Chapter 4
Putting General Relativity to the Test: Twentieth-Century Highlights and Twenty-First-Century Prospects

Clifford M. Will

4.1 Introduction

During the late 1960s, it was frequently said that "the field of general relativity is a theorist's paradise and an experimentalist's purgatory." The field was not without experiments, of course: Irwin Shapiro, then at MIT, had just measured the relativistic retardation of radar waves passing the Sun (an effect that now bears his name); Robert Dicke of Princeton was claiming that the Sun was flattened by rotation in an amount whose effects on Mercury's perihelion advance would put general relativity in jeopardy, and Joseph Weber of the University of Maryland was busy building gravitational wave antennas out of massive aluminum cylinders. Nevertheless the field was dominated by theory and by theorists. The field *circa* 1970 seemed to reflect Einstein's own attitudes: although he was not ignorant of experiment, and indeed had a keen insight into the workings of the physical world, he felt that the bottom line was the *theory*. As he once famously said, if experiment were to contradict the theory, he would have "felt sorry for the dear Lord."

Since that time the field has been completely transformed, and today experiment is a central component of gravitational physics and in some aspects is setting the agenda for the field. The breadth of current experiments, ranging from tests of classic general relativistic effects, to tests using gravitational waves, to ideas for testing the theory in astrophysical settings or on cosmic scales, attest to the ongoing vigor of experimental gravitation.

The great progress in testing general relativity during the latter part of the twentieth century featured three main themes:

C. M. Will (✉)
Department of Physics, University of Florida, Gainesville, FL, USA
e-mail: cmw@phys.ufl.edu

D. E. Rowe et al. (eds.), *Beyond Einstein*, Einstein Studies 14,
https://doi.org/10.1007/978-1-4939-7708-6_4

- The use of advanced technology. This included the high-precision technology associated with atomic clocks, laser and radar ranging, cryogenics, and delicate laboratory sensors, as well as access to space.
- The development of general theoretical frameworks. These frameworks allowed one to think beyond the narrow confines of general relativity itself, to analyze broad classes of theories, to propose new experimental tests, and to interpret the tests in an unbiased manner.
- The synergy between theory and experiment. To illustrate this, one needs only to note that the LIGO-Virgo Scientific Collaboration, engaged in one of the most important general relativity investigations – the detection of gravitational radiation – consists of over 1000 scientists. This is big science, reminiscent of high-energy physics, not general relativity!

Today, because of its elegance and simplicity, and because of its empirical success, general relativity is the standard model for our understanding of the gravitational interaction. Yet developments in particle theory and observations in cosmology suggest that it is probably not the entire story and that modifications of the basic theory may be required at some level. String theory generally predicts a proliferation of additional fields that could result in alterations of general relativity similar to that of the Brans-Dicke theory of the 1960s. In the presence of extra dimensions, the gravity that we feel on our four-dimensional "brane" of a higher-dimensional world could be somewhat different from a pure four-dimensional general relativity. And the observation that the expansion of the universe is accelerating has opened the possibility that modifications of general relativity on the largest scales might be required. However, any theoretical speculation along these lines *must* abide by the best current empirical bounds. Still, most of the current tests involve the weak-field, slow-motion limit of gravitational theory.

Putting general relativity to the test during the twenty-first century is likely to involve three main themes:

- Tests of strong-field gravity. These are tests of the nature of gravity near black holes and neutron stars, far from the weak-field regime of the solar system.
- Tests using gravitational waves. The detection of gravitational waves in 2015 has initiated a new form of astronomy, but it has also provided new tests of general relativity in the strong-field, radiative regime.
- Tests of gravity at extreme scales. The detected acceleration of the universe, the observed large-scale effects of dark matter, and the possibility of extra dimensions with effects on small scales have revealed how little is known about gravity on the largest and smallest scales.

In this paper we will review selected highlights of testing general relativity during the twentieth century and will discuss the potential for new tests in the twenty-first century. We begin in Section 4.2.1 with the "Einstein Equivalence Principle," which underlies the idea that gravity and curved spacetime are synonymous, and describe its empirical support. Section 4.2.2 describes solar system tests of gravity in terms of experimental bounds on a set of "parametrized post-Newtonian" (PPN) parameters.

In Section 4.2.3 we discuss tests of general relativity using binary pulsar systems. Section 4.3.1 describes tests of gravitational theory that can be carried out using observations of gravitational radiation, and Section 4.3.2 describes the possibility of performing strong-field tests of general relativity. Tests of gravity at cosmological and submillimeter scales are significant topics in their own right and are beyond the scope of this paper. Concluding remarks are made in Section 4.4. For further discussion of topics in this paper, and for references to the primary literature, the reader is referred to *Theory and Experiment in Gravitational Physics* (Will 1993) and to the "living" review articles by Will (2014), Stairs (2003), Psaltis (2008), Mattingly (2005), Yunes and Siemens (2013), and Gair et al. (2013).

4.2 Twentieth-Century Highlights

4.2.1 The Einstein Equivalence Principle

The Einstein equivalence principle (EEP) is a powerful and far-reaching principle, which states that (i) test bodies fall with the same acceleration independently of their internal structure or composition (weak equivalence principle or WEP); (ii) the outcome of any local non-gravitational experiment is independent of the velocity of the freely falling reference frame in which it is performed (local Lorentz invariance or LLI); and (iii) the outcome of any local non-gravitational experiment is independent of where and when in the universe it is performed (local position invariance or LPI).

The Einstein equivalence principle is central to gravitational theory, for it is possible to argue convincingly that if EEP is valid, then gravitation must be described by "metric theories of gravity," which state that (i) spacetime is endowed with a symmetric metric, (ii) the trajectories of freely falling bodies are geodesics of that metric, and (iii) in local freely falling reference frames, the non-gravitational laws of physics are those written in the language of special relativity.

General relativity is a metric theory of gravity but so are many others, including the Brans-Dicke theory.

To illustrate the high precisions achieved in testing EEP, we shall review tests of the weak equivalence principle, where one compares the acceleration of two laboratory-sized bodies of different composition in an external gravitational field. A measurement or limit on the fractional difference in acceleration between two bodies yields a quantity $\eta \equiv 2|a_1 - a_2|/|a_1 + a_2|$, called the "Eötvös ratio," named in honor of Baron von Eötvös, the Hungarian physicist whose experiments carried out with torsion balances at the end of the nineteenth century were the first high-precision tests of WEP (see Figure 4.1). Later classic experiments by Dicke and Braginsky in the 1960s and 1970s improved the bounds by several orders of magnitude. Additional experiments were carried out during the 1980s as part of a search for a putative "fifth force" that was motivated in part by a reanalysis

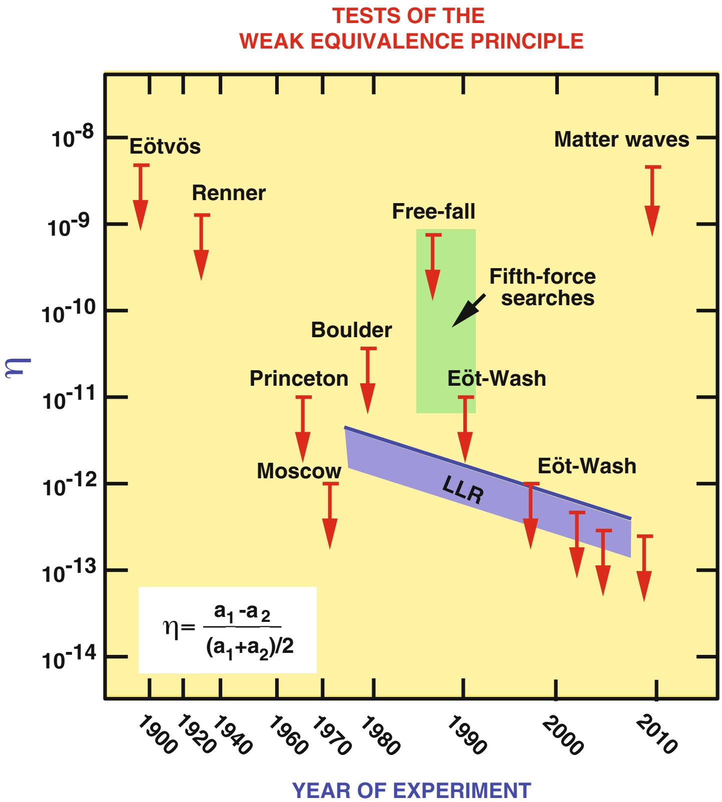

Fig. 4.1 Selected tests of the weak equivalence principle, showing bounds on the fractional difference in acceleration of different materials or bodies. Blue line and shading show evolving bounds on WEP for the Earth and the Moon from lunar laser ranging (LLR).

of Eötvös' original data (the range of bounds achieved during that period is shown schematically in the region labeled "fifth force" in Figure 4.1).

The best limit on η currently comes from the "Eöt-Wash" experiments carried out at the University of Washington, which used a sophisticated torsion balance tray to compare the accelerations of bodies of different compositions toward the Earth, the Sun, and the galaxy. Another strong bound comes from lunar laser ranging (LLR), which checks the equality of free fall of the Earth and Moon toward the Sun. The results from laboratory and LLR experiments are:

$$\eta_{\text{Eöt–Wash}} < 3 \times 10^{-13}\,, \quad \eta_{\text{LLR}} < 3 \times 10^{-13}\,. \tag{4.1}$$

In fact, by using laboratory materials whose composition mimics that of the Earth and Moon, the Eöt-Wash experiments permit one to infer an unambiguous bound from lunar laser ranging on the universality of acceleration of gravitational binding energy at the level of 9×10^{-4} (test of the Nordtvedt effect – see Section 4.2.2 and Table 4.1.)

High-precision WEP experiments can test superstring inspired models of scalar-tensor gravity or theories with varying fundamental constants in which weak violations of WEP can occur via nonmetric couplings. The project MICROSCOPE, designed to test WEP to a part in 10^{15}, was launched by the French space agency CNES in 2016. Other concepts for future improvements include advanced space experiments, experiments on suborbital rockets, lunar laser ranging, binary pulsar observations, and experiments with antihydrogen. For an update on past and future tests of WEP, see the series of articles introduced by Speake and Will (2012).

Very stringent constraints on local Lorentz invariance have been placed, notably by experiments that exploited laser-cooled trapped atoms to search for variations in the relative frequencies of different types of atoms as the Earth rotates around our velocity vector relative to the mean rest frame of the universe (as determined by the cosmic background radiation). For reviews, see Mattingly (2005), Will (2006), and Liberati (2013). Local position invariance has also been tested by gravitational redshift experiments and by tests of variations with cosmic time of fundamental constants. For a review of such tests, see Uzan (2011).

4.2.2 Solar System Tests

It was once customary to discuss experimental tests of general relativity in terms of the "three classical tests," the gravitational redshift, which is really a test of the EEP, not of general relativity itself; the perihelion advance of Mercury, the first success of the theory; and the deflection of light, whose measurement in 1919 made Einstein a celebrity. However, the proliferation of additional tests as well as of well-motivated alternative metric theories of gravity made it desirable to develop a more general theoretical framework for analyzing both experiments and theories.

This "parametrized post-Newtonian (PPN) framework" dates back to Eddington in 1922 but was fully developed by Nordtvedt and Will in the period 1968–1972. When we confine attention to metric theories of gravity and further focus on the slow-motion, weak-field limit appropriate to the solar system and similar systems, it turns out that, in a broad class of metric theories, only the values of a set of numerical coefficients in the expression for the spacetime metric vary from theory to theory. The framework contains ten PPN parameters: γ, related to the amount of spatial curvature generated by mass; β, related to the degree of nonlinearity in the gravitational field; ξ, α_1, α_2, and α_3, which determine whether the theory violates local position invariance or local Lorentz invariance in *gravitational* experiments (violations of the strong equivalence principle); and ζ_1, ζ_2, ζ_3, and ζ_4, which describe whether the theory has appropriate momentum conservation laws. In general relativity, $\gamma = 1$, $\beta = 1$, and the remaining parameters all vanish. For a complete exposition of the PPN framework, see Will (1993).

To illustrate the use of these PPN parameters in experimental tests, we cite the deflection of light by the Sun, an experiment that made Einstein an international celebrity when the sensational news of the Eddington-Clark eclipse measurements was relayed in November 1919 to a war-weary world. For a light ray which passes a distance d from the Sun, the deflection is given by

$$\begin{aligned} \Delta\theta &= \left(\frac{1+\gamma}{2}\right)\frac{4GM}{dc^2} \\ &= \left(\frac{1+\gamma}{2}\right) \times 1.7505 \left(\frac{R}{d}\right) \text{ arcsec}, \end{aligned} \tag{4.2}$$

where M and R are the mass and radius of the Sun and G and c are the Newtonian gravitational constant and the speed of light. The "1/2" part of the coefficient can be derived by considering the Newtonian deflection of a particle passing by the Sun, in the limit where the particle's velocity approaches c; this was first calculated independently by Henry Cavendish around 1784 and Johann von Soldner around 1803. The second "$\gamma/2$" part comes from the bending of "straight" lines near the Sun relative to lines far from the Sun, as a consequence of space curvature. A related effect, called the Shapiro time delay, an excess delay in travel time for light signals passing by the Sun, also depends on the coefficient $(1+\gamma)/2$.

To illustrate the dramatic progress of experimental gravity since the dawn of Einstein's theory, Figure 4.2 shows a history of results for $(1+\gamma)/2$. "Optical" denotes measurements using visible light, made mainly during solar eclipses, beginning with the 1919 measurements of Eddington and his colleagues. Arrows denote values well off the chart from one of the 1919 eclipse expeditions and from others through 1947. "Radio" denotes interferometric measurements of radio-wave deflection, and "VLBI" denotes very long baseline radio interferometry, culminating in a global analysis of VLBI data on over 540 quasars and compact radio galaxies distributed over the entire sky, which verified GR at the 0.02 percent level. "Hipparcos" denotes the European optical astrometry satellite. "Galactic lensing" denotes a 2006 measurement of γ using stellar velocity-dispersion measurements and gravitational lensing data on 15 elliptical galaxies taken from the Sloan Digital Sky Survey. The GAIA astronomical observatory, launched in 2013, is a high-precision astrometric telescope (a successor to Hipparcos), which could, among other science goals, measure light deflection and γ to the 10^{-6} level.

Shapiro time delay measurements began in the late 1960s, by bouncing radar signals off Venus and Mercury; the most recent test used tracking data from the *Cassini* spacecraft on its way to Saturn, yielding a result at the 0.001 percent level.

Other experimental bounds on the PPN parameters, all consistent with general relativity, came from measurements of the perihelion shift of Mercury, bounds on the "Nordtvedt effect" (a possible violation of the weak equivalence principle for self-gravitating bodies) via lunar laser ranging, and pulsar measurements. Table 4.1 summarizes the current bounds.

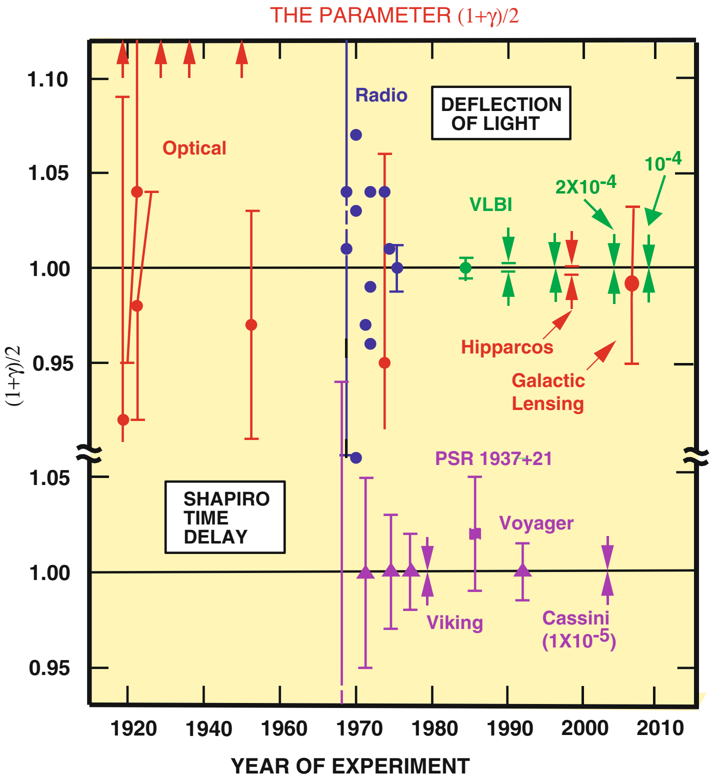

Fig. 4.2 Measurements of the coefficient $(1+\gamma)/2$ from observations of the deflection of light and of the Shapiro delay in propagation of radio signals near the Sun. The general relativity prediction is unity.

The perihelion advance of Mercury, the first of Einstein's successes, is now known to agree with observation to better than a part in 10^4, largely from improved data on the planet's orbit provided by the Mercury Messenger orbiter. Although there was controversy during the 1960s about this test because of Dicke's claims of an excess solar oblateness $J_{2\odot}$, which would result in an unacceptably large Newtonian contribution to the perihelion advance, it is now known from helioseismology that the oblateness is 2.2×10^{-7}, as expected from standard solar models, and too small to affect Mercury's orbit, within the experimental error.

Scalar-tensor theories of gravity are characterized by a coupling function $\omega(\phi)$ whose size is inversely related to the "strength" of the scalar field relative to the metric. In the solar system, the parameter $|\gamma - 1|$, for example, is equal to $1/(2 + \omega(\phi_0))$, where ϕ_0 is the value of the scalar field today outside the solar system. Solar system experiments (primarily the Cassini results) constrain $\omega(\phi_0) > 40000$.

Future space missions could provide improved measurements of PPN parameters. BepiColombo, a Mercury orbiter mission planned for launch around 2018,

Table 4.1 Current limits on the PPN parameters

Parameter	Effect	Limit	Remarks
$\gamma - 1$	(i) time delay	2.3×10^{-5}	Cassini tracking
	(ii) light deflection	2×10^{-4}	VLBI
$\beta - 1$	(i) perihelion shift	8×10^{-5}	$J_{2\odot} = (2.2 \pm 0.1) \times 10^{-7}$
	(ii) Nordtvedt effect	2.3×10^{-4}	$\eta = 4\beta - \gamma - 3$ assumed
ξ	spin precession	4×10^{-9}	millisecond pulsars
α_1	orbital polarization	10^{-4}	lunar laser ranging
		4×10^{-5}	PSR J1738+0333
α_2	spin precession	2×10^{-9}	millisecond pulsars
α_3	pulsar acceleration	2×10^{-20}	pulsar $\dot{P}$ statistics
ζ_1	–	2×10^{-2}	combined PPN bounds
ζ_2	binary acceleration	4×10^{-5}	$\ddot{P}_p$ for PSR 1913+16
ζ_3	Newton's 3rd law	10^{-8}	lunar acceleration
ζ_4	–	–	not independent

could, among other measurements, determine J_2 of the Sun to 10^{-8} and improve bounds on a time variation of the gravitational constant. The Laser Astrometric Test of Relativity (LATOR) and the Astrodynamical Space Test of Relativity using Optical Devices (ASTROD), involving laser ranging to one or more satellites on the far side of the Sun, could measure γ to a part in 10^8 and could possibly detect second-order effects in light propagation. The Apache Point Observatory for Lunar Laser-ranging Operation (APOLLO) project, a joint effort by researchers from the Universities of Washington, Seattle, and California, San Diego, is using enhanced laser and telescope technology, together with a good, high-altitude site in New Mexico, with the goal of improving the lunar laser-ranging bound by as much as an order of magnitude.

The NASA mission called Gravity Probe B completed its measurement of the Lense-Thirring and geodetic precessions of gyroscopes in Earth orbit. Launched on April 20, 2004, for a 16-month mission, it consisted of four spherical fused quartz rotors coated with a thin layer of superconducting niobium, spinning at 70–100 Hz, in a spacecraft containing a telescope continuously pointed toward a distant guide star (IM Pegasi). Superconducting current loops encircling each rotor measured the change in direction of the rotors by detecting the change in magnetic flux through the loop generated by the London magnetic moment of the spinning superconducting film. The spacecraft orbited the Earth in a polar orbit at 650 km altitude. In 2011, GPB reported a 20 percent measurement of the 41 milliarcsecond per year frame dragging or Lense-Thirring effect caused by the rotation of the Earth and a 0.3 percent measurement of the larger 6.6 arcsecond per year geodetic precession caused by space curvature (Everitt et al. 2011).

A complementary test of the Lense-Thirring precession involved measuring the precession of the orbital planes of two Earth-orbiting laser-ranged satellites called LAGEOS, using up-to-date models of the gravitational field of the Earth in order to subtract the dominant Newtonian precession with sufficient accuracy to yield a measurement of the relativistic effect, good to about 10 percent. Including data from a third laser-ranged satellite called LARES, launched in 2012, could improve the measurement, possibly to the one percent level.

As one final illustration of the precision and depth of the constraints on alternative theories of gravity, we mention the strange case of Alfred North Whitehead's 1922 theory of gravity. Uncomfortable with the fact that, in general relativity, the causal relationships in spacetime are not known *a priori* until the field equations have been solved, Whitehead proposed a Lorentz invariant theory of gravity in which the physical metric was constructed algebraically from a flat background Minkowski metric and matter variables (mass and velocities). The theory had the unusual property that, for a point mass at rest, the metric was mathematically equivalent to the Schwarzschild metric. As a consequence, Whitehead's theory could not be experimentally distinguished from general relativity using the standard tests of the day, such as light deflection, perihelion advance, or gravitational redshift. This led to a conundrum of how to select between competing theories that equally satisfy experimental observations.

But in 1971, we pointed out that when the theory was extended in a natural way to more than one gravitating body or to extended bodies, then the gravitational attraction between any pair of masses in the presence of a third body would be anisotropic, that is, dependent upon the orientation of the pair relative to the distant body (an effect additional to the normal tidal gravitational effects). This effective "anisotropy in Newton's constant G" would result in anomalous tide-like distortions of the Earth in the presence of the mass of the galaxy that were ruled out by precise measurements made with gravimeters. This however did not totally end the fascination with Whitehead's theory among some philosophically oriented scholars and so Gibbons and Will (2008) embarked on a serial killing of Whitehead's theory, pointing out that it actually fails the test of experiment in five different ways:

- *Anisotropy in G.* A reanalysis of Will's 1971 result verified that the theory violates gravimeter data on anomalous Earth tides by a factor of at least 100.
- *Nordtvedt effect.* The theory predicts a violation of the equivalence principle for gravitating bodies, leading to a Nordtvedt effect in lunar laser ranging 400 times larger than the data will permit.
- *Birkhoff's theorem and LAGEOS data.* It was already known in the 1950s that the theory predicts that the metric of a static, spherically symmetric *finite-sized* body has an additional size-dependent contribution. This contributes an additional advance of the perigee of the LAGEOS II satellite, in disagreement with observations by a factor of 10.
- *Momentum non-conservation.* The theory predicts an acceleration of the center of mass of a binary system, an effect ruled out by binary pulsar data by a factor of a million.

- *Gravitational radiation damping.* The theory predicts anti-damping in binary orbits due to gravitational radiation, in violation of binary pulsar data by four orders of magnitude.

The purpose of this list of failures is not to gang up on Whitehead but rather to illustrate that matching the Schwarzschild geometry is no longer sufficient to match experimental tests and that the current generation of empirical data strongly and deeply constrains the theoretical possibilities.

4.2.3 Binary Pulsars

The binary pulsar PSR 1913+16, discovered in 1974 by Joseph Taylor and Russell Hulse, provided important new tests of general relativity, specifically of gravitational radiation and of strong-field gravity. Through precise timing of the pulsar "clock," the important orbital parameters of the system were measured with exquisite precision. These include nonrelativistic "Keplerian" parameters, such as the eccentricity e, and the orbital period (at a chosen epoch) P_b, as well as a set of relativistic "post-Keplerian" parameters (see Table 4.2). The first PK parameter, $\langle\dot{\omega}\rangle$, is the mean rate of advance of periastron, the analogue of Mercury's perihelion shift. The second, denoted γ', is the effect of special relativistic time dilation and the gravitational redshift on the observed phase or arrival time of pulses, resulting from the pulsar's orbital motion and the gravitational potential of its companion. The third, $\dot{P}_b$, is the rate of decrease of the orbital period; this is taken to be the result of gravitational radiation damping (apart from a small correction due to galactic differential rotation). Two other parameters, s and r, are related to the Shapiro time

Table 4.2 Parameters of selected binary pulsars

Parameter	Symbol	Value[1] in PSR1913+16	Value[1] in J0737-3039
Keplerian parameters			
Eccentricity	e	0.6171334(5)	0.0877775(9)
Orbital period	P_b (day)	0.322997448911(4)	0.10225156248(5)
Post-Keplerian parameters			
Periastron advance	$\langle\dot{\omega}\rangle$ ($^\circ$yr^{-1})	4.226598(5)	16.8995(7)
Redshift/time dilation	γ' (ms)	4.2992(8)	0.386(3)
Orbital period derivative	$\dot{P}_b$ (10^{-12})	−2.423(1)	−1.25(2)
Shapiro delay (range)	r (μs)		6.2(3)
Shapiro delay ($\sin i$)	s		0.9997(4)

[1]Numbers in parentheses denote errors in last. digit

delay of the pulsar signal if the orbital inclination is such that the radio signal from the pulsar passes the companion in close proximity; s is a direct measure of the orbital inclination $\sin i$. According to general relativity, the first three post-Keplerian effects depend only on e and P_b, which are known, and on the two stellar masses which are unknown. By combining the observations of PSR 1913+16 with the general relativity predictions, one obtains both a measurement of the two masses and a test of the theory, since the system is overdetermined. The results are

$$m_1 = 1.4398 \pm 0.0002 M_\odot \,, \qquad m_2 = 1.3886 \pm 0.0002 M_\odot \,,$$
$$\dot{P}_b^{\rm GR} / \dot{P}_b^{\rm OBS} = 0.997 \pm 0.002 \,. \tag{4.3}$$

The accuracy in measuring the relativistic damping of the orbital period is now limited by uncertainties in our knowledge of the relative acceleration between the solar system and the binary system as a result of galactic differential rotation.

The results also test the strong-field aspects of metric gravitation in the following way: the neutron stars that comprise the system have very strong internal gravity, which contributes as much as several tenths of the rest mass of the bodies (compared to the orbital energy, which is only 10^{-6} of the mass of the system). Yet in general relativity, the internal structure is "effaced" as a consequence of the strong equivalence principle (SEP), a stronger version of EEP that includes *gravitationally* bound bodies and local *gravitational* experiments. As a result, the orbital motion and gravitational radiation emission depend *only* on the masses m_1 and m_2 and not on their internal structure, apart from standard tidal and spin-coupling effects. By contrast, in alternative metric theories, SEP is not valid in general, and internal structure effects can lead to significantly different behaviors, such as the emission of dipole gravitational radiation. Unfortunately, in the case of scalar-tensor theories of gravity, because the neutron stars are so similar in PSR 1913+16 (and in other double-neutron star binary pulsar systems), dipole radiation is suppressed by symmetry; the best bound on the coupling parameter $\omega(\phi_0)$ from PSR 1913+16 is in the hundreds. By contrast, the close agreement of the data with the predictions of general relativity constitutes a kind of "null" test of the effacement of strong-field effects in that theory. On the other hand, strong bounds on dipole radiation have been placed using pulsars with white-dwarf companions, including J1738+0333 and J11416545.

The "double pulsar" J0737-3039, discovered in 2003, is a binary system with two detected pulsars, in a 147-minute orbit seen almost edge on, with eccentricity $e = 0.09$, and a periastron advance of 17° per year (the secondary pulsar has since undergone sufficient precession of its spin axis that it no longer beams a signal toward the Earth). A variety of novel tests of relativity, neutron star structure, and pulsar magnetospheric physics have been carried out in this system. And because of its relative proximity to the Earth and favorable location in the galaxy, measurement of the orbital period decrease will not be limited by galactic acceleration effects, and so this system will eventually surpass the Hulse-Taylor system in a precision test of gravitational-wave damping.

The remarkable triple system, J0337+1715, was reported in 2014. It consists of a 2.73 millisecond pulsar ($M = 1.44 M_\odot$) with extremely good timing precision, accompanied by *two* white dwarfs in coplanar circular orbits. The inner white dwarf ($0.1975 M_\odot$) has an orbital period of 1.629 days, with $e = 6.918 \times 10^{-4}$, and the outer white dwarf ($0.41 M_\odot$) has a period of 327.26 days, with $e = 3.536 \times 10^{-2}$. This is an ideal system for testing the Nordtvedt effect in the strong-field regime. For reviews of binary pulsar tests, see Stairs (2003) and Damour (2009).

4.3 Twenty-First-Century Prospects

4.3.1 Gravitational-Wave Tests of Gravitation Theory

The 2015 detection of gravitational radiation by the ground-based laser interferometer observatory LIGO in the USA (Abbott et al. 2016) has ushered in a new era of gravitational-wave astronomy (for a pre-discovery review, see Hough and Rowan 2000). Over time, other advanced detectors – Virgo in Europe, KAGRA in Japan, and INDIGO in India – will join LIGO in a worldwide network of gravitational-wave observatories. The European Space Agency, with support from the NASA, plans to launch a space-based detector around 2030, nominally called LISA, to explore the low-frequency band of gravitational waves around a millihertz. And in the very low-frequency, nanohertz regime, coherent timing of millisecond pulsars distributed around the sky could also detect waves.

Furthermore, such observations promise to yield new and interesting tests of general relativity in its radiative regime. Indeed, the LIGO detection of waves from the inspiraling binary black hole system, dubbed GW150914, already yielded one such test.

This test concerns the speed of gravitational waves. According to general relativity, in the limit in which the wavelength of gravitational waves is small compared to the radius of curvature of the background spacetime, the waves propagate along null geodesics of the background spacetime, *i.e.,* they have the same speed, c, as light. In other theories, the speed could differ from c because of coupling of gravitation to “background” gravitational fields. For example, in some theories with a flat background metric $\boldsymbol{\eta}$, gravitational waves follow null geodesics of $\boldsymbol{\eta}$, while light follows null geodesics of $\boldsymbol{g}$. In brane-world scenarios, the apparent speed of gravitational waves could differ from that of light if the former can propagate off the brane into the higher-dimensional “bulk.” Another way in which the speed of gravitational waves could differ from c is if gravitation were propagated by a massive field (a massive graviton), in which case v_g would depend on the wavelength λ of the gravitational waves according to $v_g/c \approx 1 - \lambda^2/2\lambda_g^2$, where $\lambda_g = h/m_g c$ is the graviton Compton wavelength ($\lambda_g \gg \lambda$ is assumed).

The most obvious way to test for a massive graviton is to compare the arrival times of a gravitational wave and an electromagnetic wave from the same event, *e.g.,* a supernova. For a source at a distance of hundreds of megaparsecs, and for a relative arrival time between the two signals of order seconds, the resulting bound would be of order $|1 - v_g/c| < 5 \times 10^{-17}$. A bound such as this cannot be obtained from solar system tests at current levels of precision.

However, there is a situation in which a bound on the graviton mass can be set using gravitational radiation alone, with no need for an electromagnetic counterpart. That is the case of the inspiraling compact binary, the final stage of evolution of systems like the binary pulsar, in which the loss of energy to gravitational waves has brought the binary to an inexorable spiral toward a final merger. Because the frequency of the gravitational radiation sweeps from low frequency at the initial moment of observation to higher frequency at the final moment, the speed of the gravitational waves emitted will vary, from lower speeds initially to higher speeds (closer to c) at the end. This will cause a distortion of the observed phasing of the waves as a function of wavelength. Furthermore, through the technique of matched filtering, whereby a theoretical model of the wave and its phase evolution is cross-correlated against the detected signal, the parameters of the compact binary can be measured accurately, along with the parameter, the graviton Compton wavelength, that governs the distortion.

This analysis was performed on the signal from GW150914, the merger of two black holes, yielding a lower bound on λ_g of the order of several times 10^{13} km, slightly better than the bound $\lambda_g > 2.8 \times 10^{12}$ km, derived from solar system dynamics, which limits the presence of a Yukawa modification of Newtonian gravity of the form $V(r) = (GM/r)\mathrm{e}(-r/\lambda_g)$. The space-based LISA antenna could place a bound on λ_g on the order of 10^{16} km.

Another test that will be possible involves verifying the polarization content of the waves; general relativity predicts only two polarization modes out of a possible six, irrespective of the source.

4.3.2 Tests of Gravity in the Strong-Field Regime

One of the main difficulties of testing GR in the strong-field regime is the possibility of contamination by uncertain or complex physics. In the solar system, weak-field gravitational effects can in most cases be measured cleanly and separately from non-gravitational effects. The remarkable cleanliness of most binary pulsars permitted precise measurements of gravitational phenomena in a strong-field context. Unfortunately, nature is rarely so kind. Still, under suitable conditions, qualitative and even quantitative strong-field tests of GR could be carried out.

One example is the exploration of the spacetime near black holes and neutron stars via accreting matter. Studies of certain kinds of accretion known as advection-dominated accretion flow (ADAF) in low-luminosity binary X-ray sources may yield the signature of the black hole event horizon. The spectrum of frequencies of quasiperiodic oscillations (QPO) from galactic black hole binaries may permit measurement of the spins of the black holes. Aspects of strong-field gravity and frame dragging may be revealed in spectral shapes of iron fluorescence lines from the inner regions of accretion disks around black holes and neutron stars. Measurements of the detailed shape of the infrared image of the accretion flow around the 4 million solar-mass black hole SgrA* in the center of our galaxy or of the orbits of stars or pulsars in close proximity to the hole could also provide tests of the spacetime black hole metric. Because of uncertainties in the detailed models, the results to date of studies like these are suggestive at best, but the combination of higher-resolution observations and better modeling could lead to striking tests of strong-field predictions of GR. For a review of such tests, see Psaltis (2008).

The best tests of GR in the strong-field limit may come from gravitational-wave observations (see Yunes and Siemens (2013) and Gair et al. (2013) for reviews). The ground-based interferometers have detected and will continue to detect the gravitational waves from the final inspiral and merger of pairs of stellar-mass black holes or neutron stars. Comparison of the observed waveform with the predictions of GR from a combination of analytic and numerical techniques can test the theory in the most dynamical, strong-field limit. In fact a preliminary test of this kind was carried out using data from the *second* LIGO detection, GW151226. The space antenna LISA may observe as many as 100 mergers of massive black holes per year, with large signal-to-noise ratio. Such observations could provide precise measurements of black hole masses and spins and could test the "no hair" theorems of black holes by detecting the spectrum of quasi-normal "ringdown" modes emitted by the final black hole. Observations by LISA of the hundreds of thousands of gravitational-wave cycles emitted when a small black hole inspirals onto a massive black hole could test whether the geometry of a black hole actually corresponds to the "hair-free" Kerr metric.

4.4 Conclusions

Einstein's general theory of relativity altered the course of science. It was a triumph of the imagination and of theory, but in the early years following its formulation, experiment played a secondary role. In the final four decades of the twentieth century, we witnessed a second triumph for Einstein, in the systematic, high-precision experimental verification of general relativity. It has passed every test with flying colors.

But the work is not done. During the twenty-first century, we may look forward to the possibility of tests of strong-field gravity in the vicinity of black holes and neutron stars. Electromagnetic and gravitational-wave astronomy will play a critical role in probing this largely unexplored aspect of general relativity.

General relativity is now the "standard model" of gravity. But as in particle physics, there may be a world beyond the standard model, beyond Einstein. Quantum gravity, strings, and branes may lead to testable effects beyond standard general relativity. Experimentalists and observers can be counted on to continue a vigorous search for such effects using laboratory experiments, particle accelerators, gravitational-wave detectors, space telescopes, and cosmological observations, well into the twenty-first century.

Acknowledgements This work was supported in part by the US National Science Foundation, Grant No. PHY 06-52448, and by the National Aeronautics and Space Administration, Grant No. NNG-06GI60G. The 2016 update of the article was supported by NSF Grant Nos. PHY 13-06069 and 16-00188. We are grateful for the hospitality of the Institut d'Astrophysique de Paris, where the initial version of this paper was prepared.

References

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., et al. (The LIGO-Virgo Scientific Collaboration)(2016). Observation of gravitational waves from a binary black hole merger. *Physical Review Letters, 116*, 061102.

Damour, T. (2009). Binary systems as test-beds of gravity theories. In M. Colpi, P. Casella, V. Gorini, U. Moschella, & A. Possenti, *Physics of relativistic objects in compact binaries: From birth to coalescence: Astrophysics and space science library* (Vol. 359, p. 1). New York: Springer.

Everitt, C. W. F., DeBra, D. B., Parkinson, B. W., Turneaure, J. P., Conklin, J. W., Heifetz, M. I., et al. (2011). Gravity probe B: Final results of a space experiment to test general relativity. *Physical Review Letters, 106*, 221101.

Gair, J. L., Vallisneri, M., Larson, S. L., & Baker, J. G. (2013). Testing general relativity with low-frequency, space-based gravitational-wave detectors. *Living Reviews in Relativity, 16*, 7 [Online article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2013-7.

Gibbons, G., & Will, C. M. (2008). On the multiple deaths of Whitehead's theory of gravity. *Studies in History and Philosophy of Modern Physics, 39*, 41–61.

Hough, J., & Rowan, S. (2000). Gravitational wave detection by interferometry (ground and space). *Living Reviews in Relativity, 3*, 3 [On-line article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2000-3.

Liberati, S. (2013). Tests of Lorentz invariance: A 2013 update. *Classical Quantum Gravity, 30*, 133001.

Mattingly, D. (2005). Modern tests of Lorentz invariance. *Living Reviews in Relativity, 8*, 5 [On-line article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2005-5.

Psaltis, D. (2008). Probes and tests of strong-field gravity with observations in the electromagnetic spectrum. *Living Reviews in Relativity, 11*, 9 [On-line article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2008-9.

Speake, C. C., & Will, C. M. (2012). Tests of the weak equivalence principle. *Classical and Quantum Gravity, 29*, 180301.

Stairs, I. H. (2003). Testing general relativity with pulsar timing. *Living Reviews in Relativity, 6*, 5 [On-line article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2003-5.

Uzan, J.-P. (2011). Varying constants, gravitation and cosmology. *Living Reviews in Relativity, 14*, 2 [Online article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2011-2.

Will, C. M. (1993). *Theory and experiment in gravitational physics*. Cambridge: Cambridge University Press.

Will, C. M. (2006). Special relativity: A centenary perspective. In T. Damour, O. Darrigol, B. Duplantier & V. Rivasseau (Eds.), *Einstein 1905-2005: Poincaré Seminar 2005* (p. 33). Basel: Birkhäuser.

Will, C. M. (2014). The confrontation between general relativity and experiment. *Living Reviews in Relativity, 17*, 4 [Online article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2014-4.

Yunes, N., & Siemens, X. (2013). Gravitational-wave tests of general relativity with ground-based detectors and pulsar-timing arrays. *Living Reviews in Relativity, 16*, 9 [Online article]: Cited 15 Nov. 2016, https://doi.org/10.12942/lrr-2013-9.