplane at 900 km/h is more compelling: unless one hits air turbulence or looks out the window, it is impossible to tell how fast one is going.

1.2.2 Basic Ingredients of General Relativity

1.2.2.1 Special Relativity. Newtonian gravity already incorporated the Galilean principle of relativity, that the laws of physics were independent of the state of motion of the experimenter. In SR, this becomes formalized into the idea of an *observer*, who is basically an entire reference frame with an information-gathering mechanism to determine what happens in experiments. We usually restrict ourselves to so-called Lorentz observers, who move at a uniform speed relative to one another, and who normally use a rectangular coordinate system for locating things in space. Their time coordinate consists of a set of synchronized clocks.

What Einstein added to Galileo's principle in order to get SR was the central role of the speed of light c . In SR, the speed of light has the status of a physical law: it is invariant under a change of experimenter. If two different experimenters measure the speed of the *same* beam of light, they will *each* measure $c = 2.998 \times 10^8 \text{ m s}^{-1}$, regardless of their speed relative to each other.

Special relativity made many conceptual changes in physics, which in turn carry over to any relativistic theory of gravity. Space and time are no longer distinct, but form a four-dimensional *spacetime*, whose points are called *events*. The set of all events experienced by a given particle through all its history is called its *world line*. Different observers make different measurements of simultaneity, of distance (the *Lorentz contraction*), and of time (*time dilation*), but there is still an invariant measure of distance, called the *spacetime interval* ds^2 :

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2. \quad (12)$$

(There are many conventions for this definition. We adopt the one most commonly used in GR.)

This is a generalization of the Pythagorean theorem of Euclidean geometry. Just as in Euclidean geometry, the physical length of the hypotenuse of the right triangle is given by the square root is the sum of the squares of the sides, so also in SR is the square root of the

interval a physical observable. For two events separated by a positive (*spacelike*) interval, its square root is the length of a physical ruler that would stretch between the events in a frame in which the events are simultaneous. This is called the *proper distance* between the events. For two events separated by a negative (*timelike*) interval, we define the *proper time* $d\tau$ by $d\tau^2 = -ds^2/c^2$. The proper time is the time elapsed on a clock that experiences both events. Events separated by a zero (*null* or *lightlike*) interval have zero proper time between them.

Because spacetime is four dimensional, vectors have four components and are called *four-vectors*. A simple four-vector is a tangent vector to a timelike world line. All such tangent vectors at a given event are parallel to one another. A useful one, conventionally called the *four-velocity* U , is defined to have components

$$(U^t, U^x, U^y, U^z) = \left(\frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau} \right). \quad (13)$$

(Here we have introduced some conventions: vector indices are written as superscripts, and the time dimension is normally taken first.) Because τ is an invariant, the components of U change in the same way as the coordinates do when the observer changes. This is part of the definition of a four-vector. Note that for small velocities, the interval $d\tau$ is nearly equal to the time interval dt , and the spatial components of U are nearly equal to the ordinary velocity. At the other extreme, the four-velocity of a null world line (e.g., that of a photon) is undefined, because $d\tau = 0$ along it.

Although SR is normally described in terms of Lorentz observers with rectangular coordinates, one can allow other coordinate systems for both space and time. This complicates the mathematics, such as the form of Eq. (12), but is a price worth paying in particular circumstances, such as when dealing with an accelerated particle. What distinguishes SR from GR is, fundamentally, that it is *possible* to find a single Lorentz frame that covers all of spacetime. The notion that Eq. (12) describes the way rulers and clocks behave everywhere in spacetime is very restrictive; in particular, since the interval depends only on the coordinate differences between events and not on their absolute location (no explicit dependence on x, y, z, t), it follows that SR can only

hold in a spacetime that is perfectly uniform, or *homogeneous*.

Special relativity also unifies energy and momentum. Just as time is dilated under a change of observer, so is energy increased. Just as there is no invariant separation between space and time, so is there none between energy and momentum. They are parts of a single vector in spacetime, the energy-momentum four-vector, which equals the rest mass of the particle times its four-velocity. Its time component is the particle's energy, and its spatial components are the momentum. For photons, the energy-momentum vector exists, even though the four-velocity does not. Since this vector is null, photons have energy equal to c times their momentum.

The introduction of the energy-momentum vector was a great conceptual advance for physics. In Newtonian mechanics, neither energy nor momentum had any observer independence: both changed when the observer's velocity changed. In SR, there is a single energy-momentum four-vector, unique to the particle. Its time component (the energy) and spatial components (momentum) do change, but they are subsidiary to the unique four-vector.

The identification of an invariant four-vector is a way of thinking introduced by SR that has great importance for GR. It is easy to understand what it means for a scalar quantity to be invariant, such as the interval in Eq. (12): its value is the same for all observers. A four-vector is likewise an invariant under spacetime coordinate transformations, just as the electric field vector \mathbf{E} , for example, is under spatial rotations; its orientation with respect to other vectors is unchanged, even though its components change. (Many authors call such a thing a *geometrical object*.)

1.2.2.2 The Einstein Equivalence Principle. Special relativity requires a new formulation of the equivalence principle. In the Galilean principle, we traded off gravity against acceleration. But what does acceleration mean for light, whose speed cannot change? Fortunately, there is a new way to state the principle that is equally at home in Galilean or Einsteinian relativity.

The key is to focus on an experimenter who falls freely in a uniform gravitational field. This experimenter could, say, be in a spaceship that is coasting through this field. Any other particle that falls freely (has no non-

gravitational forces on it) will keep a constant velocity with respect to the experimenter. This can then also apply to light: if a photon passes by the falling experimenter, it will maintain a constant relative velocity, maintaining not only its speed c but also its direction.

More generally, any observer who works in a *freely falling reference frame* in a uniform gravitational field will not be able to determine that there is a gravitational field at all. The results of any experimenters are just as in SR. This is the *Einstein equivalence principle* (EEP).

The EEP implies that SR can cope with *uniform* gravitational fields: just use freely falling observers and the effects of the fields go away. More importantly, it also means that SR is incompatible with nonuniform gravity, which is what one has in every real situation. Near the Earth, for example, the local acceleration of gravity changes from place to place. Since SR requires a Lorentz frame to exist and be rigid everywhere, it is not possible to identify such a frame in the presence of real gravity. Thus, SR cannot incorporate nonuniform gravity.

However, even in a nonuniform field, a *local* freely falling observer sees no gravity locally. Weightless astronauts feel no gravity, even though gravity holds them in orbit around the Earth. The EEP implies that gravity can still be removed *locally*, even if not globally. It is a powerful principle. Three of its consequences are explored in the next three sections.

1.2.2.3 Tidal Forces Are the Real Gravity.

The EEP implies that all the real, irremovable effects of gravity must therefore have to do with the differences between nearby freely falling observers. Any relativistic theory of gravity must give special importance to these differences between local gravitational accelerations.

These differences are called the *tidal forces*. They get their name from the tides raised on the Earth by the Moon and Sun. The Earth's oceans bulge out on both the near side and the far side with respect to the Moon. The near side of the Earth wants to fall towards the Moon slightly faster than the average, while the far side falls slightly slower. The size of the bulge is determined by the *difference* in the acceleration of gravity due to the Moon across the Earth. This gradient is called the tidal force of the Moon. If we could not see the

Moon, we could still infer its existence from observations of the tides: tidal forces are the observer-independent gravitational forces.

1.2.2.4 Gravitational Redshift. Photons are redshifted (lose energy) as they climb out of a gravitational field. A simple thought experiment shows how this follows from the EEP. We send a photon up from the ground to the top of a tower of height h , where it is detected by equipment at rest with respect to the Earth. During the time h/c it takes the photon to reach the top of the tower, a local freely falling frame has acquired a speed $v = gh/c$ downwards, where g is the acceleration of gravity on the surface of the Earth. In this frame, the EEP tells us that the photon's frequency does not change, as measured by this observer. Since the top of the tower is *rising* at speed gh/c relative to this observer, there is a Doppler redshift that produces a lengthening of the wavelength λ by a fractional amount

$$\frac{\delta\lambda}{\lambda} = \frac{v}{c} = \frac{gh}{c^2}. \quad (14)$$

The EEP implies that an experimenter who remains at rest with respect to the Earth measures this redshift of the photon as it climbs upwards in the gravitational field. Since the product gh is the change in the Newtonian gravitational potential between the emitter and the absorber, the redshift is just a function of the potential difference.

The gravitational redshift has another important interpretation. Since the frequency of the photon is measured essentially by clocks located at the emitting and absorbing locations, it is impossible for these clocks to remain synchronized. During a given number of oscillations of the photon's electromagnetic field, the clock on the ground ticked less time than the one higher up. The clock nearer the ground runs more slowly. *The gravitational redshift implies a gravitational time dilation.* This is now routinely measured by the GPS of satellites (Sec. 3.1.3).

1.2.2.5 Gravitational Deflection of Light.

Consider light passing a star. Since light must travel on a straight line with respect to a local freely falling observer, and since these observers all fall towards the center of the star, the light must continually bend its direction of travel in order to go on a straight line with

respect to each observer it happens to pass. We can estimate the size of the effect roughly by the following argument, which uses Newtonian concepts.

Let us consider just one freely falling observer who is at rest with respect to a star of mass M at the point where the light beam makes its closest approach to the star as it passes it by. Let this closest distance be R . The observer's acceleration towards the star is $g = GM/R^2$. Traveling at speed c , the photon will experience most of its deflection in a time of order R/c , the time it takes for the photon to move significantly further away from the star. During this time, the observer has acquired a speed $v = gR/c = GM/Rc$ perpendicular to the motion of the photon. By the EEP, the photon must also have acquired roughly this same speed transverse to its original direction. This means it changes its direction of travel relative to the star by an angle $\alpha = v/c = GM/Rc^2$ rad.

The total deflection should be double this, since the photon will experience the same deflection in to the point of nearest approach as going out. A more careful integration of this effect along the whole of the light path gives, coincidentally, exactly the same as our rough estimate, a total Newtonian deflection of $2GM/Rc^2$. This calculation was first performed by Cavendish in 1784 and independently by von Soldner in 1801, on the assumption that light was a corpuscle moving at speed c .

In relativistic gravity, however, there is an added effect that increases the deflection. This is, as we will see later, that space is curved near the star, and so the actual natural path of light as it passes the star, even if it were to go at an infinite speed, is deflected. The result in GR is that light is deflected by a net angle of

$$\alpha = 4GM/Rc^2, \quad (15)$$

double the Newtonian result.

For a light beam passing the Sun, just grazing its surface, the effect is small: α is less than 1.75 seconds of arc. This is measurable, as described in Sec. 3.2.1.

More dramatically, if the gravitational field is that of a faint cluster of galaxies, and the source of the light is a more distant but brighter quasar, then light from the quasar might reach the Earth by two or more different routes, giving separate images of the source separated by angles that are easy to resolve.

This is called *gravitational lensing*. By studying the distribution and relative brightness of multiple images, and the shape of distorted arclike images of a distant quasar, produced by an intervening galaxy or galactic cluster, astronomers can infer the distribution of mass in the cluster, particularly that of dark matter, which produces gravity, but not light. Indeed, one astronomer has dubbed this method *gravitational tomography*. By looking for differences in the arrival time of temporal features in the quasar signal (such as flare-ups) in the different images, a result of the light having traveled different path lengths, it may be possible to determine the distance of the lensed quasars and thereby obtain cosmological information. This is described more fully in Sec. 2.4.3.

1.2.3 General Relativity: Gravity as Geometry

1.2.3.1 Why Gravity Is Geometrical. The EEP tells us that the effect of gravity on a body does not depend on its internal properties. In a given gravitational field, a freely falling body will follow a straight line with respect to local freely falling observers. This means that its trajectory depends only on its initial position and velocity. Any body with the same initial position and velocity will follow the identical trajectory.

Einstein reasoned that the trajectories were more a property of the gravitational field itself than of its interaction with specific bodies. Gravity is not a force like electromagnetism or the strong interactions. It is a property of space and time.

It is natural to try to relate gravity to curvature. If the trajectories of freely falling bodies—bodies on which no nongravitational forces are acting—are curved lines, then maybe there is a natural way to associate these lines with the locally straight lines, called *geodesics*, of a curved space. This is exactly what Einstein did, but he had to include the curvature of space *and* time together. Consider a curved surface, such as that of an apple. By following as straight a line as possible, one traces out a curve that meanders over the surface. This curve is determined by one's initial position and direction. It is not determined, however, by the speed with which one travels along the apple. Since gravitational trajectories depend not only upon direction but also upon speed,

they cannot be modeled by the curvature of space alone.

In a curved four-dimensional spacetime, the natural trajectories (the geodesics) will, by analogy, be determined by the initial position and the initial direction of the tangent vector, but not by the length of the tangent. Now, as we saw in our discussion of SR in Sec. 1.2.2.1, the four-velocity is a tangent vector to a particle's world line. Its components do contain the information about the ordinary velocity of the particle. For example, the ratio of the x component to the time component is $(dx/d\tau)/(dt/d\tau) = dx/dt = v^x$. Since any other tangent is a multiple of U , this ratio will be the same for any tangent: it depends only on the direction and not the length of the tangent four-vector. Therefore, the information needed to determine a geodesic in a curved spacetime is exactly that needed to determine a gravitational trajectory: the position and ordinary velocity of the particle.

In this way, the EEP leads to the representation of gravity by the curvature of spacetime.

1.2.3.2 The Metric of Gravity. Curved spaces usually have no *natural* coordinate system. Special coordinates, such as Lorentz observers, do not exist in nonuniform gravitational fields, as we have seen. So we must admit arbitrary spacetime coordinate systems $\{x^\alpha, \alpha=0,1,2,3\}$ and we do not necessarily attach any physical significance to particular coordinate values or intervals. (Another convention here: we use Greek letters for spacetime indices, and the index 0 stands for the coordinate t .) Physical significance attaches only to things that can be measured. Coordinates are usually matters only of definition, not of independent measurement.

The intrinsic geometry of a curved spacetime is completely defined if one gives the distances between points (events, in our case). One only needs local distances, which may be integrated up along curves in the spacetime to give global ones. The distance information is contained in the *metric tensor* $g_{\alpha\beta}$:

$$ds^2 = \sum_{\alpha=0}^3 \sum_{\beta=0}^3 g_{\alpha\beta} dx^\alpha dx^\beta. \quad (16)$$

This equation gives a prescription for calculating the invariant length ds^2 between nearby events separated by coordinate intervals dx^α in terms of the given set of functions $g_{\alpha\beta}$. The

interval is an *invariant*; this property determines how the components $g_{\alpha\beta}$ change when the coordinates change. The term “tensor” refers to the fact that the 4×4 matrix of components must transform this way. (See GEOMETRICAL METHODS for more information on tensors.)

The metric tensor of SR is conventionally called $\eta_{\alpha\beta}$. By comparing Eq. (12) with Eq. (16), we find $\eta_{00} = -c^2$, $\eta_{xx} = \eta_{yy} = \eta_{zz} = +1$, and all other components of $\eta_{\alpha\beta}$ vanish. The spacetime of SR is called *Minkowski spacetime*, and the metric tensor $\eta_{\alpha\beta}$ is called the Minkowski metric.

The metric contains observable information. For example, the proper time on a clock that stays at fixed spatial coordinates ($dx = dy = dz = 0$ on its world line) is related to the coordinate time interval dt by $d\tau^2 = -g_{00}dt^2/c^2$. Given two clocks at different locations \mathcal{A} and \mathcal{B} , then the ratio $g_{00}(\mathcal{A})/g_{00}(\mathcal{B})$ determines the time dilation between them.

We saw in Newtonian gravity that the gravitational time dilation depends on the difference between the potentials at the two points (Sec. 1.2.2.4). Since Newtonian theory only applies when gravity is weak and the metric is close to the Minkowski form, we have that $g_{00} = -(1 + \epsilon)c^2$, where ϵ is small. Then the ratio of the two values of g_{00} depends to first order on the *difference* between the values of ϵ . Therefore, there is a close connection between ϕ and ϵ . If one follows the algebra through, one finds that, for a given Newtonian-type field, the metric needs to have $g_{00} \approx -(1 + 2\phi/c^2)c^2$. This establishes the role of the metric tensor in relativistic gravity: it is the analog of the Newtonian *potential*.

The metric has some simple properties that one can infer from Eq. (16). First, since exchanging dx^α and dx^β in this equation can be accomplished by relabeling the indices, there cannot be any difference between $g_{\alpha\beta}$ and $g_{\beta\alpha}$: the 4×4 matrix of components $g_{\alpha\beta}$ is *symmetric*. This means there are 10 independent components. Second, since one has complete freedom to change the coordinates, there are essentially four degrees of arbitrariness among the 10 components: a given curved spacetime can be described by a variety of metric tensors, all related by coordinate transformations. There are essentially $10 - 4 = 6$ geometrically independent functions that one can choose in order to determine the spacetime.

1.2.3.3 Local Flatness of Spacetime and the Local Freely Falling Frames. The freely falling observers are a special class of coordinate systems that make spacetime look as much like SR as possible in the region of one event. This idea is very similar to the observation that any smooth surface is nearly flat if we look at only a small region. The *local flatness theorem* states that we can choose any point as the origin of a locally flat (or in spacetime, freely falling) coordinate system and adjust the coordinates in such a way that if we make a Taylor expansion of the metric components about this point, the first departure from flatness will be at *second order* in the coordinates. This means that the coordinates can be chosen to make the first derivatives of the metric vanish at any desired point. Since the acceleration of gravity also vanishes in a freely falling frame, the local flatness theorem suggests that *the acceleration of gravity resides in the first derivatives of the metric tensor components* at the point. This is consistent with the relation we saw earlier between the metric and the Newtonian potential.

1.2.3.4 Curvature and Tidal Forces. The *curvature* of the spacetime is a function of the metric tensor. The relation is not immediately straightforward, since even in the flat (Minkowski) spacetime of SR, one can choose a funny coordinate system and make the metric components look very complicated. The local flatness theorem tells us that the curvature is not defined by the first derivatives of the metric, since these can be made to vanish. Therefore, the curvature information must reside in the second derivatives of the metric components. Now, if the acceleration of gravity is given by the first derivatives, then the true gravitational forces—the tidal forces—are given by the gradients of the acceleration of gravity, which are the second derivatives of the metric. Therefore, we conclude that *the tidal forces are represented mathematically by the curvature of spacetime*.

The fundamental measure of curvature is the *Riemann curvature tensor*, denoted by $R^\alpha{}_{\mu\beta\nu}$. It is an object with four indices, essentially because it must be a function of the second derivatives of the metric, which have four indices: $\partial^2 g_{\alpha\beta} / \partial x^\mu \partial x^\nu$. There are a number of symmetries among the indices, so that, among its $4^4 = 256$ components, only 20 are algebraically independent. This is far fewer than the number of independent second de

rivatives of the metric tensor components ($10^2 = 100$) because most of the second derivatives can be adjusted or even set equal to zero by appropriate coordinate choices. The 20 Riemann components represent the irreducible, coordinate-free geometrical information about the curvature. It is a theorem that the spacetime is flat (Minkowskian) if and only if the Riemann tensor vanishes everywhere.

One can derive from this tensor another called the *Ricci curvature tensor* $R^{\alpha\beta}$ by taking a trace. (It turns out that there is only one independent trace of the Riemann tensor, so that the Ricci tensor is essentially unique). The Ricci tensor is a symmetric tensor, with 10 independent components. It has two indices. One can likewise take its trace to get an object with no indices: this is called the *Ricci scalar* R .

A further two-index tensor is the so-called *Einstein tensor* $G^{\alpha\beta}$, which is formed by subtracting one-half of the product of the Ricci scalar and the metric tensor from the Ricci tensor. Its derivatives in a locally flat coordinate system satisfy a remarkable identity, called the *Bianchi identity*: for any metric, the Einstein tensor is divergence-free. (The divergence of a symmetric two-index tensor is defined as the trace of its first derivative, the trace being taken on the derivative index and one of the tensor indices.) This identity is central to Einstein's theory of gravity, as we will see in the next section. Notice also that there is one other divergence-free two-index tensor: the metric tensor itself, whose first derivatives all vanish in a locally flat coordinate system.

1.2.4 Sources of Gravity: How Matter Creates the Geometry

1.2.4.1 The Stress-Energy Tensor. In Newtonian gravity, the source of gravity is simple: the density of mass. It creates the field (the Newtonian potential ϕ) through Newton's field equation, Eq. (11). Because of the interconvertibility of mass and energy, we would expect the density of total mass-energy to be the source. This is, however, incompatible with SR for two reasons. First, in SR, mass is a form of energy, and energy is not a scalar: it is only one component of the energy-momentum four-vector. Second, in a relativistic theory, the notion of a density is not frame independent. The Lorentz contraction

means that volumes and therefore densities depend on the observer.

These two transformations of mass-energy density (one acting on *mass-energy*, the other on *density*) mean that the mass-energy density is actually only one component of a two-index tensor. The other components of this tensor include the momentum density and the *stress tensor*. Just as SR links momentum and energy in a single four-vector, so too does it link stress and energy-momentum density into a single four-tensor with two indices. The unified tensor is called the *stress-energy tensor*. It is denoted by $T^{\alpha\beta}$. The mass density is measured by T^{tt} , the density of x -momentum by T^{tx} , and the stresses by the spatial components.

Energy and momentum are conserved in ordinary physics, and by the EEP this must be true in a local freely falling frame. For energy, for example, the conservation law says that the rate of change of energy in any volume equals the total net rate at which energy enters the volume across its sides. This implies that the (four-)divergence of the stress-energy tensor vanishes.

If mass density creates gravity in Newton's theory, then it is natural to expect the stress-energy tensor to be the source for a relativistic theory of gravity. A simple counting argument reinforces this idea. The stress-energy tensor can be shown to be symmetric on its two indices, so it has 10 independent components. This matches the 10 components of the metric tensor that describe the gravitational field: 10 unknowns $g_{\alpha\beta}$ determined by the 10 sources $T^{\alpha\beta}$.

However, this simple argument is too hasty: we saw above that there are effectively only six components of the metric that can be determined by the physics, since there are four degrees of coordinate freedom. Would such a theory be overdetermined? The answer is no, because there are as well four identities among the components of the stress-energy tensor: the laws of conservation of energy and momentum. So the counting is indeed encouraging: there are as many independent "sources" of gravity as there are "fields."

1.2.4.2 Einstein's Field Equations. To turn the counting argument into a full theory of gravity, Einstein had to find the appropriate differential equations that would allow the fields to be found in terms of the sources. He made what seemed to be the simplest choice.

We shall see below that there are more complicated alternatives.

One wants a generalization of Eq. (11), which has the form of a differential equation involving second derivatives of the potential set equal to the mass density as a source. The analog of the potential in a geometrical theory is, as we have seen, the metric tensor. So we need an equation involving second derivatives of the metric. The natural tensorial object involving second derivatives is the curvature tensor. If we take the hint of the preceding section and assume that the source will be the stress-energy tensor, then what we want is to equate the stress-energy tensor to a two-index curvature tensor that, like the stress-energy tensor, is divergenceless. As we pointed out earlier, there is such a tensor: the Einstein tensor. Einstein's choice was, then, to adopt the following form of the field equations:

Einstein tensor = $\kappa \times$ stress-energy tensor ,

where κ is a constant of proportionality that is not fixed by the argument so far.

We fix κ by demanding that the solution of the field equations when the density is low and the velocities of the sources are small should have geodesics that are the paths of freely falling particles in the Newtonian gravitational field of the same sources. This gives $\kappa = 8\pi G/c^4$. In mathematical notation, the field equations of GR are

$$G^{\alpha\beta} = (8\pi G/c^4) T^{\alpha\beta} . \quad (17)$$

These equations are determined only by the correspondence with Newtonian gravity: there are no free parameters. Therefore, experimental constraints on GR can be very strict. One new consequence of Einstein's field equations after another has been calculated: the perihelion shift of Mercury's orbit (Sec. 2.4.4); the gravitational deflection of light (Sec. 2.4.3); the energy radiated in gravitational waves by a binary star system (Sec. 2.4.4); the existence of black holes (Sec. 2.4.2); the Big Bang and its consequences (Sec. 2.4.5). Each one could have proved the undoing of the theory. Yet in each case, there has been excellent quantitative agreement with the predictions of GR.

Because the Einstein equations are exceedingly nonlinear in the metric tensor, it has not been easy to find solutions of the field equations. Some exact solutions are known, which

fortunately correspond to some of the most interesting simple cases: the field outside an isolated neutron star, the field of a black hole, and the metric of the homogeneous universe as a whole. In other cases, approximations can be made to good effect, such as when looking for small corrections to Newtonian gravity or for gravitational waves. And increasingly, as computers improve, relativists are attempting to solve more complicated systems numerically.

1.2.5 Other Theories of Gravity. General relativity is not the only geometrical theory of gravity possible. By the Einstein equivalence principle (EEP), the spacetime metric is the only field that directly affects geodesics or the behavior of stress-energy. Nevertheless, there could be other gravitational fields in the universe. Most alternative metric theories introduce such auxiliary fields; they could be scalar fields, vector fields, tensor fields, and so on. These fields may mediate the manner in which matter generates the spacetime metric, but they do not act back directly on the matter. Theories of gravity in which auxiliary fields act directly on matter are called *nonmetric* theories and typically violate the EEP. Experimental tests of that principle, described in Sec. 3.1, place tight constraints on such theories.

The prototypical example of an alternative metric theory of gravity is the Brans-Dicke theory, in which a scalar field exists in addition to the metric. This field plays the role of allowing the gravitational "constant" to vary in space and time. The differences in the predictions for observable effects between Brans-Dicke theory and GR depend on how strongly coupled the scalar field is to the metric—the weaker the coupling, the smaller the differences.

When one focuses on the weak-field, slow-motion regime that is appropriate to the Solar System (the so-called post-Newtonian limit because it includes the first corrections to Newtonian gravity) one finds that, in a broad class of metric theories, the metric looks the same, except for the numerical values of coefficients in front of various terms. A framework for studying metric theories in general has been developed, called the parametrized post-Newtonian (PPN) framework, in which parameters take the place of the numerical coefficients, parameters whose values depend

on the theory under study. We shall encounter the use of some of these parameters in Sec. 3.2.1.

1.2.6 A Cosmological Term in Einstein's Equations. When we derived the form of Einstein's equations, Eq. (17) above, we reasoned that we needed a curvature tensor that had two indices and was divergenceless, and so we followed Einstein and used the Einstein tensor. But we remarked in Sec. 1.2.3.4 that there is another divergenceless tensor, the metric tensor itself. So it is possible to amend Einstein's equations to the following form:

$$\begin{aligned} &\text{Einstein tensor} + \Lambda \times \text{metric tensor} \\ &= \kappa \times \text{stress-energy tensor}, \end{aligned} \quad (18)$$

which satisfies all the properties we needed for consistency and which will reduce to Newton's field equation in the appropriate limit provided the constant Λ is sufficiently small. Einstein introduced this modification into his theory when he found that his original equations did not admit static cosmologies, because the astronomical prejudices of his day held that the Universe was static and unchanging. Einstein found that he could choose a value of Λ that would indeed allow a static cosmological solution. He called Λ the *cosmological constant*.

There was no experimental evidence for Λ in Einstein's time, and when the expansion of the Universe was established a decade or so later, Einstein disavowed the cosmological term, calling it one of his biggest blunders. This is understandable: had he stuck to the original form of the equations, he would have predicted the expansion of the Universe and the associated Big Bang long before there was astronomical evidence for it.

Today there is a different view of Λ . If we carry the cosmological term over to the right-hand side of the equation, it then plays the role of an effective stress-energy tensor of the form

$$T_{\alpha\beta}^{\text{cosmol}} = -\Lambda g_{\alpha\beta}. \quad (19)$$

Because $g_{tt} < 0$, this means that Λ is an effective energy density, constant everywhere. Provided Λ is much smaller than the mean energy density of the solar system, its effects will not have been noticeable. The values discussed today are much smaller. Interestingly, since the stresses are given by the

spatial parts of Eq. (19), we find that the effective pressure is $-\Lambda$: if the energy density is positive, the pressure is negative.

Strange as this may look for ordinary matter, it seems almost inevitable in some schemes for quantizing gravity. When a classical field is quantized, there are renormalization terms that affect the residual energy density of the field (see Birrell and Davies, 1982). If one quantizes the vacuum gravitational field, the stress-energy tensor that survives renormalization will have to be proportional to the metric tensor, since there are no other tensors around, and no preferred frames. (This effect has an electromagnetic analog, called the *Casimir effect*.) It could happen, of course, that the cosmological term turns out to be zero; but there is no reason for this to be inevitable, and there is an expectation among many cosmologists today that the term may be comparable to the smoothed-out energy density of the present Universe.

2. SOME CONSEQUENCES OF EINSTEIN'S FIELD EQUATIONS

2.1 Momentum and Stress Also Make Gravity

The fact that all components of the stress-energy tensor contribute to gravity has important consequences.

2.1.1 Gravitomagnetism. In the Solar System, Einstein's equations will give the Newtonian potential in g_{tt} , using the mass density (tt component of the stress-energy tensor) as the source (Sec. 1.2.3.2). But the momentum density is also a source, so that *moving* bodies create new parts of the gravitational field. These couple to the velocities of particles, so they are called *gravitomagnetic effects*.

An important example is rotation. The angular momentum density of the Earth generates a metric term whose value is proportional to the angular momentum and which falls off as $1/r^2$, just as the electromagnetic potential of a rotating conductor does.

This term has a number of consequences. A free particle falling straight down will acquire a small motion in the same sense as the Earth's rotation. This is sometimes called the "dragging of inertial frames" or the *Lense-Thirring effect*, after its discoverers. For the

same reason, the orbital period of a satellite orbiting from west to east is slightly less than that of one going in the opposite direction. A spinning gyroscope in orbit will couple to this effect in much the same way that the spin of an electron in an atom couples to the magnetic moment of the nucleus, producing a precession of the spin.

These effects are weak compared to the dominant Newtonian orbital effects and hence hard to measure, for two reasons. First, the angular momentum density is small compared to the mass density, basically by a factor of v/c , which for the Earth's rotational motion is about 1.5×10^{-6} ; the gravitomagnetic metric terms are therefore smaller than ϕ by this factor. But the effect is smaller still because, like magnetic effects, it only couples to the speed of the orbiting body, not simply to its mass. This introduces yet another factor of v/c . Such effects that are predicted by Einstein's equations but are of order $(v/c)^2$ smaller than the Newtonian effects are called *post-Newtonian* effects. Measuring them is one way of testing the validity of GR (Sec. 3.4.1).

2.1.2 Gravitational Collapse. In the same way, the stresses inside a body also make a contribution to gravity. These have an even smaller effect in nonrelativistic bodies like the Earth, being down by order $(v/c)^4$. But in a relativistic star, like the neutron stars we will describe in Sec. 2.4.1 below, they can be important. The dominant stress inside a star is the ordinary pressure of the fluid. In a neutron star, pressure and density are comparable, in the sense that $p/c^2 \sim \rho$. Such a large pressure has the effect of *increasing* the effective gravitational field inside the star. The result in GR is that there is actually a limit on how compact and relativistic a star can be, regardless of what it is made of: no star of mass M and radius R can have $GM/Rc^2 > 2.25$.

This leads to a very interesting conclusion. Suppose a star is right at the limiting size, and then it is perturbed, to make its radius marginally smaller. It cannot restore itself to equilibrium, and so it will collapse. Once it begins to collapse, it cannot stabilize itself at a smaller radius, for this would violate the inequality. This is called *gravitational collapse*: the unstoppable collapse of a star. The result of such collapse is a *black hole*.

2.2 Black-Hole Theory

2.2.1 Black Holes in the Eighteenth Century. Probably the most dramatic prediction of Einstein's equations is the existence of black holes. It may seem surprising, therefore, that essentially the same concept was discussed in the context of Newtonian gravity two centuries ago!

Black holes are regions of space in which gravity is so strong that light is trapped. The boundary of this region is the *horizon*: no light emitted from inside the horizon can reach the outside. The idea that light could be trapped by gravity occurred to eighteenth-century physicists as well. The first to suggest it was the amateur scientist John Michell (1724–1793), who in 1784 reasoned that, since light traveled at a finite speed, it would in principle be possible for an astronomical body to exist whose escape velocity was greater than the speed of light. Similar conclusions were drawn by Pierre Laplace in 1796.

Since the escape velocity is $(2GM/R)^{1/2}$, it follows that a body will trap light in Newtonian gravity if $2GM/Rc^2 > 1$. This is, remarkably, exactly the condition for the formation of a black hole in Einstein's theory as well. [Michell's suggestion was no fluke of speculation: it was Michell who suggested to his friend Henry Cavendish (1731–1810) that Cavendish should perform the famous experiment to measure the value of G that now bears his name.]

In the Newtonian picture, the trapping star is dark, because light cannot reach us from it, but its gravitational field is unchanged, and it can be discovered and identified by its gravitational effects. The same is true for the relativistic black hole.

2.2.2 Black Holes in General Relativity. The modifications needed to turn the Newtonian picture into a black hole in GR arise because of the special role played by the speed of light in relativity. In the Newtonian picture, light would still leave the surface of the trapping body, but it would slow down progressively and eventually turn around and fall back to the surface. Light emitted somewhat further out would escape, passing at some point the previous photon where it turns around. In relativity, all photons travel at the same speed. If a photon leaves the surface of the body, it cannot turn around later while

other photons pass it, still moving outwards. Instead, there is a *horizon*: any photons outside the horizon can escape to infinity, while those inside are trapped.

Moreover, in the Newtonian picture, an interested astronaut could still in principle explore the dark star, getting away on a sufficiently powerful rocket. In relativity, nothing can go faster than light, so anyone or anything that crosses inside the horizon is trapped there forever.

There are no markers at the horizon, no signs saying "Point of no return". The horizon is defined as the boundary between what is trapped and what is not, but this boundary may not be known until the whole dynamics of the collapse that formed the black hole has occurred. The horizon is generally a place in the vacuum gravitational field, locally flat, in every way locally unremarkable.

2.2.3 Singularities Inside the Hole. We saw in Sec. 2.1.2 that a star with a radius smaller than $9/8$ of the horizon size of a black hole of the same mass as the star cannot support itself against gravity. This means that the mass that has collapsed to form the black hole cannot find an equilibrium radius inside the hole: the material must continue to move inwards. The details of what happens to it are conjectural at present, but there is a set of *singularity theorems*, proved within GR by the mathematical physicists Roger Penrose and Stephen Hawking (see Hawking and Ellis, 1973), that show that matter obeying rather weak conditions (such as that its energy density should be positive) will always generate some sort of singularity inside the horizon. The nature of this singularity is, in general, poorly understood. Quantizing gravity (or even the matter fields) may change or eliminate the singularity, but it is hard to see how quantum effects could get outside the horizon and change the external aspect of the hole of macroscopic size.

It is also not clear whether gravitational collapse can lead to a singularity that is not inside the horizon. Such an event would prove much more damaging to notions of causality: singularities inside the horizon are confined, but those outside might make unpredictable effects on their surroundings. Such *naked singularities* are an active subject of research today. Penrose has formulated the *cosmic censorship hypothesis*, which asserts that na-

ked singularities do not form. There is no proof of it at present.

2.2.4 Black Holes Have No Hair. Although the inside of a black hole may be messy, the external gravitational field is extraordinarily simple. A series of "no hair" theorems by Hawking, Brandon Carter, and others (see Wald, 1984) has shown that black holes, once formed, quickly radiate away everything possible and settle into a stationary state characterized entirely by values of the conventional conserved quantities of physics: total energy (the mass of the hole), total angular momentum, electric charge, and magnetic monopole moment (should monopoles exist). These properties of the hole are all measurable externally: the mass from orbital periods of satellites, the angular momentum from the Lense-Thirring effect (Sec. 2.1.1), and the charge from electrostatic attraction.

Higher multiple moments of the electromagnetic or gravitational field that might have characterized the collapsing star are carried away from the hole in a burst of radiation that accompanies its formation. The total number of baryons and leptons that formed the hole is not measurable from outside, because the nuclear forces have short range.

The metric of the spherical uncharged black hole was the first exact solution found for Einstein's equations, derived by Karl Schwarzschild (1873–1916) in 1916. It was not completely understood until the 1960s, after work by Martin Kruskal, Peter Szekeres, John Wheeler, and others. (See Misner *et al.*, 1973, for references.) The rotating hole is described by the Kerr metric. The term "black hole" was coined by Wheeler in the 1960s.

2.2.5 Black-Hole Thermodynamics. Forming a black hole entails a considerable loss of information: by the "no hair" theorems, most of what goes down the hole is lost forever. It is not surprising, then, that one can assign an *entropy* to the black hole. In fact, the whole of thermodynamics can be generalized to include black holes, and *black-hole thermodynamics* is one of the most beautiful parts of black hole theory (see Thorne *et al.*, 1986). The entropy of the hole is proportional to the surface area of its horizon. A theorem of Hawking establishes that this area cannot decrease with time (Hawking and Ellis, 1973).

There is a corresponding temperature, which is proportional to the surface gravity of the hole.

This temperature is the most remarkable black-hole property of all. In black-hole thermodynamics, it plays a formal role in determining the energy (mass) change associated with a change in the entropy (surface area) of a black hole: $dM = TdS$. But in ordinary thermodynamics, objects also exhibit their temperature by emitting blackbody radiation. Despite the fact that black holes trap all photons, they also give off a *thermal radiation* whose spectrum is exactly that of a blackbody of the thermodynamic temperature. This is called the *Hawking radiation* (Hawking, 1974).

Hawking's demonstration of this radiation relies on quantum mechanics, but it requires only the quantum field theory of the electromagnetic field near the hole, not any quantum theory of gravity. The Hawking temperature is inversely proportional to the mass of the hole, and for a hole that has the mass of the Sun, it is entirely negligible. This means that the radiation is likely to be exceedingly difficult to observe. Its existence is accepted, however, partly because well-established techniques of quantum field theory predict it, and partly because it allows black-hole thermodynamics to fit so beautifully into standard thermodynamics.

Black holes have other macroscopic properties as well (Thorne *et al.*, 1986). For example, the horizon has a finite electrical conductivity. This comes essentially from the "no hair" theorems as well: if a charge is lowered near the horizon, it will have to look from a distance as if it is smeared uniformly over the horizon; otherwise there would be an electric dipole moment. The horizon therefore effectively conducts charge along its surface.

2.2.6 Wormholes. The simplest black-hole solution, the one found by Schwarzschild, describes a stationary black hole, i.e., one that has existed for all time. It has an interesting topology: it actually involves *two* "exterior" spaces connected by a throat, or wormhole. Although the gravitational field outside the hole in both exterior regions is static, the throat is not. It opens up to a maximum size and then closes off in a finite proper time. In fact, it closes off so fast that a particle falling into the hole would not have time to get through it: the wormhole is not a

channel for communication between the two exteriors. Moreover, a black hole formed by the collapse of a star has a simpler topology, with no wormhole, and only one exterior region.

Nevertheless, wormholes have intriguing properties. If physicists could manipulate quantum fields to create and maintain a region of negative energy (as happens in the Casimir effect referred to in Sec. 1.2.6), then they could in principle maintain a wormhole open long enough for a particle to travel through it. If the two ends of the wormhole are open in the same space, then a particle could circle through the throat many times. If, in addition, the ends of the wormhole were made to move relative to one another, then, as Thorne (1991) has shown, the time dilation effect of special relativity could be arranged to allow the particle to travel backwards in time.

The realization that GR allows such behavior, even under what present technology suggests are implausible conditions, has created considerable interest. If such *closed timelike curves* exist, they raise questions about causality and especially the consistency of initial-value formulations of the law of physics, and they have to be taken into account in formulations of quantum gravity. The study of these rather exotic conditions may in the future shed new light on fundamental physics.

2.3 Gravitational Waves

2.3.1 The Necessity of Gravitational Waves.

Gravitational waves (GWs) are present in any relativistic theory of gravity. When fields are weak and motions slow, the relativistic equations must reduce to Newton's field equation, Eq. (11). This involves the differential operator ∇^2 . If fields are weak but motions may be rapid, then one expects the theory to be relativistically invariant. The operator ∇^2 will be replaced by its relativistic generalization

$$-\frac{1}{c^2} \frac{\partial^2}{\partial t^2} + \nabla^2.$$

This is the wave operator, with the characteristic wave speed c . We should expect wave effects in relativistic gravity, propagating at the speed of light. Different theories of gravity will differ in the details of the waves, but GWs will be present in all of them.

In GR, the waves bear a striking mathematical resemblance to electromagnetic waves. They are transverse, have two independent polarizations, and fall off in amplitude as $1/r$. Electromagnetic waves do not carry monopole radiation: for slow-motion sources they are dominated by dipole emission. Gravitational waves in Einstein's theory do not carry either monopole or dipole radiation, but are dominated by quadrupole emission (Sec. 2.3.3).

2.3.2 The Interaction of Gravitational Waves with Matter. Waves can only be detected through their time-dependent tidal forces. If a wave passes a single particle, one can always choose comoving coordinates, so that the particle remains at a fixed coordinate position. By the EEP, the particle will not feel anything locally from the wave: bowls of soup will not spill, pocket calculators will work normally. The particle will only be able to detect the wave if it looks at other things sufficiently far away for the tidal effects of gravity to be noticeable.

The absence of single-particle effects of GWs led to much misunderstanding in the early development of GR. The intuition developed in electromagnetism, where single charged particles can detect electromagnetic waves, was not helpful. Many relativists, including at times Einstein himself, believed that GWs were a mathematical illusion. Work

in the 1950s and 1960s by Hermann Bondi, Richard Isaacson, Penrose, Joseph Weber, and many others showed that to detect GWs one needs at least two particles, but that the waves are very real, transferring energy and angular momentum away from their sources to their detectors. [See Misner *et al.* (1973) or Schutz (1985) for more details.]

The simplest way to detect a GW is to monitor the proper distance between two nearby free particles, particles that have no external forces on them. The distance can be monitored, for example, by measuring the round-trip light-travel time between the particles. A GW can affect this distance. Gravitational waves are transverse, and so if the wave moves in the z direction, then only distances in the x - y plane will be affected.

The pattern of distance changes produced by a given wave is shown in Fig. 1. The dots represent a ring of free particles. The circles and ellipses are not physical connections: they only illustrate the pattern of the placement of the particles. The first ring is the undisturbed position: a circular ring of particles. The second diagram shows the effect of a GW on proper distances. It has lengthened the x axis of the circle and shortened the y axis by the same fraction, preserving the area of the ring. When the wave reaches its opposite phase (final diagram), the effects on the axes are reversed. At the bottom of the diagram are

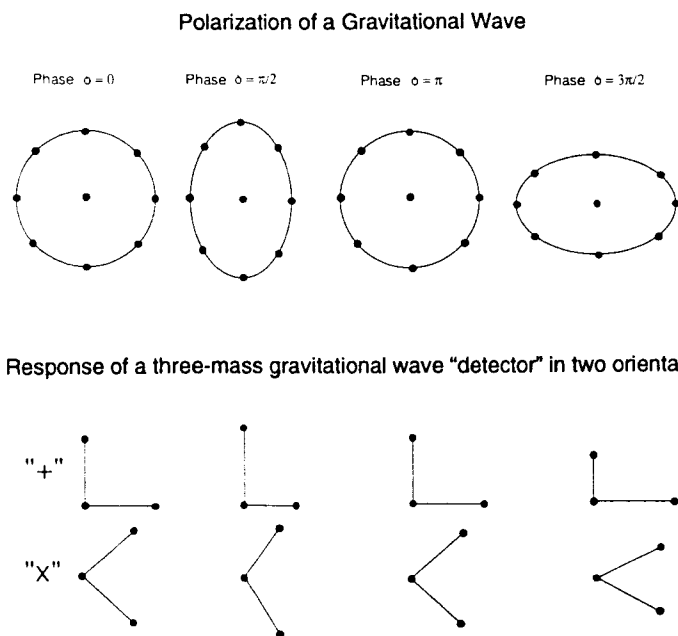


FIG. 1. Illustration of the action of a GW on a ring of free particles transverse to the direction of propagation.

shown the lines joining certain particles to the center of the ring. The top lines are for the particles at 12 o'clock and 3 o'clock (particles oriented in the "+" configuration relative to the ellipse's principal axes). Their lines remain perpendicular but change their relative length. The bottom lines are for the particles at 1:30 and 4:30 (particles oriented in the "x" orientation relative to the ellipse's principal axes). Their lines remain the same length but the angle between them changes.

Nothing is preferred about the x axis in this diagram. The entire ellipse can be rotated about the z axis to give possible patterns. A rotation of 90° effectively carries the pattern back into itself, which is characteristic of a tensor (spin-2) wave. The second independent polarization is obtained by a rotation of 45° .

Because tidal effects grow linearly with distance, *the shape of the ellipse in Fig. 1 is a property only of the wave, independent of the size of the ring.* The relative change in the axes is, therefore, a measure of the amplitude of the wave itself. It is conventionally called h , and can be defined from Fig. 1 as

$$h := 2(\delta l)_{\max}/l, \quad (20)$$

where l is the length of the semimajor axis of the ellipse. (The factor of 2 is conventional.) The relative distortions shown in the figure are, of course, exaggerated in order to make them easy to see. Modern GW detectors are being built to detect amplitudes as small as 10^{-22} . We shall consider the technical challenges of this in Sec. 4.2.

This tiny amplitude means that the effect of matter on the gravitational wave will also be of this order; the wave will lose or scatter only a fraction of order 10^{-22} of its energy as it passes through a detector. By extension, this means that the waves that arrive at our detectors have lost little of their original form: they carry information from their sources uncorrupted by scattering or absorption. In this they are a unique carrier of information in astronomy; even the neutrinos detected from the supernova event in 1987 (SN1987A) were from a thermal distribution that had scattered many times before leaving the collapsed core (Sec. 2.4.1.4). Traveling at the speed of light, gravitational waves follow geodesics through the Universe; they can be gravitationally lensed, but scattering and absorption are negligible.

2.3.3 Wave Emission: The Quadrupole Formula. Sources of gravitational waves respect the conservation laws for energy and momentum that are built into Einstein's equations. For weak GWs, these have a similar effect to that produced by the conservation of charge in electromagnetism, which ensures that there are no monopole electromagnetic waves: the lowest order of radiation is dipole. In GR, the law of conservation of energy similarly ensures there is no spherical gravitational radiation. Any oscillating spherical mass leaves its exterior gravitational field undisturbed in GR (just as in Newtonian gravity), and so it does not radiate GWs. Conservation of momentum has the same effect on dipole radiation: there is no dipole gravitational radiation.

From a fundamental point of view, charge conservation (and its consequence, the absence of monopole radiation) in electromagnetism follows from the gauge invariance of the theory. (See SYMMETRY AND CONSERVATION LAWS.) In GR, energy-momentum conservation (and the absence of monopole and dipole radiation) follow from the general coordinate invariance of the theory.

At quadrupole order, the formulas for the radiation emitted by a slow-motion system are remarkably similar to those for electric quadrupole radiation in electromagnetism. (For highly relativistic, fast-motion systems, the formulas below can only be used approximately.) We define the quadrupole moment tensor of the mass distribution:

$$I_{jk}(t) = \int \rho(t, \mathbf{r}) x_j x_k d^3x, \quad (21)$$

where ρ is the mass density (rest mass dominates in a nonrelativistic system), and where we follow the usual convention of employing Latin indices for purely spatial values (1,2,3). This equation can only be used in a Cartesian spatial coordinate system. The trace-free or reduced quadrupole moment is derived from this by

$$\mathcal{I}_{jk} := I_{jk} - \frac{1}{3} \delta_{jk} I, \quad (22)$$

where I is the trace of I_{jk} .

The amplitude of the radiation at a distance r from the source is roughly

$$h \sim (4G/c^4) (\ddot{\mathcal{I}}/r), \quad (23)$$

where \ddot{x} stands for the second time derivative of a typical component of the quadrupole moment. If the motions that are responsible for the time derivatives are driven by the internal gravitational fields of the source, which is the case for all the realistic sources of GWs that we will consider later, then there is a convenient upper bound on this amplitude, given by

$$h \leq (GM/rc^4)\phi_{\text{internal}}, \quad (24)$$

where ϕ_{internal} is the typical size of the Newtonian gravitational potential within the source (Schutz, 1984).

Although the local conservation laws of energy and momentum eliminate the monopole and dipole radiation, it does not follow that GWs remove no energy or angular momentum from the source. Unlike electromagnetism, which is a linear theory, GR is nonlinear, and GWs can act as sources for gravity. So when they leave the system, there is a gradual decrease in its energy. This is compensated by a well-defined energy flux carried by the waves (Isaacson, 1968). Energy from this flux can be transferred to other systems, such as GW detectors. The GW luminosity (Landau and Lifshitz, 1962) is probably the most useful of the quadrupole formulas:

$$L_{\text{GW}} = \frac{G}{5c^5} \left\langle \sum_{j=1}^3 \sum_{k=1}^3 |\ddot{x}_{jk}|^2 \right\rangle, \quad (25)$$

where the angle brackets denote an average over one period of the motion of the source. Notice that this depends on the square of a third time derivative, just as the analogous formula for electric quadrupole radiation does. It is therefore very sensitive to the size of the velocities inside the source. As a source becomes more relativistic, the power radiated goes up very rapidly, as $(v/c)^6$.

2.4 Applications of General Relativity

For its first 50 years, GR was mainly the property of mathematicians and theoretical physicists who valued it for its intellectual challenge and its aesthetic beauty. It seemed to have little practical relevance to the rest of physics. But astounding discoveries in astronomy, beginning with the realization in 1963 that quasars have enormous luminosities, attracted physicists to the challenge of explaining the new phenomena, and they found that

they needed to use GR as an everyday tool in making models of astronomical phenomena. In this section we describe some of the ways that GR is used. For more details of many of these phenomena, the reader is referred to the article on ASTROPHYSICS.

2.4.1 Relativistic Stars (Pulsars) and Gravitational Collapse

2.4.1.1 Stellar Evolution. An ordinary star nearing the end of its nuclear-burning lifetime undergoes many changes as it switches from its primary fuel (hydrogen) to short-lived secondary sources of energy (helium, then carbon, and so on). When it finally exhausts its nuclear fuel, it can no longer generate energy to replace that which it radiates from its surface, with the result that gas pressure can no longer hold the star up against gravity. The star then has only two options: it can rely on *degeneracy pressure* (the quantum-mechanical resistance of identical fermions to being squeezed into too small a volume), or it can collapse to a black hole. Which option it takes depends on its mass.

2.4.1.2 White Dwarfs. If the mass of the star when it reaches this point is less than about 1.4 times the mass of the Sun, then it can end its days quietly, supported by electron degeneracy pressure. It is called a *white dwarf*, and has a radius roughly that of the Earth, about 10^4 km.

However, Chandrasekhar (1939) showed that electron degeneracy cannot support more than about $1.4M_{\odot}$. The burnt-out core of a star that is larger than this will inevitably grow to exceed this size, and then it must gravitationally collapse. During the collapse, which lasts less than a tenth of a second, essentially all electrons and protons combine to form neutrons. This process is called *neutronization*. When the density is high enough for neutron degeneracy pressure to become effective, the collapse will halt, provided again that the star is not too massive. This *neutron star* has the density of a typical atomic nucleus, so that a $1M_{\odot}$ star has a radius of about 10 km.

2.4.1.3 Neutron Stars. We have met neutron stars (NSs) in Sec. 1.1.2.1. The upper mass limit of NSs is sensitive to poorly understood high-density nuclear physics, and to the details of the relativistic theory of gravity. It is probably between $2M_{\odot}$ and $3M_{\odot}$ in GR (Hartle, 1978).

The result of halting the collapse is a bounce; a shock travels outwards through the rest of the envelope of the star, blowing it away in an explosion called a *supernova*. Since the collapsing core originally had a rest mass of $1.4M_{\odot}$, one expects the supernova to leave behind a NS of about that rest mass or a little more. Because the gravitational redshift reduces all energies as measured from far away, and because the strong equivalence principle (Sec. 3.2.3) ensures that all energies contribute to gravity, it follows that the gravitational mass of the star as measured by orbits in its gravitational field will be some 10% or so less than its rest mass.

2.4.1.4 Black Holes. Collapse to a NS is likely to be complicated by a number of factors. One is rotation. If the collapsing core has as much angular momentum as the Sun (which is a slow rotator), centrifugal effects will halt the collapse before neutron-star densities. Smaller amounts of rotation may therefore have significant dynamical effects. Work is currently in progress to model rotating collapse on computers.

Another complication is neutrino physics. The collapse generates many neutrinos, first from the neutronization reaction $p + e \rightarrow n + \nu_e$, but more copiously from the thermal equilibrium reaction $\gamma + \gamma \leftrightarrow \nu + \bar{\nu}$ that develops as the collapse halts. These appear to exert a crucial pressure to power the shock outwards. The neutrinos detected from the nearby supernova event in 1987 (called SN1987A—see ASTROPHYSICS) came primarily from the thermal distribution.

Whatever the complications, it seems that on occasion the shock does not blow away the whole star, and enough further matter falls on the collapsed core to push it over the upper mass limit. It must then collapse further, and it forms a black hole (BH). In the next section we will see that there is good evidence for a number of $10M_{\odot}$ BHs in our Galaxy.

2.4.1.5 Pulsars. When a NS is formed, it is likely to be rotating. By what seems like a great stroke of luck, we are able to observe this rotation in hundreds of cases. During the collapse, not only is the rotation speed amplified, but so is the magnetic field of the star. It happens that the magnetic and rotation axes do not line up; in fact, they seem to prefer to be perpendicular to each other. The result is that rotation carries the magnetic poles around. Charged particles spiral around the magnetic

field lines and crash into the star at its magnetic poles, in a colossal version of our terrestrial auroral displays. The result is a beam of radiation that emerges from the poles and is spun by the star's rotation in the same way a lighthouse beam turns. If we happen to sit in the direction of the beam, we see the radiation (radio, optical, and x-ray) turning on and off. This is a *pulsar*.

Pulsar periods must be longer than the orbital period at the surface of a NS, which is about 0.5–1 ms. In fact, the fastest known pulsar has a 1.6-ms period. However, all pulsars that we know are young (for example, those that can be identified with supernova explosions whose remnant clouds of gas are still visible) are rather slow rotators. For example, the pulsar in the Crab nebula spins “only” 30 times per second. We now believe that this is typical of new NSs, and that the really rapid rotators have been spun up by accreting gas from a companion star in a binary system (Sec. 2.4.2).

Pulsars in binaries are not uncommon, but they are hard to find because the orbital motion keeps changing their apparent pulse period. However, binary pulsars are prized discoveries, because by monitoring the Doppler shifts of the period one gets dynamical information about the system. In a few cases, we get enough information to determine the masses of the individual stars, as we describe in Sec. 3.3. Interestingly, all six neutron stars with well-determined masses are within 10% of $1.4M_{\odot}$.

2.4.2 Black Holes in X-Ray Binaries and in Galactic Centers

2.4.2.1 X-Ray Binaries. When gravity overwhelms neutron degeneracy, the star must collapse. Even though high-density nuclear physics is poorly understood, simple constraints (such as that the speed of sound in nuclear matter must be less than the speed of light) are enough to guarantee that nuclear physics cannot support more than about $3M_{\odot}$ at these densities (Hartle, 1978). This implies that a compact object with a significantly larger mass, such as $7M_{\odot}$ or $10M_{\odot}$, must be a BH. Since BHs do not emit observable radiation themselves, practically the only way we can identify holes of this size is by their effects in binary systems.

When NSs or BHs orbit ordinary stars, mass can sometimes flow from the compan-

ion onto the compact star. The angular momentum of the infalling gas forces it into a disk around the star, and friction in the disk leads to a more or less steady accretion of matter onto the compact object. As we noted above, NSs can be spun up by this process. Black holes also spin up, but this may not be observable. But the hot accretion disk is itself a source of observable radiation: x rays.

One of the big surprises of the early 1970s was the discovery, by x-ray satellite observatories, that our Galaxy has dozens of such binary x-ray sources. If the companion can be identified by optical observations and its mass estimated from its spectrum, then the mass of the compact object can be estimated. In this way, we now have three or four fairly secure identifications of BHs. The best identification, in a system called V404-Cyg, has a minimum mass of $8M_{\odot}$ (Casares *et al.*, 1992).

This certainly underestimates the number of BHs in the Galaxy. Most mass determinations are uncertain by 50% or so, and, given the uncertainty in the upper mass limit of NSs, there are a number of binaries where we are unable to be sure about the nature of the compact object. Therefore, while we are confident that about 10% of the stars in the Galaxy are white dwarfs and 1% are NSs, the fraction that are BHs is very uncertain. X-ray binary observations suggest a plausible lower limit of about one in 10^4 .

2.4.2.2 Galactic Black Holes. Giant BHs seem to be relatively more common than this: it is possible that every normal galaxy has in its center a BH of mass between 10^4M_{\odot} and $10^{10}M_{\odot}$. These would probably have been formed at about the same time as the galaxy itself formed, either from a dense gas cloud in the center or from the collective collapse of a swarm of NSs and small BHs formed near the center. Recall that the density of a 10^8M_{\odot} cloud of gas when it forms a BH is only the density of water.

For general information about galaxies, the reader is referred to the article on GALAXIES AND COSMOLOGY. The evidence for BHs in the centers of galaxies is manifold, and in many ways better than that for individual x-ray binary BHs. The strongest evidence is velocity information. By measuring the Doppler shifts of spectral features in many wave bands, astronomers have accumulated evidence for rotation, collapse, and/or expansion with velocities of typically hundreds of

km/s inside volumes no larger than a parsec (3×10^{16} m, which is a typical distance between individual stars in the neighborhood of the Sun). Using the order-of-magnitude relation $GM/R = v^2$ gives masses of the order we have quoted (Rees, 1990). There do not seem to be any mechanisms that would allow such concentrations of mass to remain stable for long without collapsing to BHs. The inference is that the collapse has already occurred in most cases, and we are seeing a small amount of gas swirling around in the resulting gravitational field.

Another line of evidence also points to BHs: the jets of relativistic particles that are expelled from the centers of galaxies. Some jets are narrow (opening angles of less than 1°) and run straight for millions of light years, which argues that their emission has been constant over millions of years. Jets are often found in opposing directions. The phenomenon is ubiquitous and at the same time highly variable. Some galaxies seem quite ordinary except for the immense radio emission that takes place at the ends of the jets, far outside the galaxies themselves. Other galaxies contain *quasars*, which are very small sources of immense luminosity, from which emerge jets (Blandford *et al.*, 1982). Quasars in particular seem to be associated with young galaxies: a far larger fraction of galaxies exhibited the quasar phenomenon when the Universe was $\frac{1}{3}$ of its present age than today.

These facts fit the BH model too, chiefly because only BHs seem to be able to form stable centers of activity, to supply sufficient energy, and (through their rotation) to provide a consistent direction for jets. Gas probably forms an accretion disk around a massive BH, just as it does in binary systems. Something then forms jets of relativistic particles and expels them perpendicular to the disk in both directions. That something probably involves twisted magnetic fields, but it may also involve the BHs directly. Since BHs have an electrical conductivity (Sec. 2.2.5) and will in general rotate, they can interact with magnetic fields to produce large voltage differences from pole to equator, which can accelerate and indeed even create pairs of charged particles. The energy would come from spinning down the rotating BH (Blandford and Znajek, 1977). Modeling the quasar and jet activity of galaxies is one of the most active areas of theoretical astrophysics today.

2.4.3 Gravitational Lensing. In Sec. 1.2.2.5 we saw that gravitational fields bend light, and that this can lead to a form of lensing. The amount of deflection depends on the theory of gravity, but the qualitative features are the same in all theories. Lensing was predicted more than 50 years ago, but the phenomenon does not seem to have been taken very seriously by astronomers until the now-famous “double quasar” 0957+561 was discovered in 1979 (Walsh *et al.*, 1979). Since then, dozens of examples of gravitational lenses have been found.

Gravitational lensing is one of the most dramatic confirmations of relativistic gravity. Lensed images come in many forms. Some are simply multiple images of a single point-like source, each with a different magnification. (The number of such images must always be odd, although they may not all be bright enough to be observed.) Sometimes images are smeared out into arcs. Some images do not change position but are “micro-lensed” by the gravitational field of a single star of an intervening galaxy that happens to move across the image, magnifying it briefly.

Lensing takes place at all scales, and to some extent in every astronomical image, but to be observable it must stand out from the confusion of other properties of images. The ideal situation is a compact image that is lensed into two or more similar images separated by a few arc seconds, far enough apart to be distinguished but close enough to be noticed as unusual. Spectra provide the “fingerprints” that convince one that the images are of the same object and not neighboring distinct objects. The lensing mass is typically a cluster of galaxies somewhere between the source and us.

Unless the galaxies in the cluster can all be seen, it is hard to model the lens. However, modeling has potential rewards. The biggest is to predict the Shapiro time delay (the excess travel time, as described in Sec. 3.2.1.2) along each of the image paths. Since the only data we have about the distance to and thus the linear size of the lens and the source are the redshifts produced in their spectra by the cosmological expansion, we need the *Hubble constant* (Sec. 2.4.5) to predict the actual time delays. Conversely, the measurement of the time delays between various images of a variable source in a well-modeled lens can provide a determination of the Hubble constant,

and from it an estimate of the age of the Universe. This is a very active area of research today.

2.4.4 Solar System and Stellar Orbits. General relativity makes small corrections to the orbits of bodies in the solar system. The effect may be described by a small $1/r^3$ correction to the gravitational acceleration produced by the Sun. It is smaller than the Newtonian term by the ratio of the Sun's gravitational radius to r , which is of order 10^{-7} or less for the planets. Its effect is to make the planet's ellipsoidal orbits *precess*: the direction from the Sun to the point of a planet's closest approach to the Sun (its *perihelion*) rotates with time.

The precession of Mercury's perihelion provided an early confirmation of GR, as we describe in Sec. 3.2.2. But at 43 seconds of arc per century, it is hard to measure, since the gravitational fields of other planets produce similar but much larger effects on Mercury's orbit. In fact, the measurement of the relativistic precession of any other planet is hopeless, partly because of the dominance of planetary perturbations.

When stars orbit each other, there are similar effects. They are easiest to observe in binaries containing pulsars, where we have excellent dynamical information about the orbit. Indeed, the periastron shift (as it is called for binaries) is necessary if we are to determine the masses of the individual stars. Fortunately, the orbits of such stars bring them much closer to each other than Mercury gets to the Sun, so the effect is larger. For the pulsar binary system PSR1913+16 (Sec. 3.3), the rate of periastron advance is 4.2° per year, easily measurable.

Another dramatic effect of GR on the orbits of stars is caused by the emission of GWs by the orbital motion. The energy lost to waves brings the stars gradually closer together, and also tends to make orbits more circular. As the stars approach, their orbital period decreases, and the system actually speeds up. This effect has been observed in PSR1913+16. It is also thought to control the evolution of certain classes of binaries, where mass is transferred from one star to another when they get too close. The transfer pushes the stars apart, but they come closer together because of gravita-

tional radiation reaction. The net effect is a steady transfer of mass from one star to another.

2.4.5 Cosmology, Inflation, and the Origins of the Universe. General relativity opened the door to the study of cosmology, the Universe in the large. In Newtonian gravity it is not even possible to formulate the gravitational force consistently for a Universe of infinite size, and if one takes the Universe to be finite but large, then the exact shape of its boundary affects the gravitational field everywhere in a measurable way. General relativity gives a consistent picture of the Universe. The reader should consult the article GALAXIES AND COSMOLOGY for a complete survey of cosmology; we concentrate here on the general relativistic aspects of the study.

2.4.5.1 Cosmological Models. Observations suggest that we can make a good first approximation to describing the Universe by assuming that it is homogeneous on the largest scales and isotropic around us. Since the Universe is also expanding, these assumptions imply the existence of a special choice of time everywhere: “now” has to be chosen so that the density, say, is the same everywhere at that time. This choice of time is unique: the Universe has a preferred reference frame. When cosmologists speak of the distance between galaxies, they are referring to proper distance at a particular time in this preferred frame. Similarly, the age of the Universe is the proper time since the Big Bang, as measured in this frame.

There are only three possible geometries for the spaces of constant time in a homogeneous and isotropic cosmology: flat, open, and closed. In the flat model, space at a constant time is Euclidean. In the closed model it is a three-sphere, which is the locus of all points equidistant from a fixed point in four Euclidean dimensions (equidistant as measured by the four-dimensional Euclidean distance). In the open model it is a three-hyperbola, which cannot be described as a subsurface of Euclidean space; rather, it has the same geometry as the hyperbola in Minkowski spacetime that consists of all points at a given timelike interval from the origin.

Without a cosmological constant, the field equations of GR imply that all models have a singularity at some time: they expand from a singular point of infinite density, contract to

one, or both (the closed model). Since we observe the Universe to be expanding (see Sec. 2.5.4.2), we infer that there was a singular point a finite time in the past. This is the *Big Bang*.

If the initial expansion velocity was sufficiently great compared to the overall gravitational attraction, the expansion will continue forever. The critical measure here is the local density of mass–energy. It turns out that models that expand forever have hyperbolic space sections; those that recontract are three-spheres; and the marginal case is flat. Therefore, measuring the actual mass–energy density has implications for the large-scale structure of the Universe.

Although much is sometimes made of the philosophical implications of this, it must be emphasized that we can only observe a small section of our Universe, and any assertion about the large-scale topology of space depends on our assuming that the parts of the Universe that are too far away to observe are in fact identical to the section we can observe. This can, of course, never be proved.

2.4.5.2 Hubble’s Law: The Big Bang. The expansion velocity of the Universe is given by the Hubble constant. Edwin Hubble was the first to show, by painstaking systematic observation of many galaxies, that all distant galaxies recede from us, and more distant galaxies recede faster (see GALAXIES AND COSMOLOGY). He expressed this as a linear relation between the distance d of a galaxy and the recession velocity v :

$$v = Hd. \quad (26)$$

We call the proportionality constant H the Hubble constant. This law is what one expects in a Universe that expands homogeneously: in any given interval of time, galaxies twice as far away should have moved by twice as much, leading to twice the recession velocity.

A frequently suggested analogy is an expanding balloon. If one draws dots on the balloon and measures their separations, then when the balloon is twice as large the points will have receded from one another in exactly this way. The balloon is also an excellent two-dimensional analog of the geometry of the three-spherical (closed) universe.

2.4.5.3 Cosmic Microwave Background Radiation. The Big Bang picture provides a simple model for our history. At a finite time

in the past, the Universe was infinitely dense. It began expanding and maintained thermal equilibrium among all the exotic species of particles for a short time. Gradually, as the temperature dropped, different species froze out. When nuclei froze out, about 20% of the mass was in helium, the rest in hydrogen. Later, hydrogen atoms froze out and the expanding cloud of gas became neutral. The mean free path of photons increased dramatically, and most photons in existence at the time have not scattered since. This epoch of “decoupling” of photons and matter left behind a thermal distribution of photons that has remained thermal as the Universe has expanded. The temperature of this distribution is inversely proportional to the expansion. Today this temperature is about 2.7 K, and the resulting *microwave background radiation*, discovered by Penzias and Wilson (1965), is one of the strongest pieces of evidence favoring the Big Bang model.

2.4.5.4 The Age of the Universe. The measurement of the Hubble constant is fraught with difficulties. The velocity v is easy to obtain from spectra, but measuring the distance is hard. Astronomers look for “standard candles”: objects of known intrinsic brightness, whose apparent brightness can be used to infer their distance. No really reliable standard candles exist, and even today astronomers differ by up to a factor of 2 in their determinations of H . In the astronomers’ customary units, where distances between galaxies are measured in megaparsecs (1 Mpc = 3×10^{22} m = 3×10^6 light-years) and velocity is measured in km/s, H is likely to lie in the following range:

$$45 \text{ km s}^{-1} \text{ Mpc}^{-1} < H < 100 \text{ km s}^{-1} \text{ Mpc}^{-1} .$$

It is usual to parametrize H by $h = H / (100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

If the expansion were at a constant rate, then the age of the Universe (time since the Big Bang) would be $1/H$. If $h=1$, then $1/H = 10^{10}$ yr. Smaller values of H lead, of course, to older Universes. In fact, gravity has been slowing down the expansion rate, so $1/H$ overestimates the age of the Universe. Therefore one can also constrain H by measuring the age of, say, the Earth, or better, of an old (hopefully first-generation) star or cluster of

stars. The Earth seems to be about 4.5×10^9 years old, but some clusters may be older than 10^{10} years.

2.4.5.5 Mass and Missing Mass. Once we know H we know the so-called critical density, the density of mass-energy just needed to stop the expansion, to make the Universe flat. In terms of h , the critical density is

$$\rho_c = 2 \times 10^{-26} h^2 \text{ kg m}^{-3} .$$

Measuring the actual mass-energy density is even harder than measuring H ; the uncertainty seems to be a factor of 10 at present. The problem is that the mass that is radiating light, mainly in the form of stars, accounts for less than 1% of the critical density, and there is much evidence for at least 10 times this amount of dark or hidden mass. The evidence is dynamical: the rotation rates of spiral galaxies and the internal motions of large clusters of galaxies are simply much too large for these objects to stay together if their masses were only what we could infer from the light they emit. This dark mass may consist of BHs or other low-brightness remnants of the early evolution of stars, or (and this is the more popular idea at present) it may consist of a cosmological distribution of elementary particles of an unknown kind (Knapp and Kormendy, 1986). Searches are now under way for such particles randomly passing through the laboratory. Their detection would have important implications for high-energy physics as well as for astronomy.

Indirect evidence of this dark matter may be sought in the dynamics of galaxy formation. It turns out that density perturbations in ordinary baryons do not grow rapidly in the early Universe, mainly because these particles are charged and couple too strongly to the equilibrium radiation field. Once the Universe has expanded and cooled off enough for hydrogen to become neutral, the perturbations may grow, but that does not leave enough time for galaxies as we observe them—well separated objects even at early times—to have evolved.

A pervasive background of neutral particles could provide seeds for galaxy formation that grew unhindered in the early Universe. If the particles had reasonably large rest mass, they would cool off rapidly and form strong condensations. Computer simulations performed within this *cold dark matter* hypothe-

sis provide reasonable consistency with the present distribution of galaxies and their clustering. It must be stressed, however, that there are many other possibilities for providing seeds for galaxy formation. Cosmic strings, mentioned below, are also attractive. The real story of galaxy formation may involve several such factors.

2.4.5.6 Microwave Anisotropies. The microwave background ought, if the Big Bang is correct, to be a very good black-body distribution. Indeed, measurements by the recent COBE spacecraft show it to be (Smoot *et al.*, 1992). However, its temperature need not be the same in all directions. For one thing, the Earth may be moving with respect to the cosmological rest frame, the mean rest frame of the matter at decoupling. This would produce a dipole anisotropy in the temperature, and this has been measured. The inferred velocity of the Sun is 370 km s^{-1} , again from the COBE measurements.

In 1992 COBE also measured another effect on the microwave temperature: fluctuations that arise from the early density irregularities of the Universe. While not surprising, these provide striking “baby pictures” of the condensations that eventually evolved into large aggregates of clusters of galaxies.

2.4.5.7 Inflation and the Homogeneity Problem. The measurements performed by COBE highlight a conceptual problem for cosmology: the structures seen by it were on a scale so large that light could not have traveled across the structure in the time between the Big Bang and the time of decoupling, when the radiation was produced. Yet the structures seen by COBE were relatively consistent right around the sky. How could they be causally related? And if they were not causally related, why are they so similar?

The same problem exists for other measurements. We can see quasars in opposite directions on the sky that are today so far apart that they could not have communicated since the Big Bang, and yet the distributions of quasars and other galaxies near them look similar at both places. How indeed can the Universe be homogeneous over distances that could not communicate?

A simple answer is that it is all in the initial conditions of the Big Bang. But such an explanation is unsatisfying to many scientists, and an attractive alternative has been proposed: inflation. If, in the early universe, just

after the Big Bang, the laws of high-energy physics at energy scales inaccessible to present-day experimentation were of a certain form, then the Universe could at some point have undergone a phase transition that pushed it into very rapid expansion, much more rapid than one would infer by working backwards from today's expansion. The expansion would in fact have been exponential, hence the term *inflation*. If it halted after many *e*-foldings, then parts of the Universe that now seem too far apart to have communicated would in fact have been much closer before inflation, and could have been causally connected. Inflation can solve, at least in this sense, the homogeneity problem (Guth and Steinhardt, 1984; Blau and Guth, 1987).

Inflation has side benefits, one of which is that it might produce exactly the spectrum of seed perturbations that COBE saw. Observational evidence is slim at present, but by the year 2000 one can expect much stronger constraints on inflationary models.

3. TESTS OF GRAVITATIONAL THEORIES AND THEIR TECHNOLOGICAL DEMANDS

For the first half-century of GR's existence, experimental tests were infrequent, largely because the smallness of the predicted effects made them extremely difficult to measure with any accuracy. The validity of the theory rested on two tests, the deflection of light, measured in 1919, and the resolution of an anomaly in Mercury's perihelion advance. But beginning in the 1960s, advances in technology made high-precision tests of the theory possible, and there followed a systematic effort to put the theory to the test (for a review of these tests see Will, 1993).

3.1 Tests of the Einstein Equivalence Principle

3.1.1 Tests of Special Relativity. The Einstein equivalence principle described in Sec. 1.2.2.2 demands that in any local, freely falling reference frame (in which gravity is absent locally), the nongravitational laws of physics (such as mechanics, electromagnetism, quantum mechanics) must be compatible with special relativity. As we have seen, a

consequence of this principle is that gravity must be described by spacetime curvature.

Special relativity has become such a successful and integral part of such areas of modern physics as quantum field theory, nuclear physics, and particle physics that physicists often take its validity for granted. But in many of these subdisciplines of physics, the experiments are designed to test particular models of fields and interactions rather than the underlying special relativistic framework. The Michelson–Morley experiment and its modern-day descendants provide clean tests of SR in that they can constrain directly and quantitatively possible violations of SR (see RELATIVITY, SPECIAL).

3.1.1.1 Michelson–Morley Experiments.

One way to understand the significance of these tests is to imagine an explicit violation of SR in electrodynamics, by permitting the limiting speed of material particles, c_m , to differ from the speed of electromagnetic waves, c_e . Since moving frames are no longer equivalent, this assumption establishes a preferred universal rest frame. From the field-theoretic point of view, such a violation is likely to be induced by some long-range field created by the matter in the universe; hence the preferred frame is likely to be that of the smoothed out distribution of matter, or equivalently of the cosmic microwave background. The resulting observable violations of SR in such a model depend on the fact that the Earth is moving through the universe at about 370 km/s, and are parametrized by $\delta = (c_m/c_e)^2 - 1$. By placing a limit on a difference in the speed of light in two perpendicular directions using an interferometer, the 1887 Michelson–Morley experiment set a limit $|\delta| < 10^{-4}$. A 1979 laser-interferometric version of the experiment improved the limit to 10^{-9} (Brillet and Hall, 1979).

3.1.1.2 Hughes–Drever Experiment. But in 1960, a substantial improvement in the limit on δ resulted from experiments that tested the “isotropy of inertia”, done independently by Hughes *et al.* (1960) and Drever (1961). If SR is violated in the way described above, then the energy levels of an atom or nucleus can depend both on the velocity of the nucleus through the preferred frame and on the orientation of the quantization axis relative to the direction of motion. Anomalous shifts of energy levels can then occur. The Hughes–Drever experiments used nuclear mag-

netic resonance techniques to look for such anomalous shifts as the Earth rotated in space. More recently, new techniques of atomic physics, such as laser-cooled ion and atom traps, have yielded improved limits on such variations with direction in the energy levels of beryllium ions, neon atoms, and mercury isotopes, ranging from 2×10^{-19} to 2×10^{-21} eV. These results give the bound $|\delta| < 10^{-21}$, a truly high-precision confirmation of SR (Prestage *et al.*, 1985; Lamoreaux *et al.*, 1986; Chupp *et al.*, 1989).

3.1.2 The Eötvös Experiment, the Weak Equivalence Principle, and the Fifth Force

3.1.2.1 Eötvös and the Weak Equivalence Principle. Another experiment that helped lay the foundation for GR was the Eötvös experiment (1889, 1908), which verified what we have called the weak equivalence principle (WEP), the equality of gravitational acceleration of objects of different composition. The precision achieved was a few parts in 10^9 . Two new experiments, by Dicke at Princeton University in the early 1960s (Roll *et al.*, 1964), and by Braginski at Moscow State University in 1970 (Braginski and Panov, 1971), improved the accuracy by two orders of magnitude.

3.1.2.2 Fifth-Force Experiments. In 1986, there was renewed interest in the Eötvös experiment. As a result of a detailed reanalysis of Eötvös’s original data, Fischbach *et al.* (1986) suggested the existence of a “fifth force” of nature, with a strength of about a percent of that of gravity, but with a range (as defined by the range λ of a Yukawa potential $e^{-r/\lambda}/r$ augmenting the usual Newtonian potential) of a few hundred meters. This proposal dovetailed with earlier hints of a deviation from the inverse-square law of Newtonian gravitation derived from measurements of the gravity profile down deep mines in Australia.

During the next four years, over a dozen new experiments looked for evidence of the fifth force by searching for composition-dependent differences in acceleration, with variants of the Eötvös experiment or with free-fall Galileo-type experiments. Many of the Eötvös-type torsion-balance experiments took advantage of new high- Q torsion suspension systems, advanced seismic isolation, and sophisticated arrangement of component masses to reduce gravity-gradient couplings.

One of the free-fall experiments used laser interferometry to measure the differential acceleration of two chambers, fitted with corner retroreflectors, containing different materials. Although two early experiments reported positive evidence, all others yielded null results. Over the range between 1 and 10^4 m, the null experiments produced upper limits on the strength of a postulated fifth force of between 10^{-3} and 10^{-6} the strength of gravity. Interpreted as tests of the WEP (corresponding to the limit of infinite-range forces), the results of a University of Washington experiment (dubbed Eöt-Wash) were comparable to that of Dicke (Adelberger *et al.*, 1991).

At the same time, tests of the inverse-square law of gravity were carried out by comparing variations in gravity measurements up tall towers or down mines or boreholes with gravity variations predicted using the inverse-square law together with Earth models and surface gravity data mathematically "continued" up the tower or down the hole. Despite early reports of anomalies, three independent tower measurements now show no evidence of a deviation (see, for example, Jekeli *et al.*, 1990). The consensus at present is that there is no credible experimental evidence for a fifth force of nature (for recent reviews see Adelberger *et al.*, 1991; Fischbach and Talmadge 1992; and Will 1990).

3.1.3 Gravitational Redshift. Although the gravitational redshift of light is a simple consequence of the EEP (Sec. 1.2.2.2), as Einstein found some eight years before he completed the full theory, it was not confirmed experimentally until the Pound-Rebka experiment of 1960, in which the frequency shifts of γ rays rising and falling in a tower were observed, making use of the Mössbauer effect to reduce recoil broadening of the emission and absorption γ -ray lines (Pound and Rebka, 1960).

Other tests included transporting atomic clocks on jet aircraft, and measuring the shift of solar spectral lines. The most accurate confirmation to date has been a 1976 rocket experiment, in which a hydrogen maser clock was launched on a Scout rocket to an altitude of 10 000 km, and its rate compared with an identical clock on the ground, resulting in a 0.02% test (Vessot *et al.*, 1980). Recently, a measurement of the shift of the rate of oscillator clocks on the Voyager spacecraft caused

by Saturn's gravitational field yielded a 1% test (Krisner *et al.*, 1990).

The gravitational redshift now has practical consequences. In satellite-based navigation systems, such as the U.S. Air Force's Global Positioning System, the atomic clocks on the satellites tick faster, as a consequence of the gravitational frequency shift and the special relativistic time dilation, than do clocks on the ground by over 30 000 ns per day. Yet the navigational accuracy requirement of 10 m demands timekeeping accuracy to 30 ns at all times, and therefore general relativistic corrections *must* be taken into account in order for the system to function. Since GPS is now used extensively both for military and civilian navigation, this represents a new "applied" side of GR.

3.2 Solar-System Tests of General Relativity

3.2.1 The Deflection and Retardation of Light

3.2.1.1 Light Deflection. One of the first calculations that Einstein performed in November of 1915, when he had the final (vacuum) field equation of GR, was the deflection of light. Earlier, in 1911, he had determined the deflection in a preliminary theory based essentially purely on the EEP (Sec. 1.2.2.2) and got the answer, $2GM/Rc^2$, as in a Newtonian gravity theory in which light was treated as a corpuscle (see Sec. 1.2.2.5).

The result of Einstein's 1915 calculation was to double the prediction. For a light ray that grazes the Sun, for example, the deflection would be $1.75''$ instead of $0.875''$. The difference can be understood as follows: half the deflection indeed comes directly from the Einstein principle, or equivalently from a Newtonian ballistic calculation; the remaining part derives from the curvature of space near the Sun relative to space far away. The first contribution is the same in any theory of gravity that is a metric theory. The second, space-curvature contribution varies from one metric theory to another, and is conventionally represented by the PPN parameter γ , whose general-relativistic value is 1 (see Sec. 1.2.5). In this parametrized language, the deflection of a light ray by the Sun is given by

$$\Delta\theta = \frac{1}{2}(1 + \gamma)1.75''/d, \quad (27)$$

where d is the distance of closest approach of the ray to the Sun, in units of a solar radius.

The measurement of this effect by British astronomers during a total solar eclipse in 1919 (Dyson *et al.*, 1920) catapulted Einstein and GR to worldwide fame. However, the accuracy was 20% at best. A few measurements during the next 45 years failed to yield substantial improvements. The development of radio interferometry during the 1960s coupled with the discovery of quasars led to dramatically better accuracy (see RADIO TELESCOPES). The technique involved monitoring the relative angle between a pair or group of quasars as they passed near the Sun as seen from Earth. During the passage, the light from a quasar whose image is closer to the Sun would be deflected more than that from a more distant one, leading to a displacement of one image relative to the other. Between 1969 and 1975 a dozen measurements of this sort were carried out systematically, culminating in confirmations of GR at about the 1.5% level (Fomalont and Sramek, 1976). After 1975, further direct measurements of the deflection of light to test relativity essentially ceased.

However, in the early 1980s systematic efforts were initiated by geophysicists to monitor the Earth's rotation state accurately by means of very-long-baseline interferometric (VBLI) measurements of the positions of radio galaxies and quasars. Using transcontinental and intercontinental baselines, and improved timing accuracy made possible by hydrogen maser and other atomic clocks, these measurements reached the several hundred *microarcsecond* level in accuracy, making it necessary to take the relativistic deflection of light into account over the entire celestial sphere, not just near the Sun. For a ray that approaches the Earth from a direction 90° away from the Sun, for example, the deflection is 4 milliarcseconds and is readily detectable. A by-product of this effort was a 0.1% confirmation of GR (Robertson *et al.*, 1991).

The European Space Agency (ESA) satellite HIPPARCOS, which is completing its mission as this is being written (1993), has similarly made precise determinations of the positions of some 10^5 stars, for which it has to take into account (and therefore measure) the light deflection over the whole celestial sphere. This will measure γ to similar accuracy.

3.2.1.2 Shapiro Time Delay. There is another important test of the propagation of light through curved spacetime, which was not around in 1915, yet which is closely related to the deflection of light. It was first predicted as a consequence of GR by Shapiro (1964), and is now commonly called the Shapiro time delay. It is an excess propagation delay of light passing through a region of curved space near a body compared to the analogous propagation time if the ray passes far from the body. A light ray that passes the Sun on a round-trip, say, from Earth to Mars at superior conjunction (when Mars is on the far side of the Sun) suffers a delay given by

$$\Delta t \approx \frac{1}{2}(1+\gamma)250(1-0.16 \ln d) \mu\text{s}. \quad (28)$$

The close relationship between this effect and the deflection of light is reflected in the factor $\frac{1}{2}(1+\gamma)$ and is to be expected, since any phenomenon that bends light (refraction, curved space) may be expected to alter its propagation time as well. Observations of the Shapiro time delay began in the mid-1960s with the use of radar echoes from Mercury and Venus. Later, use was made of interplanetary spacecraft equipped with radar transponders, such as Mariners 6, 7, and 9 and the Viking landers and orbiters. Data from Viking yielded a 0.1% test (Reasenberg *et al.*, 1979).

The result of these measurements on light is $\gamma = 1.000 \pm 0.002$, in agreement with GR. This precise determination of the parameter γ is one of the crowning achievements of experimental gravitation.

3.2.2 Mercury's Perihelion Advance. The first effect that Einstein calculated in November 1915 using his new field equations was the advance of the perihelion of Mercury. The discrepancy between the observed advance and the amount that could be accounted for from the Newtonian gravitational perturbations of Mercury by the other planets was a problem that had bedeviled celestial mechanics for the latter half of the nineteenth century. As mentioned in Sec. 2.4.4, GR predicted an amount that neatly accounted for the discrepancy. Einstein wrote later that he had palpitations of the heart upon finding this result. For another half a century, this stood as one of the triumphs of general relativity.

The predicted rate of advance of the perihelion of Mercury (excluding the part from planetary perturbations) can be written in the

following form, in seconds of arc per century:

$$d\omega/dt = 42.98''\lambda_p, \quad (29)$$

$$\lambda_p = \frac{1}{3}(2 + 2\gamma - \beta) + 0.0003(J_2/10^{-7}). \quad (30)$$

The first term in the coefficient λ_p is the relativistic contribution to the advance, in a form that encompasses a wide class of alternative metric theories of gravity. The parameter γ is the same parameter that appeared in the deflection of light and the Shapiro time delay, while the parameter β is a rough measure of how "nonlinear" gravity is in a given theory. Both parameters are unity in GR. The second term comes from the Newtonian effect of a possible oblateness of the Sun, which will alter its external gravitational field from the pure inverse-square form of a spherical body. The oblateness is measured by the quantity $J_2 = (I_3 - I_1)/(M_\odot R_\odot^2)$, where I_3 is the Sun's moment of inertia about its rotation axis and I_1 is the same about an equatorial axis; for a Sun that rotates uniformly with its observed surface angular velocity, so that the oblateness is caused by centrifugal flattening, J_2 is estimated to be of order 10^{-7} .

Now, the measured perihelion shift of Mercury is known very accurately because of the combination of two factors: improved radar ranging to Mercury since 1966, leading to a more accurate determination of its orbit, and improved data on the masses and orbits of the other planets from radar ranging and spacecraft encounters, leading to improved values for the planetary perturbations. After those perturbing effects have been accounted for, the excess shift is known to about 0.1%, with the result that $\lambda_p = 1.000 \pm 0.001$ (Shapiro, 1990).

If J_2 were indeed as small as 10^{-7} this would be in complete agreement with GR. However, in 1966, a value for J_2 of 2.5 ± 10^{-5} was inferred from visual solar-oblateness measurements made by Dicke and Goldenberg (1974), a result that, if confirmed, would have disagreed strongly with GR. Between 1966 and 1980 J_2 values ranging over two orders of magnitude were reported. Beginning around 1980, however, the observation and classification of modes of oscillation of the Sun ("helioseismology") made it possible to obtain information about its internal rotation rate, thereby constraining the possible centrifugal flattening; current results favor a value $J_2 \approx 1.7 \pm 10^{-7}$ (Brown *et al.*, 1989), making the peri-

helion shift of Mercury another success for general relativity.

3.2.3 Test of the Strong Equivalence Principle. Another important Solar-System experiment tests a generalization of the EEP, known as the strong equivalence principle (SEP). The strong equivalence principle states, for example, that *all* bodies should fall with the same acceleration in an external gravitational field, including bodies with significant internal gravitational binding energy, such as planets, stars, and so forth. In the WEP, one considers only laboratory-sized bodies, whose internal structures are dominated by nongravitational energies. Different theories of gravity can treat the effect of gravity on gravitational energy differently, and so could predict violations of the SEP by massive, self-gravitating bodies. General relativity is one of the few theories that actually obeys the SEP. The Brans-Dicke theory, for example, does not.

Since 1969, this principle has been tested using lunar laser ranging (LURE), in which laser pulses sent from Earth bounce off corner retroreflectors deposited on the Moon during U.S. and Soviet lunar landings. The goal is to look for the orbital effects of a possible difference in acceleration between the Earth and Moon toward the Sun (the *Nordtvedt effect*: Nordtvedt, 1968a,b). No orbital perturbation of this type has been found to date down to the 6-cm level, placing a limit of 7 parts in 10^{13} on a difference in acceleration between the two bodies (Dickey *et al.*, 1989; Shapiro, 1990; Müller *et al.*, 1991). The accuracy of lunar laser ranging is approaching the level of several millimeters, at which point the accuracy of this experiment as a test of the effect of gravity on gravitational energy (test of the SEP) will be limited by the accuracy of laboratory tests of the weak principle, because the composition of the Earth (iron rich) and Moon (iron poor) differ.

3.3 The Binary Pulsar: An Astronomical Relativity Laboratory

Until 1974, the solar system provided the principal testing ground for GR, because it is a "clean" system (few uncertain or messy physical processes to complicate the gravitational effects) and it is accessible to high-precision tools. However, the discovery of the binary pulsar PSR1913+16 in 1974 by

astronomers at the Arecibo Radio Telescope in Puerto Rico (Hulse and Taylor, 1975) showed that certain kinds of distant astronomical systems may also provide precision laboratories for testing GR. The system consists of a 59-ms-period pulsar in an 8-h orbit with a companion that has not been seen directly, but that is generally believed to be another NS. The unexpected stability of the pulsar "clock" and the cleanliness of the orbit allowed radio astronomers to determine the orbital and other parameters of the system to extraordinary accuracy, by analyzing the variations in pulse arrival times caused by the orbital displacements (for the most recent data, see Taylor *et al.*, 1992).

Furthermore, the system is highly relativistic ($v_{\text{orbit}}/c \approx 10^{-3}$). Observation of the relativistic periastron advance ($4.226\,628 \pm 0.000\,018$ deg per year) and of the effects on pulse arrival times of the gravitational redshift caused by the companion's gravitational field and of the special relativistic time dilation caused by the pulsar's orbital motion (0.15% accuracy) have been used, assuming that GR is correct, to constrain the nature of the system. In GR, these two effects depend in a known way on measured orbital parameters and on the unknown masses m_p and m_c of the pulsar and companion (assuming that the companion is sufficiently compact that tidal and rotational distortion effects can be ignored), and consequently the two masses may be calculated with these two pieces of data, with the result $m_p = (1.4411 \pm 0.0007)M_\odot$ and $m_c = (1.3873 \pm 0.0007)M_\odot$. These are the most accurately known masses of any astronomical bodies outside the Solar System. It is interesting how GR plays a crucial role in this high-precision determination of astrophysical parameters.

This binary system should radiate GWs, which, unfortunately, are of too low a frequency to be detected directly by a ground-based GW detector. On the other hand, the loss of energy in GWs should cause the orbiting stars to spiral together, and the orbital period to *decrease*. The "quadrupole formula" (Sec. 2.3.3) determines the rate of loss of energy and the consequent orbital damping rate. It was first detected successfully in 1979, with about 10% precision. Using the measured orbital elements and the two masses, one can predict the rate of decrease of the period to be $dP/dt_{\text{predicted}} = -(2.402\,43 \pm 0.000\,05) \times 10$

-12 s/s. Because of the extraordinary stability of the pulsar, and with the added ability to transfer stable atomic time from the world's system of atomic clocks to the Arecibo Radio Telescope via the Global Positioning System, the observations of this period decrease have now reached 0.5% in accuracy, giving $dP/dt_{\text{observed}} = -(2.408 \pm 0.015) \times 10^{-12}$ (Taylor *et al.*, 1992). This agrees completely with the prediction and is very strong indirect evidence for the correctness of GR's predictions about GWs.

3.4 Future Work in Experimental Gravitation

3.4.1 Search for Gravitomagnetic Effects.

According to GR, moving or rotating matter should produce a contribution to the gravitational field that is the analog of the magnetic field of a moving charge or a magnetic dipole (Sec. 2.1.1). Although gravitomagnetism plays a role in a variety of measured relativistic effects, it has not been seen to date, isolated from other post-Newtonian effects.

The Relativity Gyroscope Experiment at Stanford University (also known by the NASA terminology Gravity Probe B, or GP-B) is in the advanced stage of developing a space mission to detect this phenomenon directly (Everitt, 1988). A set of four superconducting-niobium-coated, spherical quartz gyroscopes will be flown in a low polar Earth orbit, and the precession of the gyroscopes relative to the distant stars will be measured. The predicted effect of gravitomagnetism is about 42 milliarcseconds per year, and the accuracy goal of the experiment is about 0.5 milliarcseconds per year.

To achieve this accuracy, which corresponds to a precession rate of 10^{-16} rad/s, numerous technical challenges have had to be met, including fabricating gyroscopes that are homogeneous and spherical to better than a part per million; developing and testing a "London moment" readout system that exploits the magnetic dipole moment developed by a spinning superconductor and uses SQUIDS to read out the varying currents in superconducting loops surrounding the gyroscope; and developing a magnetic shield of novel design to reduce the ambient magnetic field of the Earth below 10^{-7} G. Recently, a full-size flight prototype of the instrument package was tested as an integrated unit. Current plans call for a

test of the final flight hardware on the Space Shuttle followed by a shuttle-launched science experiment.

Another proposal to look for an effect of gravitomagnetism is to measure the relative precession of the line of nodes of a pair of laser-ranged geodynamics satellites (LAGEOS), with supplementary inclination angles; the inclinations must be supplementary in order to cancel the dominant nodal precession caused by the Earth's Newtonian gravitational multipole moments (Ciufolini, 1989). A third proposal envisages orbiting an array of three mutually orthogonal, superconducting gravity gradiometers around the Earth to measure directly the contribution of the gravitomagnetic field to the tidal gravitational force (see, for example, Mashhoon *et al.*, 1989).

3.4.2 Tests of the Einstein Equivalence Principle. The concept of an Eötvös experiment in space has been developed as a possible joint NASA-ESA mission, with the potential to test the WEP to 10^{-17} , a millionfold improvement over current ground-based results (Worden, 1988). Such an experiment could also look for additional long-range, fifth-force-type interactions, with ranges in excess of about 40 km, and could lead to improvements in the value of the Newtonian constant G . It would also map the multipole moments of the Earth's Newtonian field to high accuracy.

The accuracy of measurements of gravitational redshift could be improved to the 10^{-9} level, and higher-order effects could be seen for the first time by placing a hydrogen maser clock on board Solar Probe, a proposed spacecraft that would travel to within four solar radii of the Sun (Vessot, 1989).

3.4.3 Further Fifth-Force Searches. Because they are relatively inexpensive and because they have the potential to constrain certain classes of particle-physics models, laboratory fifth-force experiments are likely to continue at some level for many years.

4. GRAVITATIONAL-WAVE DETECTION: A TECHNOLOGICAL FRONTIER

4.1 Likely Sources of Detectable Waves

In Sec. 2.3 we learned about GWs. Here we discuss their detection. Although detectors

have been under development since Weber (1960) built the first one at the University of Maryland in the early 1960s, it is only recently that technology has permitted the design of a detector that meets theoretician's predictions about the likely strengths of expected waves. It is not unreasonable to expect that, by the year 2000, the first direct detection of a GW will have occurred.

Simple detection is not the main goal of the present detector development. GWs carry information about their sources that is obtainable in no other way. In order to extract maximum information from the waves, one needs to be able to infer their amplitude and direction. Since GW detectors are not directional, this can only be done with a worldwide network of detectors, which infer directions from the relative times of arrival of waves at different locations. Three detectors is the minimum for extraction of full information from the waves.

The first aspect of designing a detector is to estimate what amplitudes and frequencies one might expect from astronomical sources. These set targets for the experimental development.

We can dispose of one source right away: laboratory generators. It is not possible to build a laboratory generator of detectable GWs. One can make an apparatus that disturbs a detector by generating a time-dependent gravitational field; this has been done in several laboratories (for example, Astone *et al.*, 1991a). But it is always the near-zone Newtonian field, and the disturbance would be the same in Newtonian gravity. To detect waves one must be at least one wavelength λ away from the source, and at this distance all reasonable sources are too weak.

We look therefore to relativistic sources in astronomy. All of these are powered by their internal gravitational forces, and so the key equation for estimating the strength of their emissions is Eq. (24), which sets a relatively simple upper bound on the amplitude we can expect. Here is a brief review of the principal candidates and the astronomical information we might expect to get from them. It is by no means an exhaustive list, but it represents the most conservative predictions that have been made about possible sources.

4.1.1 Supernovae. Supernovae are rare events, occurring once in perhaps 50 years in any galaxy. We would like to be able to detect them, therefore, in a volume of space containing perhaps 2000 galaxies, so that we have a reasonable chance of seeing one. This means that realistic detectors must reach as far as the Virgo cluster, a cluster of galaxies about 15 Mpc away, containing more than 1000 galaxies.

Supernovae will give off GWs if the collapse event is very nonspherical. The mass involved will be about a solar mass, and the size of the emission region is about 10 km. These give, from Eq. (24), an upper limit of a few times 10^{-21} for waves from the Virgo cluster. Given that this is an upper bound, the usual target is 10^{-21} . A detector with that sensitivity has some chance of seeing an occasional supernova explosion, unless they are all very symmetrical.

The expected frequency can be inferred from the collapse time scale. The whole collapse takes about 10 ms, but not much radiation comes off during most of this time. The bounce time scale is about 1 ms, and subsequent oscillations may also have this time scale. The relevant frequency is the inverse, about 1 kHz. One expects a broadband burst of this central frequency from a supernova explosion.

If such bursts are seen, they may allow us to identify the object (NS or BH) formed by the collapse; they may provide crucial information about high-density nuclear physics; they may allow rapid notification of other astronomers that a supernova has occurred at a particular position; and they will enlighten our understanding of the late stages of stellar evolution.

Bursts of radiation will also be emitted if galactic-size BHs are formed in a single event or if compact stars fall into such BHs. The frequency of the emitted radiation is inversely proportional to the mass of the hole, and for holes in the centers of galaxies this can be in the millihertz region or lower. Such events are not observable from the ground since ground vibrations and the near-zone Newtonian gravitational disturbances produced by atmospheric mass motions in this frequency range are too strong to screen out. But studies have shown (Bender *et al.*, 1989) that space-based detectors could be very sensitive at these frequencies. A suitable detector could see a

giant BH formation event anywhere in the Universe; it could verify or rule out that mode of formation of galactic BHs.

4.1.2 Coalescing Binaries. The remarkable binary pulsar described in Sec. 3.3 will be even more remarkable in about 10^8 years, when the orbit has shrunk to the point that the two NSs in the system are orbiting 10 or more times per second. The system will be a strong source of GWs at relatively high frequency during the remaining few seconds before the stars coalesce. The frequency of the waves is twice the orbital frequency.

Such events are, of course, rare, occurring perhaps once every 10^5 to 10^7 years in any galaxy (Narayan *et al.*, 1991; Phinney, 1991; Tutukov and Yungelson, 1993). We therefore would need to be able to detect them in a volume of space up to 200 000 times larger than that for supernovae, or as much as 60 times further away. Fortunately, this is made possible by the regularity and predictability of the signal from such a system.

In contrast to supernovae, which are likely to be messy, binaries of compact objects emit a steady, almost monochromatic “chirp” signal, whose frequency increases with time in a predictable way. If one has N cycles of such a signal in one’s data stream, then one can use pattern-matching techniques (such as matched filtering) to find it at an amplitude that is smaller by a factor of \sqrt{N} than one could find if one had only a single cycle. Supernovae are basically single-cycle signals, so we can see binaries 60 times further away if we can detect 3600 cycles. The binary wave train will be followed by a much less predictable burst associated with the coalescence of the two stars.

The detector has to have a broad bandwidth at a few tens or hundreds of hertz to permit such filtering and following of the signal. In fact, the broadband interferometers described below are less noisy at the lower frequencies of binary signals, but on the other hand the intrinsic signal from a binary is a bit weaker than the strongest possible supernova (because ϕ_{internal} is smaller when the stars are still several radii apart), so that the calculation must be done carefully. The result is that detectors will have about 40 times the range for binaries as they have for moderate supernova explosions. Detectors that can just barely see supernovae in Virgo (such as the first-

stage interferometers described below) are unlikely to see coalescing binaries. But the second stage of development, with 10 times better amplitude sensitivity, ought to see hundreds per year.

The binary signal contains much information, including the individual masses of the component stars. By observing them, we will directly identify NSs and BHs and get much better statistics on their mass distributions. Since they occur at great distances, these events sample the cosmological mass distribution on very large scales. The signal contains enough information to allow the absolute distance to the binary to be estimated: they are true standard candles. They therefore contain cosmological information, including the possibility of measuring the Hubble constant (Schutz, 1986). The actual coalescence of the two objects (NSs and/or BHs) may be associated with the mysterious γ -ray bursts seen by satellite detectors (Meegan *et al.*, 1992). If so, they will be easier to model and understand. Finally, the detection of the radiation from the merger of two BHs will provide a strong test of GR itself; modeling this event on computers is an area of current research.

4.1.3 Pulsars. A NS that is axially symmetric will not emit GWs when it rotates, but we know that pulsars are not symmetrical; they have off-axis magnetic fields. If they have other asymmetries, perhaps mass deformations that help to pin the magnetic poles in one place, then they may give off detectable radiation. There is an upper limit on the strength of this radiation for any pulsar whose spin-down rate has been measured: GWs probably do not carry away more than the rotational energy loss of the pulsar.

The signal will be steady, so one's ability to find it increases with the square root of time, for exactly the same reason as for coalescing binaries. Again, a broadband detector is desirable, so that signals of any frequency can be detected. Given a few months' observation, second-stage interferometers are expected to be able to beat the upper limits on several pulsars by factors of 10 or more.

It is possible to conduct searches for unknown pulsars through their GW emission. Here, as for known pulsars, one must remove the Doppler effects caused by the Earth's motion before the signal becomes periodic. This correction depends on the position of the

pulsar. To perform a sensitive all-sky search is a demanding computational task, and the sensitivity will be limited by the capacity of available computers (Schutz, 1991).

4.1.4 Ordinary Binaries. Binary star systems emit radiation at twice the frequency of their orbit. For all except the coalescing binaries considered before, this is a very low frequency, not observable from the ground (Sec. 4.1.1). But space-based detectors would be able to reach as low as 10^{-4} Hz and could detect steady radiation from many binaries. One of the most prominent would be PSR1913 + 16. Other compact binaries might be discovered this way.

Binaries containing white dwarfs are so numerous that at the lower frequencies they provide a chaotic background of waves that might mask other sources, at least if only one detector is used in space.

4.1.5 Cosmological Background. Just as there is a cosmological background of electromagnetic radiation, one expects a background of GWs. The thermal background may be too weak to see, but many other potential sources have been discussed. Most rely on aspects of high-energy physics that give rise to inflation or to topological anomalies.

Theories that produce cosmic strings have been studied extensively. Strings are massive linear regions of trapped field, which can seed galaxy formation and can emit GWs. If strings do provide the seeds for galaxy formation, then they ought to produce a background of GWs that is detectable by interferometers (Allen and Shellard, 1992).

Searches for such a background rely on cross-correlation of the output of two detectors, since in a single detector the background appears simply as another source of noise. If the detectors are close enough, the background will produce a correlated response in both that can stand out against other local sources of noise.

4.1.6 Unexpected Sources. One of the most exciting prospects is that there might be significant radiation from unpredicted sources. Whenever new windows on astronomy have been opened, such as radio or x-ray astronomy, completely unexpected objects have been found. This will undoubtedly be true for GWs as well, but of course one does not know what sensitivity will be required to reveal them. By

going 10 times deeper in amplitude (100 times in energy) than the strongest predictions for bursts, second-stage interferometers must stand a good chance of making such discoveries.

4.2 Detectors

The reason for the current interest in GWs is technological: it now seems possible to build a detector that meets theoretical predictions about sources. But detectors of various types have been under development since 1960, and six or seven detectors around the world are able to make observations today. In this section we review the techniques, with an emphasis on the areas of technology that are relevant.

4.2.1 Bar-Type Detectors. Weber designed and built the first GW detector in the 1960s. It was based on a cylinder of aluminum, suspended and isolated from ground vibrations (Weber, 1967). When a GW hits the bar broadside, the tidal forces represented in Fig. 1 stretch it along its axis. By monitoring the excitation of the fundamental longitudinal mode of vibration of the bar, one can look for GWs.

Weber chose a bar that had a resonant frequency in the kilohertz region to look for supernova events. He showed that, even for a burst event, it is desirable for the vibrational mode to have a large Q ; such a mode is weakly coupled to the thermal bath of the other modes, making it easier to recognize the rapid change in its excitation caused by a GW.

The dominant background is thermal vibrations. Subsequent generations of bars have been cooled to 4.2 K to reduce this background. A further generation of bars now under construction (1993) may reach 10 mK (for example, see Astone *et al.*, 1991b). Weighing several tons, such bars will be the coldest large objects the Universe has ever seen! Where Weber's original bar had a sensitivity limit of $h \sim 10^{-16}$ for a broadband burst around 1 kHz, cryogenic bars today reach about 10^{-18} , and the new bars may go to 10^{-20} .

For bars to go below this to the interesting level of 10^{-21} means defeating the *quantum limit*. This refers to the fact that, in classical terms, the excitation energy deposited in a bar by a GW of that amplitude will be smaller than the energy of one phonon of longitudinal excitation. Although one might think this would

be an absolute limit, GW theorists showed that one can use the uncertainty principle to "squeeze" the uncertainty in a particular observable (not the phonon number) and deduce from that the amplitude of the incident GW (Caves *et al.*, 1980).

Incidentally, when the same theory was applied to the laser interferometric detectors, the notion of "squeezed light" was developed (Caves, 1981). Although it had been discussed before, GW detectors stimulated the current strong interest in the quantum optics and communications community in this technique for reducing photon noise.

Despite more than a decade of research into squeezing for bars, no practical scheme has emerged, and one cannot be hopeful that bars will break the 10^{-21} barrier. For the next five years, while interferometers are being constructed, the ultracryogenic bars will have the best chance of detecting GWs. If a supernova occurs in our Galaxy or in a nearby one, they may well succeed.

4.2.2 Laser-Interferometric Detectors.

An interferometer is designed to measure the relative difference in two optical paths. If one places an interferometer in the center of the ring in Fig. 1, with the ends of two perpendicular arms on the ring, then the subsequent motion of the ends relative to the center can be detected by interferometry.

The relative motions are small, however, and so the technical challenge is large. For the end mirrors to respond as free masses, they must be hung from supports and isolated from ground vibration. Because the tidal gravitational forces scale with distance, it is desirable to make the arms as long as possible. Present prototypes (see the various articles in Blair, 1991) are in the 10–40-m range, but detectors now under construction (see below) will be as long as 4 km. It is this ability to take advantage of the tidal scaling that gives interferometers the edge over bars.

Even over 4 km, a disturbance of 10^{-21} translates into a mirror motion of 4×10^{-16} cm, less than 0.01 fm. To sense average displacements of the surface of a macroscopic object to this accuracy with optical or infrared photons whose wavelengths are 12 orders of magnitude larger requires many photons. In turn, this requires excellent mirrors and high-power continuous-wave lasers. The physical

principles of such detectors have been reviewed by Giazotto (1989).

The detectors will be built in stages. The first stage will aim at a sensitivity to broadband bursts of 10^{-21} . The second stage will aim for 10^{-22} . Current designs anticipate mirrors with losses well below 10^{-4} per reflection, and neodymium-doped yttrium aluminum garnet (YAG) lasers as sources for the second stage. (See LASERS, SOLID STATE.) For the second stage, light will have to be conserved, and designers like R. W. P. Drever and the late B. J. Meers have devised clever methods of optimizing the use of light. It is now clear that, although the detectors are intrinsically broadband, they can also be tuned to narrow frequency ranges if desired.

The light must travel along the arms in a good vacuum, better than 10^{-8} Torr for the second stage. This is not hard in principle, but the volume to be evacuated is large: tubes 4 km in length and 1.5 m in diameter. The cost of this vacuum system is the dominant cost of the detectors.

Data will flow from these detectors at an enormous rate. Recent short observing runs using prototypes at Glasgow University and the Max Planck Institute for Quantum Optics in Garching, Germany, produced data rates approaching 1 Gbyte/h. The storage and analysis of data from a network of interferometers are formidable problems.

There are proposals for a number of interferometric detectors around the world. A project called LIGO to build two in the United States (Abramovici *et al.*, 1992) is funded, and should begin construction in 1993. If present (1993) schedules hold, LIGO could begin observing by 1998. A French-Italian detector called VIRGO, to be built near Pisa, Italy, is likely to get final approval in 1993 (Bradaschia *et al.*, 1990). There are further proposals for a British-German detector near Hannover, Germany, and an Australian detector near Perth that are awaiting funding. Japan is planning an intermediate-sized detector (300 m). Because detectors have broad quadrupolar antenna patterns, an accurate direction to a detected source can only be obtained by triangulation, using the time delays among various detectors. For this reason, a minimum of three detectors worldwide is required for the extraction of full information from detected signals.

4.2.3 Space-Based Detectors. As we noted in Sec. 4.1.1, the interesting frequency range around 1 mHz is only accessible to space-based detectors. Space-based searches for GWs have already been made using transponding data from interplanetary space probes. A passing GW would affect the time delays of round-trip signaling to the probes; the signature of this effect is unique, and sensitive searches can be made at very low frequencies. No positive detections have been reported (e.g., Armstrong *et al.*, 1987). As the principal limit on sensitivity is propagation disturbances of the radio communication signals, sensitivity can be increased by using multiple-frequency communication, by going to higher frequencies, and by using multiple spacecraft. A three-spacecraft experiment was performed in 1993 using NASA's Mars Observer and Galileo probes and the European Space Administration's Ulysses solar observatory. When the data are analyzed, they may improve previous limits by a factor of 10 or more.

Proposals have been made for purpose-built interferometers in space. Laser interferometers using the Earth-Moon or even the Sun-Earth Lagrangian points to stabilize their orbits look very promising.

GLOSSARY

Cosmology: The study of the large-scale properties of the observable universe as a whole.

Geodesics: Locally straight lines of a curved space or spacetime. They are also paths of extremal proper distance.

Homogeneous Cosmology: A cosmology in which there is a choice of time (a reference frame) such that all its properties are the same everywhere at a constant time.

Isotropic Cosmology: A cosmology that is homogeneous and in which, in addition, there is no preferred direction in space, such as a systematic velocity of the matter.

Reference Frame: A coordinate system for space and time that includes facilities for measuring and recording the exact location and time of any event. In special relativity, a *Lorentz* reference frame has rigid spacings between spatial coordinate locations, and its clocks are synchronized everywhere. In general relativity, a *local* reference frame makes measurements only in a restricted region.

List of Works Cited

- Abramovici, A., Althouse, W. E., Drever, R. W. P., Gürsel, Y., Kawamura, S., Raab, F. J., Shoemaker, D., Sievers, L., Spero, R. E., Thorne, K. S., Vogt, R. E., Weiss, R., Whitcomb, S. E., Zucker, M. E. (1992), *Science* **256**, 325–333.
- Adelberger, E. G., Heckel, B. R., Stubbs, C. W., Rogers, W. F. (1991), *Annu. Rev. Nucl. Particle Sci.* **41**, 269–320.
- Allen, B., Shellard, E. P. S. (1992), *Phys. Rev. D* **45**, 1898–1912.
- Armstrong, J. W., Estabrook, F. B., Wahlquist, H. D. (1987), *Astrophys. J.* **318**, 536–541.
- Astone, P., Bassan, M., Bates, S., Bizzarri, R., Bonifazi, F., Cardarelli, R., Cavallari, G., Coccia, E., Degasperis, A., Depedis, D., Frasca, S., Majorana, E., Merucci, L., Modena, I., Muratori, G., Pallottino, G. V., Patrignani, C., Pizzella, G., Price, M., Rapagnani, P., Ricci, F., Visco, M. (1991a), *Z. Phys. C* **50**, 21–29.
- Astone, P., Bassan, M., Bonifazi, F., Bronzini, F., Castellano, M. G., Coccia, E., Cosmelli, C., Fafone, V., Frasca, S., Majorana, E., Modena, I., Pallottino, G. V., Pizzella, G., Rapagnani, P., Ricci, F., Visco, M. (1991b), *Europhys. Lett.* **16**, 231–235.
- Bender, P., Faller, J. E., Hall, J. L., Hils, D., Stebbins, R. T., Vincent, M. A. (1989), in: R. W. Hellings (Ed.), *Relativistic Gravitational Experiments in Space*, NASA Conference Publication 3046, Washington, D.C.: NASA, pp. 80–88.
- Birrell, N. D., Davies, P. C. W. (1982), *Quantum Fields in Curved Space*, New York: Cambridge University Press.
- Blair, D. G. (Ed.) (1991), *The Detection of Gravitational Waves*, Cambridge, U.K.: Cambridge University Press.
- Blandford, R. D., Znajek, R. L. (1977), *Mon. Not. R. Astron. Soc.* **179**, 433–456.
- Blandford, R. D., Begelman, M. C., Rees, M. J. (1982), *Sci. Am.*, May, 84–94.
- Blau, S. K., Guth, A. H. (1987), in: S. W. Hawking, W. Israel (Eds.), *300 Years of Gravitation*, New York: Cambridge University Press, pp. 524–603.
- Bradaschia, C., DelFabbro, R., DiVirgilio, A., Giazotto, A., Kautzky, H., Montelatici, V., Passuello, D., Brillet, A., Cregut, O., Hello, P., Man, C. N., Manh, P. T., Marraud, A., Shoemaker, D., Vinet, J. Y., Barone, F., Difiore, L., Milano, L., Russo, G., Aguirregabiria, J. M., Bel, H., Duruisseau, J. P., Ledenmat, G., Tournenc, P., Capozzi, M., Longo, M., Lops, M., Pinto, I., Rotoli, G., Damour, T., Bonazola, S., Marck, J. A., Gourgoulon, Y., Holloway, L. E., Fuligni, F., Iafolla, V., Natale, G. (1990), *Nucl. Instrum. Methods Phys. A* **289**, 518–525.
- Braginski, V. B., Panov, V. I. (1971), *Zh. Eksp. Teor. Fiz.* **61**, 873–879 [*Sov. Phys. JETP* **34**, 463–466 (1972)].
- Brillet, A., Hall, J. L. (1979), *Phys. Rev. Lett.* **42**, 549–552.
- Brown, T. M., Christensen-Dalsgaard, J., Dziembowski, W. A., Goode, P., Gough, D. O., Morrow, C. A. (1989), *Astrophys. J.*, **343**, 526–546.
- Casares, J., Charles, P. A., Naylor, T. (1992), *Nature* **355**, 614–617.
- Caves, C. M. (1981), *Phys. Rev. D* **23**, 1693–1708.
- Caves, C. M., Thorne, K. S., Drever, R. W. P., Sandberg, V. D., Zimmerman, M. (1980), *Rev. Mod. Phys.* **52**, 341–392.
- Chandrasekhar, S. (1939), *An Introduction to the Study of Stellar Structure*, Chicago: University of Chicago Press. Reprint (1958), New York: Dover.
- Chupp, T. E., Hoare, R. J., Loveman, R. A., Oteiza, E. R., Richardson, J. M., Wagshul, M. E., Thompson, A. K. (1989), *Phys. Rev. Lett.* **63**, 1541–1545.
- Ciufolini, I. (1989), *Int. J. Mod. Phys. A* **4**, 3083–3145.
- Dicke, R. H., Goldenberg, H. M. (1974), *Astrophys. J. Suppl.* **27**, 131–182.
- Dickey, J. O., Newhall, X. X., Williams, J. G. (1989), *Adv. Space Res.* **9**, 75–78.
- Drever, R. W. P. (1961), *Philos. Mag.* **6**, 683–687.
- Dyson, F. W., Eddington, A. S., Davidson, C. (1920), *Philos. Trans. R. Soc. London* **220A**, 291–333.
- Einstein, A. (1915), *Preuss. Akad. Wiss. Berlin, Sitzber.* 844–847.
- Everitt, C. W. F. (1988), in: J. D. Fairbank, B. S. Deaver, C. W. F. Everitt, P. F. Michelson (Eds.), *Near Zero: New Frontiers of Physics*, New York: Freeman, pp. 587–639.
- Fischbach, E., Talmadge, C. (1992), *Nature* **356**, 207–215.
- Fischbach, E., Sudarsky, D., Szafer, A., Talmadge, C., Aronson, S. H. (1986), *Phys. Rev. Lett.* **56**, 3–6.
- Fomalont, E. B., Sramek, R. A. (1976), *Phys. Rev. Lett.* **36**, 1475–1478.
- Giazotto, A. (1989), *Phys. Rep.* **182**, 365–425.
- Guth, A. H., Steinhardt, P. J. (1984), *Sci. Am.*, May, 90–102.
- Hartle, J. B. (1978), *Phys. Rep.* **46**, 201.
- Hawking, S. W. (1974), *Nature* **248**, 30–31.
- Hawking, S. W., Ellis, G. F. R. (1973), *The Large-Scale Structure of Space-Time*, New York: Cambridge University Press.
- Hilbert, D. (1915), *Goett. Nachr.* 395–407.
- Hughes, V. W., Robinson, H. G., Beltrán-Lopez, V. (1960), *Phys. Rev. Lett.* **4**, 342–344.
- Hulse, R. A., Taylor, J. H. (1975), *Astrophys. J. Lett.* **195**, L51–L53.
- Isaacson, R. (1968), *Phys. Rev.* **166**, 1272–1280.
- Jekeli, C., Eckhardt, D. H., Romaides, A. J. (1990), *Phys. Rev. Lett.* **64**, 1204–1206.
- Knapp, J., Kormendy, J. (1986), *Dark Matter in the Universe: I. A. U. Symposium Proceedings No. 117*, Dordrecht: Reidel.
- Krisher, T. P., Anderson, J. D., Campbell, J. K. (1990), *Phys. Rev. Lett.* **64**, 1322–1325.
- Lamoreaux, S. K., Jacobs, J. P., Heckel, B. R.,

- Raab, F. J., Fortsen, E. N. (1986), *Phys. Rev. Lett.* **57**, 3125–3128.
- Landau, L. D., Lifshitz, E. M. (1962), *The Classical Theory of Fields*, Oxford: Pergamon.
- Mashhoon, B., Paik, H. J., Will, C. M. (1989), *Phys. Rev. D* **39**, 2825–2838.
- Meegan, C. A., Fishman, G. J., Wilson, R. B., Paciesas, W. S. (1992), *Nature* **355**, 143–145.
- Misner, C. W., Thorne, K. S., Wheeler, J. A. (1973), *Gravitation*, San Francisco: Freeman.
- Müller, J., Schneider, M., Soffel, M., Ruder, H. (1991), *Astrophys. J. Lett.* **382**, L101–L103.
- Narayan, R., Piran, T., Shemi, A. (1991), *Astrophys. J. Lett.* **379**, L17–L20.
- Nordtvedt, K., Jr. (1968a), *Phys. Rev.* **169**, 1014–1016.
- Nordtvedt, K., Jr. (1968b), *Phys. Rev.* **169**, 1017–1025.
- Penzias, A. A., Wilson, R. W. (1965), *Astrophys. J.* **142**, 419–421.
- Phinney, E. S. (1991), *Astrophys. J. Lett.* **380**, L17–L19.
- Pound, R. V., Rebka, G. A., Jr. (1960), *Phys. Rev. Lett.* **4**, 337.
- Prestage, J. D., Bollinger, J. J., Itano, W. M., Wineland, D. J. (1985), *Phys. Rev. Lett.* **54**, 2387–2390.
- Reasenber, R. D., Shapiro, I. I., MacNeil, P. E., Goldstein, R. B., Breidenthal, J. C., Brenkle, J. P., Cain, D. L., Kaufman, T. M., Komarek, T. A., Zygielbaum, A. I. (1979), *Astrophys. J. Lett.* **234**, L219–L221.
- Rees, M. J. (1990), *Sci. Am.*, **26** (5), 56–66.
- Robertson, D. S., Carter, W. E., Dillinger, W. H. (1991), *Nature* **349**, 768–770.
- Roll, P. G., Krotkov, R. V., Dicke, R. H. (1964), *Ann. Phys. (NY)* **26**, 442–517.
- Schutz, B. F. (1984), *Am. J. Phys.* **52**, 412–419.
- Schutz, B. F. (1985), *A First Course in General Relativity*, New York: Cambridge University Press.
- Schutz, B. F. (1986), *Nature* **323**, 310–311.
- Schutz, B. F. (1991), in: D. G. Blair (Ed.), *The Detection of Gravitational Waves*, Cambridge: Cambridge University Press, pp. 406–452.
- Shapiro, I. I. (1964), *Phys. Rev. Lett.* **13**, 789–791.
- Shapiro, I. I. (1990), in: N. Ashby, D. F. Bartlett, W. Wyss (Eds.), *General Relativity and Gravitation*, Cambridge, U.K.: Cambridge University Press, pp. 313–330.
- Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., Deamici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J., Meyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., Wilkinson, D. T. (1992), *Astrophys. J. Lett.* **396**, L1–L5.
- Taylor, J. H., Wolszczan, A., Damour, T., Weisberg, J. M. (1992), *Nature* **355**, 132–136.
- Thorne, K. S. (1991), *Ann. N. Y. Acad. Sci.* **631**, 182–193.
- Thorne, K. S., Price, R. H., Macdonald, D. A. (1986), *Black Holes: the Membrane Paradigm*, New Haven: Yale University Press.
- Tutukov, A. V., Yungelson, L. R. (1993), *Mon. Not. R. Astron. Soc.* **260**, 675–678.
- Vessot, R. F. C. (1989), *Adv. Space. Res.* **9**, 21–28.
- Vessot, R. F. C., Levine, M. W., Mattison, E. M., Blomberg, E. L., Hoffman, T. E., Nystrom, J. W., Farrel, B. F., Decher, R., Eby, P. B., Baugher, C. R., Watts, J. W., Teuber, D. L., Wills, F. O. (1980), *Phys. Rev. Lett.* **45**, 2081–2084.
- Wald, R. M. (1984), *General Relativity*, Chicago: Univ. of Chicago Press.
- Walsh, D., Carswell, R. E., Weymann, R. J. (1979), *Nature* **279**, 381.
- Weber, J. (1960), *Phys. Rev.* **117**, 306–313.
- Weber, J. (1967), *Phys. Rev. Lett.* **18**, 498–501.
- Will, C. M. (1990), *Sky Telescope* **80**, 472–479.
- Will, C. M. (1993), *Theory and Experiment in Gravitational Physics*, 2nd ed., Cambridge, U.K.: Cambridge University Press.
- Worden, P. W. (1988) in: J. D. Fairbank, B. S. Deaver, C. W. F. Everitt, P. F. Michelson (Eds.), *Near Zero: New Frontiers of Physics*, New York: Freeman, pp. 766–783.

Further Reading

- Blair, D. G. (Ed.) (1991), *The Detection of Gravitational Waves*, Cambridge, U.K.: Cambridge Univ. Press.
- Davies, P. C. W. (1980), *The Search for Gravity Waves*, Cambridge, U.K.: Cambridge Univ. Press.
- Hawking, S. W., Israel, W. (Eds.) (1987), *300 Years of Gravitation*, Cambridge, England: Cambridge Univ. Press.
- Hoffmann, B. (1983), *Relativity and its Roots*, New York: Freeman.
- Misner, C. W., Thorne, K. S., Wheeler, J. A. (1973), *Gravitation*, San Francisco: Freeman.
- Pais, A. (1982), *Subtle is the Lord: the Science and Life of Albert Einstein*, Oxford: Oxford University Press.
- Price, R. H. (1982), *Am. J. Phys.* **50**, 300–329.
- Schutz, B. F. (1980), *Geometrical Methods of Mathematical Physics*, Cambridge, U.K.: Cambridge Univ. Press.
- Schutz, B. F. (1985), *A First Course in General Relativity*, Cambridge, U.K.: Cambridge University Press.
- Shapiro, S. L., Teukolsky, S. A. (1983), *Black Holes, White Dwarfs, and Neutron Stars*, New York: Wiley.
- Taylor, E. F., Wheeler, J. A. (1992), *Spacetime Physics*, 2nd ed., New York: Freeman.
- Wheeler, J. A. (1990), *A Journey into Gravity and Spacetime*, New York: Freeman.
- Will, C. M. (1986), *Was Einstein Right?*, New York: Basic.