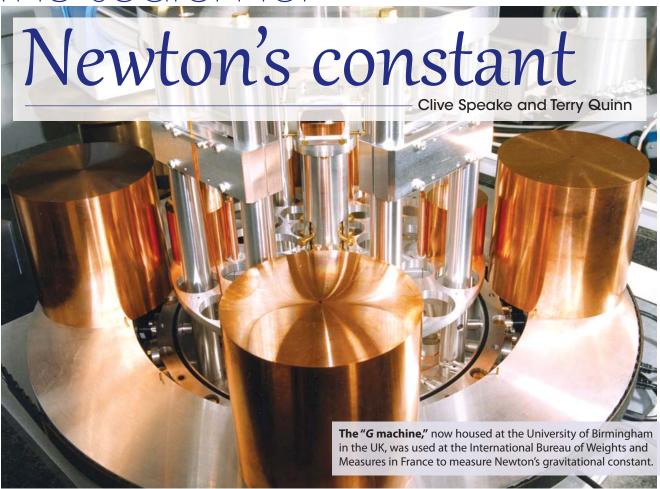
The search for



Three decades of careful experimentation have painted a surprisingly hazy picture of the constant governing the most familiar force on Earth.

ravity has a special place in physics. For starters, it is the only fundamental interaction that cannot be described by a quantum theory. Whereas the prevailing theories of gravity—Newton's law and Einstein's general relativity—consider space and time to be continuous classical quantities, the theories that describe electromagnetism and the nuclear forces are based on conserved quanta.

Gravity is also by far the weakest of the fundamental forces; its strength becomes comparable to that of the others only at energies near the Planck scale, 1.22×10^{19} GeV, some 15 orders of magnitude higher than the energies currently being explored by the Large Hadron Collider. The mismatch calls into question the validity of the standard model of particle physics, which is thought to be incompatible with such an immense fundamental energy scale.

It is fitting, then, that gravity, more than any other force, stubbornly eludes precise measurement. Newton's law, which approximates general relativity in the limit of small gravitational fields and nonrelativistic speeds, states that the magni-

tude F of the force attracting two spherical bodies of mass M_1 and M_2 , separated by a distance r, is given by $F = GM_1M_2/r^2$. The constant G is known, unsurprisingly, as Newton's constant of gravitation. It is considered to be a fundamental constant of nature. But more than three centuries after Newton's law was proposed, experiments have yet to yield a consensus on the constant's value.

According to the Committee on Data for Science and Technology (CODATA), which issues recommended values of fundamental constants once every four years, $G = 6.673\,84(80) \times 10^{-11}$ kg⁻¹ m³ s⁻². That value, from 2010, reflects the results of nearly a dozen experimental measurements made during the past three decades (see figure 1).¹ Although many of the individual measurements have an uncertainty of less than 50 parts per million (ppm),



Clive Speake is a professor of experimental physics in the school of physics and astronomy at the University of Birmingham in Birmingham, UK. **Terry Quinn** is emeritus director of the International Bureau of Weights and Measures in Sèvres, France.



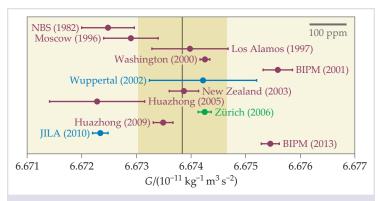


Figure 1. Measurements of Newton's gravitational constant *G* have yielded conflicting results. Here, the results of torsion-balance (maroon), pendulum (blue), and beam-balance (green) experiments discussed in the text are shown, along with the location and year in which they were measured. Error bars correspond to one standard deviation; the shaded region indicates the assigned uncertainty of the value recommended by the Committee on Data for Science and Technology in 2010. (Adapted from T. J. Quinn et al., *Phys. Rev. Lett.* **111**, 101102, 2013.)

their collective spread is nearly 10 times larger; it appears that we know G to only three significant figures! The apparent uncertainty is very large compared with that of other physical constants, many of which are known to a few parts in 10^8 . The Rydberg, which determines the electronic structure of atoms, is known to 4 parts in 10^{12} . (See the article by Peter Mohr and Barry Taylor, Physics Today, March 2001, page 29.)

Adding to the mystique is the fact that gravity is arguably the most familiar force to us here on Earth, and *G* can be taken as a measure of how well we understand it. It is not surprising that reported discrepancies between experimental determinations of Newton's constant of gravity can catch the public imagination. Why is *G* so poorly known? Before answering that question, let's consider why it's important to know *G* in the first place.

What is the value of *G*?

The actual numerical value of G is of little consequence to physics. For example, planetary orbits in our solar system are known to follow Newton's law and can be used along with G to estimate the mass of the Sun. Revising G upward by, say, 0.05% would simply reduce the Sun's estimated mass by the same fraction. At present we do not have models for the structure of the Sun that usefully constrain its mass at such small levels.

What matters, then, is not the value of the constant *G* but our ability to show that it is, in fact, constant. Abundant respectable theories predict violations of Newton's inverse square law at small length scales. Other theories predict violations of the equivalence principle—an empirical foundation of general relativity and, as such, a foundation of Newton's law—which states that the free acceleration of matter in a gravitational field does not depend on chemical composition. A growing view is that *G* may depend on matter density at astrophysical scales.

So far, both the equivalence principle² and

Newton's inverse square law3 have survived experimental scrutiny. To maximize sensitivity and to relieve the burden of metrology, however, those experimental tests are cleverly designed to give a substantial signal only if nature misbehaves in the way sought out by the experimentalists. Actual measurements of G must take stock of all relevant quantities in physical units and attack the metrology head on. Discrepant measurements of G may signal that we do not understand the metrology of measuring weak forces, which may, in turn, imply that the experimental tests establishing the inverse square law and the universality of free fall are flawed in some subtle fashion. Such a development would make for an exciting situation, which perhaps explains why so much popular interest is taken in such apparently mundane and painstaking work.

The historical perspective

The concept of a fundamental constant did not exist during Isaac Newton's time. He did not explicitly include a constant in his law; rather *G* was implied as if its value were equal to 1. Not until 1873 did Alfred Cornu and Baptistin Baille explicitly introduce a symbol for the gravitational coupling constant, which they called *f*. The constant did not take its current designation *G* until sometime in the 1890s.

The development of the concept of fundamental constants was intimately linked to the development of systems of physical units. David Newell, in his article on page 35, describes how, starting in 2018, the SI units will likely be based on fixed numerical values of seven fundamental constants, including the speed of light and Planck's constant. The latter will appear in the new definition of the kilogram.

Could we not define the kilogram in terms of G? For example, "the kilogram is the unit of mass, and its magnitude is set by fixing the numerical value of G equal to $6.67384 \times 10^{-11} \, \text{kg}^{-1} \, \text{m}^3 \, \text{s}^{-2}$." In theory, that would work fine, but in practice, any measurement of the mass of an object in terms of its gravitational attraction to another would have a precision of only a few hundred parts per million—about four orders of magnitude less than is needed in advanced metrology and can be obtained from a definition based on Planck's constant. Gravity is simply too weak at laboratory scales: The gravitational force between a pair of touching 1-kg copper spheres is roughly 10^{-8} N, about one thousandth of one millionth the weight of each.

To make even a rudimentary measurement of *G* with such a pair of spheres, one must find a way to nullify the overwhelming downward force of gravity without disturbing the spheres' sensitivity to their own mutual attraction. Near the end of the 18th century, John Michell discovered an elegant way to do just that: By placing two balls—the so-called test masses—at opposite ends of a horizontal beam suspended by a copper wire, as illustrated in figure 2a, one could neutralize Earth's downward pull while leaving the assembly free to rotate in the horizontal plane.

The assembly is, effectively, the rotational analogue of a mass suspended on a spring; assuming the wire behaves elastically, two appropriately po-

sitioned larger spheres (source masses) will cause the balance to rotate by an angle that depends on the gravitational force and the wire's torque constant κ . The torque constant can be determined by measuring the natural oscillation period T_0 of the torsion assembly and using $\kappa = I(2\pi/T_0)^2$, where I is the moment of inertia. Michell had invented the torsion balance.

Not until after Michell's death was the apparatus put to use: Henry Cavendish used it to "weigh" Earth by comparing the gravitational attraction between test and source masses with that between the source masses and Earth. Cavendish's 1798 publication describes in exquisite detail arguably the first precision experiment in physics. His torsion balance was one of the most significant pieces of physical apparatus ever invented. In a compilation of published work on measurements of *G*, George Gillies listed about 350 papers, almost all of which refer to work done with a torsion balance.⁴ Among the dozen or so experiments used in the latest CODATA evaluation, all except three were made with torsion balances.

Had Cavendish walked into any of the modern torsion-balance labs, he would have immediately known what was going on. Although today's torsion-balance assemblies are protected by vacuum chambers, not the wooden boxes used in the Cavendish experiment, the basic principle of separating the minute gravitational force between laboratory-scale masses from Earth's large, downward pull remains the same.

Cavendish would have been surprised, however, to find that after so many years, measurement accuracy has improved only modestly—not nearly as much as it has for almost every other physical quantity. We now estimate the accuracy of Cavendish's measurements to be something like 1%, which is not much worse than the spread of measurements that figure into the current CODATA value. To understand how we've arrived at this situation, let's first

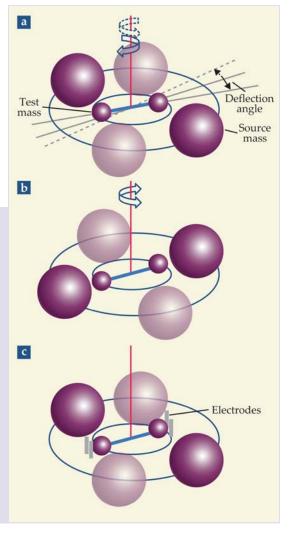
Figure 2. A torsion-balance experiment has, as its central element, two test masses balanced on a beam suspended by a thin metal wire. (a) In the original setup conceived by John Michell and later used by Henry Cavendish, two large source masses are positioned to exert a gravitational force that causes the torsion balance to turn through a small angle. The arrangements indicated by the dark and light source masses would yield clockwise and counterclockwise displacements, respectively. (b) In so-called time-ofswing experiments, G is calculated from the change in oscillation period when source masses are repositioned between arrangements lying along (dark spheres) and orthogonal to (light spheres) the resting test-mass axis. (c) In a third approach, the electrostatic servo-control technique, the gravitational force is calculated from the voltage that must be applied to nearby electrodes to hold the test assembly in place. In all three configurations, the gravitational coupling between the source masses and the whole of the torsion-balance assembly has to be calculated.

take a look at what actually has changed in the design and operation of torsion balances since the time of Cavendish.

Inspiration ...

One of the first important improvements to the Cavendish method was made in 1894 by Charles Boys, who realized that the best sensitivity would be obtained with the thinnest possible wire. That's because the torque constant increases as the fourth power of a wire's diameter, whereas the load the wire can support increases as the diameter squared. Although thinner wires require lighter test masses, the decreased gravitational force is more than compensated for by the increase in wire flexibility; the result is a larger, easier-to-measure deflection angle. Almost all torsion-balance experiments since the time of Boys have used a fine wire with a suspended mass assembly of a few grams at most. In Cavendish's original setup, the test masses were much larger, lead balls weighing some 750 grams each.

The next major advance was made in 1895 by Loránd Eötvös, who introduced the so-called time-of-swing method. In that approach, the free-oscillation period of the torsion-balance assembly is measured with the source masses positioned first along,



www.physicstoday.org July 2014 Physics Today

then orthogonal to, the test mass axis (see figure 2b). In the first configuration, gravitational attraction between the source and test masses decreases the period; in the second configuration, it increases the period. The advantage of the method is that a small change in oscillation period is easier to accurately measure than a small change in the deflection angle.

Techniques improved over the next half century, but the methods remained the same. By the 1970s Gabriel Luther of the US National Bureau of Standards (NBS, now NIST) in Gaithersburg, Maryland, and William Towler of the University of Virginia had used the time-of-swing method to measure G with an uncertainty of 70 ppm.5 That result was largely the basis for the value adopted in the CODATA's 1986 edition of fundamental constants. Later, Charles Bagley and Luther, at Los Alamos National Laboratory, repeated the NBS experiment using a different disposition of source masses.⁶ Around the same time, a team at Moscow's Tribotech Research and Development Co conducted a long series of time-of-swing measurements using various wires and various arrangements of source and test masses.7

Metrologists of the day had every reason to think that an uncertainty of 10 ppm was within reach. Attempts to improve estimates of *G* by using





extremely large masses such as mountains and reservoirs failed; although the gravitational signals were larger, so were the uncertainties, for instance, of the mass distribution inside the mountain and the shape of the bed of the reservoir. Nevertheless, there was little reason to think that the CODATA value was in serious error.

... and perspiration

During the 1990s two developments cast doubt on the 1986 CODATA value. First was the problem of anelasticity, the fact that the metal wire in a torsion balance doesn't behave as an ideal spring. A standard approach in materials science is to treat such a wire as a Maxwell material—essentially, a damped spring having both elastic and viscous components. The Maxwell model predicts an anelastic aftereffect, observed by Cavendish, wherein the spring takes a finite time to relax after an applied stress is removed.

In the early 1990s, we and our coworkers discovered that the standard Maxwell model doesn't fully explain the behavior of torsion balances. Specifically, our theory and experiments suggested that the torsion assemblies have not one characteristic relaxation time but many; the damping appears to grow stronger as the period becomes longer, and the relaxation time becomes essentially infinite.⁸ The effect is evident at periods ranging from less than a second to more than 10 minutes. We were able to relate the anelastic aftereffect to the presence of so-called 1/f noise arising from the movement of dislocations in the metal wire.

Kazuaki Kuroda then deduced that anelastic behavior would subject time-of-swing measurements to an error inversely proportional to the quality factor Q, a quantity indicating how closely the balance approximates a lossless elastic spring. He calculated corrections for many of the classic torsion-balance measurements; he revised all of them downward, in most cases by a few tenths of a percent. The NBS measurements on which the 1986 CODATA value was based were revised downward by about 50 ppm following confirmatory experiments by Bagley and Luther, who used two wires of widely different Q.

In 1996 a second development shook confidence in the CODATA value: the publication of a

Figure 3. Two twists on the torsion balance. **(a)** A group at the University of Washington used the flat plate visible at center, rather than the traditional dumbbell arrangement, as the test mass in a torsion-balance measurement of the gravitational constant *G*. (A penny at the bottom left conveys the scale.) In such a geometry, the derived value of *G* is almost completely independent of the mass distribution of the test masses. (Image courtesy of Jens Gundlach.) **(b)** Researchers at Huazhong University of Science and Technology in Wuhan, China, used a quartz slab as the test mass, which offers similar metrology advantages. The source masses are arranged in the so-called time-of-swing configuration, detailed in figure 2b. (Image courtesy of Jun Luo.)

Figure 4. A simple pendulum gravity gradiometer consists of a microwave or optical cavity formed by two hanging mirrors. When source masses are moved toward the cavity mirrors, the varying gravitational pull leads to a change in the cavity's optical length and, hence, a change in its resonant frequency. In a Fabry–Perot experiment performed at JILA, the change in the optical length was on the order of tens of nanometers.

result by Winfried Michaelis and coworkers at Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, Germany. Michaelis and his colleagues used a novel torsion balance in which the test masses were floated in a mercury bath rather than hung from a wire. Instead of measuring the displacement or change in period due to the gravitational pull of nearby source masses, the researchers used feedback control to apply an electrostatic torque just strong enough to hold the test masses in place (see figure 2c). The value of the applied voltage could then be used to infer *G*. Because there was no wire to twist, there were no anelastic effects.

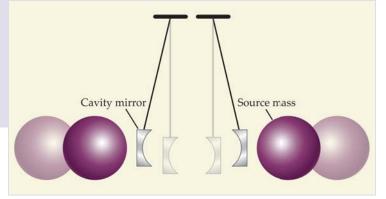
The PTB team's estimate of G differed from the accepted CODATA value at the time by some 0.7%, orders of magnitude larger than both the experiment's estimated uncertainty of around 80 ppm and the CODATA uncertainty of 127 ppm. Several groups, including one headed by one of us (Quinn) at the International Bureau of Weights and Measures (BIPM)—home of the international prototype of the kilogram—responded by embarking on their own G experiments.

The BIPM group discovered that the servocontrol technique could suffer significant errors if the electrostatic actuator was calibrated at a frequency different from the one at which it was used. Soon after, the other of us (Speake) suggested another likely source of error in the PTB experiment: A cross-capacitance term in the electrostatic calibration had been neglected. Subsequent studies at the PTB¹¹ confirmed that the omission did indeed cause measurements to overestimate *G* by about 0.7%.

Servo-control methods have since been used to make some of the most precise measurements of *G*. In 2003, Tim Armstrong and Mark Fitzgerald of the Measurement Standards Laboratory of New Zealand used the method to calculate *G* with an uncertainty of 40 ppm. ¹² The researchers used the inertial acceleration of a turntable-mounted torsion balance, rather than capacitance measurements, to calibrate their electrostatic actuator. Thus they elegantly avoided the problems encountered by the PTB workers.

Refinements and revisions

Among the biggest sources of uncertainty in a torsion-balance measurement are the source and test masses—one can be no more confident in *G* than in the properties of the objects used to measure it. Even small spatial variations in the density of a test mass can introduce sizeable error. In 2000 Jens Gundlach and Stephen Merkowitz of the University of Washington demonstrated a way around that problem.¹³ Instead of using the typical dumbbell-shaped test-



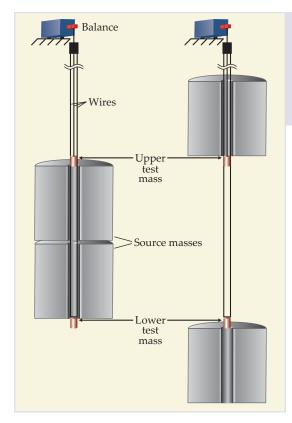
mass assembly, they used a thin, flat plate, pictured in figure 3a. The authors noted that the gravitational coupling between the test mass and neighboring spheres becomes proportional to the test mass's moment of inertia in the thin-plate limit. Because the value of *G* is calculated in terms of the ratio of the gravitational coupling to the test-mass moment of inertia, the test-mass density—and, to a good approximation, its uniformity—cancels out, provided that the field of the source masses is suitably tailored.

Gundlach and Merkowitz introduced another innovation, an adaptation of an idea that was developed by Jesse Beams for the NBS experiment but that wasn't fully exploited: They rotated their turntable-mounted torsion balance so that the plate experienced a sinusoidal gravitational coupling with the source masses. The researchers then compensated for that sinusoidal coupling by continually adjusting the turntable's rotation speed, until the wire experienced no torque. The value of G could then be inferred from the time-dependent acceleration profile of the turntable. To ensure that results weren't skewed by environmental gravity gradients, the source masses also rotated. The Washington experiment yielded the smallest uncertainty ever achieved in a G experiment, about 14 ppm. Their value, however, was significantly larger than the one obtained at NBS.

In an experiment at the Huazhong University of Science and Technology in Wuhan, China, Jun Luo and coworkers performed a time-of-swing measurement using a very long tungsten wire and a dumbbell-shaped mass assembly in which the spheres were set at different heights. ¹⁴ In their latest work, published in 2009, they followed the approach of the University of Washington group and used a solid quartz slab as a test mass—the advantage being that the slab's moment of inertia could be easily calculated to high precision (see figure 3b). Located deep inside a mountain, their lab has excellent thermal and seismic stability.

At the BIPM, we made two determinations of G, in 2001 and in 2013, using the apparatus shown on page $27.^{15}$ Instead of a traditional dumbbell-shaped test-mass assembly, we used a set of four 1-kg test masses set on the periphery of a 2-kg disk assembly suspended by a 160-mm-long, 30- μ m-thick, 2.5-mm-wide torsion strip.

Such a strip confers two important advantages. First, the restoring torque of a loaded torsion strip



has a dissipative component that depends largely on its thickness and a gravitational component that depends on its width and loading. A heavily loaded strip, much wider than it is thick, has a restoring torque that's almost wholly gravitational and hence essentially lossless. We could therefore obtain quality factors exceeding 105, meaning that it would take roughly 100 000 oscillations—or nearly five months, given the two-minute period of our balance—for the energy of our torsion balance to decay by 1/e. Second, because a torsion strip can support a much heavier load than can a wire with a similar quality factor, we were able to use four heavy test masses, which gives a bigger signal and significantly reduces the sensitivity of the balance to local gravity gradients. The resulting gravitational signal, 3×10^{-8} N·m, was some four orders of magnitude larger than in typical torsion-balance experiments.

Our experiments remain the only ones to measure *G* using the same apparatus in two significantly different methods: the classic Cavendish method, which essentially depends on an angle measurement and timing, and the servo-control method, which depends on electrical measurements. We believe such pairing of methods is a powerful way to check for systematic errors. If the results of the two methods agree, as ours did, then unknown errors in angle, timing, and electrical measurements are unlikely, and one need only look for errors in parameters that are common to both methods-mainly uncertainties in dimensional metrology and in the uniformity of source-mass densities. Had a different apparatus been used for each method—or had each experiment been performed in a different lab—errors could not have been constrained in the same way. Searching

Figure 5. In a beam-balance experiment, a Zürich team compared the weights of two 1.1-kg test masses suspended just above and just below 6.5-ton source masses. In switching between the left and right configurations, the test masses' differential weight changes by an amount equivalent to the weight of a millimeter-sized drop of water. (Adapted from ref. 18.)

for biases through a number of experimental configurations housed in the same laboratory and publishing a final result only when the measurements agree should lead to more reliable values of *G*.

Beyond the torsion balance

Since the 1990s a few groups have developed successful alternatives to the torsion balance. Among the firsts, researchers at the University of Wuppertal in Germany devised a simple pendulum gravity gradiometer, which consisted of two metal mirrors suspended by thin wires to form a hanging microwave cavity, as illustrated in figure 4. When 125-kg source masses were positioned behind each mirror, they induced a slight displacement of the mirrors, detectable as a change in the cavity resonance frequency.

By 2002 the Wuppertal group had refined the technique sufficiently to measure *G* with a reported uncertainty of 100 ppm. ¹⁶ Soon after, Harold Parks and James Faller of JILA adopted a similar approach, except they replaced the microwave cavity with a more sensitive optical cavity and used four source masses instead of two. ¹⁷ Their result, with an uncertainty of 21 ppm, was some 200 ppm smaller than the 2010 CODATA value.

In an experiment in Zürich, Stephan Schlamminger and colleagues measured *G* using the beambalance method¹⁸ depicted in figure 5. Conceptually similar to a method used by John Henry Poynting in the 1880s, the team's approach involved observing the change in the relative weights of two test masses suspended just above and just below two large source masses—steel containers each filled with 6.5 tons of mercury.

Despite the large source masses, the gravitational signal amounted to only 8 μ N. Although that's large compared with most torsion-balance experiments, it still represents a signal of only 800 μ g—roughly the mass of a millimeter-sized drop of water—superimposed on the roughly 1.1-kg weight of the test masses. The final uncertainty, less than 20 ppm, was constrained by the stability of the state-of-the-art commercial balance used to do the weighing. Although the resulting value sits squarely within the 2010 CODATA range, it differs significantly, sometimes by hundreds of parts per million, from more than half the measurements made during the past three decades.

Quo vadis?

What is one to make of all the disagreement? We mentioned above the possibility of modifications to Newton's laws. At present, however, none of the alternative theories seems compelling. More likely, the results have systematic errors much larger than their estimated uncertainties. Despite rigorous probing at

a meeting held this February at the Royal Society in the UK, no significant errors were uncovered in any of the experiments. Each group is confident of its results. What should we as a community do now?

The problem of arriving at a reliable value for *G* is unlikely to be resolved by one or two additional results obtained, as in the past, by teams working independently. Precise estimates of *G* rely on accurate measurements of such parameters as mass, density, length, time, electric current, voltage, capacitance, and angle. All of those measurements must be traced to verified national and international standards of the kilogram, meter, and second, with uncertainties evaluated with respect to the SI. In addition, future experiments must be carried out in laboratories having the highest quality of temperature and environmental control.

All those considerations strongly point to a national metrology institute, or a laboratory closely associated with one, as the most appropriate place for future *G* experiments. At the Royal Society meeting, attendees concluded that future efforts should be coordinated, so that our collective experience can be brought to bear on the design, construction, and operation of the new experiments. Strategies for achieving that goal include international collaborations. (Proposals will be discussed in more detail at an upcoming meeting at NIST; see http://pml.nist.gov/bigg.)

None of those developments, however, resolves the thorny question the CODATA Task Group must address in the forthcoming 2014 "Recommended values of the fundamental physical con-

stants": What is the best value and uncertainty to assign to *G*?

References

- P. J. Mohr, B. N. Taylor, D. B. Newell, Rev. Mod. Phys. 84, 1527 (2012).
- 2. Special issue, "Tests of the Weak Equivalence Principle," Class. Quantum Grav. 29 (2012).
- 3. R. Spero et al., Phys. Rev. Lett. 44, 1645 (1980).
- 4. G. T. Gillies, Metrologia 24(suppl.), 1 (1987).
- 5. G. G. Luther, W. R. Towler, Phys. Rev. Lett. 48, 121 (1982).
- C. H. Bagley, G. G. Luther, Phys. Rev. Lett. 78, 3047 (1997).
- O. V. Karagioz, V. P. Izmaylov, G. T. Gillies, *Grav. Cosmol.* 4, 239 (1998).
- T. J. Quinn, C. C. Speake, L. M. Brown, *Philos. Mag. A* 65, 261 (1992); C. C. Speake et al., *Meas. Sci. Technol.* 10, 430 (1999).
- 9. K. Kuroda, Phys. Rev. Lett. 75, 2796 (1995).
- W. Michaelis, H. Haars, R. Augustin, *Metrologia* 32, 267 (1995).
- W. Michaelis, J. Melcher, H. Haars, Metrologia 41, L29 (2004); 42, 67 (2005).
- T. R. Armstrong, M. P. Fitzgerald, *Phys. Rev. Lett.* 91, 201101 (2003).
- J. H. Gundlach, S. M. Merkowitz, *Phys. Rev. Lett.* 85, 2869 (2000).
- Z.-K. Hu, J.-Q. Guo, J. Luo, Phys. Rev. D 71, 127505 (2005); J. Luo et al., Phys. Rev. Lett. 102, 240801 (2009).
- T. J. Quinn et al., Phys. Rev. Lett. 87, 111101 (2001); T. J. Quinn et al., Phys. Rev. Lett. 111, 101102 (2013).
- 16. U. Kleinevoss, "Bestimmung der Newtonschen Gravitationskonstanten *G*," PhD thesis, U. Wuppertal (2002).
- 17. H. V. Parks, J. E. Faller, Phys. Rev. Lett. 105, 110801 (2010).
- 18. S. Schlamminger et al., *Phys. Rev. D* **74**, 082001 (2006). ■

