Measuring cosmic distances
WITH STANDARD SIRENS

Upgrading the Laser Interferometer Gravitational-Wave Observatory. (Courtesy of Caltech/LIGO.)
Daniel Holz is a professor of physics and of astronomy and astrophysics at the University of Chicago. Scott Hughes is a professor of physics at MIT in Cambridge, Massachusetts. Bernard Schutz is a professor of physics and astronomy at Cardiff University in the UK.

Decades of experimental effort paid off spectacularly on 14 September 2015, when the two detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) spotted the gravitational waves generated by a pair of coalescing black holes. To get a sense of the effort leading to that breakthrough, consider that the gravitational waves caused the mirrors at the ends of each interferometer’s 4 km arms to oscillate with an amplitude of about $10^{-18}$ m, roughly a factor of a thousand smaller than the classical proton radius. The detection was also a triumph for theory. The frequency and amplitude evolution of the measured waves precisely matched general relativity’s predictions for the signal produced by a binary black hole merger, even though the system’s gravity was orders of magnitude stronger than that of any system that had been precisely probed before that detection. As figure 1 shows, gravitational-wave astronomy began not with a bang but with a chirp.
Labeled GW150914, that first reported event was soon joined by other detections of binary black hole mergers. Each of those events appeared to be totally dark to traditional astronomical instruments—the matter and electromagnetic fields near the merging black holes were not sufficient to generate any signal other than gravitational. As had long been promised, gravitational waves have opened a window onto an otherwise invisible sector of the universe.

Although celebrated by the physics and astronomy communities and feted by the broader public, gravitational-wave astronomy did not initially overlap significantly with more traditional astronomy. That changed on 17 August 2017, when a gravitational-wave signal, followed by a burst of gamma rays, triggered one of the most intense observing campaigns in the history of astronomy. LIGO, joined now by the Virgo detector in Pisa, Italy, recorded a minute-long chirp (see figure 2) encoding the final several thousand orbits in the coalescence of two neutron stars.² The stars’ collision, at about ¾ the speed of light, was an astronomical cataclysm. Just 1.7 s after the end of the gravitational-wave signal, the orbiting Fermi Gamma-Ray Space Telescope and INTEGRAL observatory recorded a short gamma-ray burst.³

The LIGO–Virgo alert provided the sky position and, importantly, the distance to the event. Just 11 hours later, optical astronomers identified a violent event in the galaxy NGC 4993, a kilonova explosion that shone 1000 times brighter than a typical nova. More than 70 teams made follow-up electromagnetic observations. The effort represents the first time a source has been detected through both its gravitational and electromagnetic radiation. A significant portion of the world’s professional astronomers are coauthors with the gravitational-wave teams on the summary paper describing those observations.⁴ Observations in the x-ray and radio bands continue as we write. From the event, astronomers are learning much about gamma-ray bursts, neutron stars, and their associated physics and astronomy.

Because a gravitational wave encodes the distance to its source, GW170817 provided the astrophysical community with another advance: the first measurement of the local cosmic expansion rate—the Hubble constant—via gravitational waves.⁵ That milestone opened up a completely new way to measure the dynamics of the universe: the standard-siren technique.

A ladder to the stars

The Hubble constant has been the single most important parameter describing cosmology since Edwin Hubble discovered the expansion of the cosmos in 1929. On the largest scales, the universe expands homogeneously and isotropically, so every part of it recedes from every other part. General relativity shows that due to the cosmic expansion, radiation emitted from a distant object is redshifted as it propagates from its source to an observer.

Consider light with a wavelength \( \lambda \) emitted from a source that is a distance \( D \) away. Observers on Earth will measure the light to have a wavelength of \((1 + z) \lambda\), where \( z \) is the light’s redshift. To leading order in \( z \), the source distance and redshift are proportional:

\[
 cz = H_0 D ,
\]

where the Hubble constant \( H_0 \) is today’s value of the Hubble parameter \( H \). It has dimensions of inverse time; the reciprocal \( 1/H_0 \), known as the Hubble time, provides a rough estimate of the age of the universe. Astronomers conventionally express \( H_0 \) in units of \( \text{km s}^{-1} \text{Mpc}^{-1} \), because the megaparsec (1 Mpc = 3.26 million light-years) is convenient for intergalactic distances. As mentioned above, equation 1 is a leading-order expression. For far distant objects, it needs to be corrected with higher-order terms that depend on the nature of the matter and energy that fill the universe.⁶

In principle, just one object of known distance and cosmological redshift suffices to determine the Hubble constant. The redshift of many objects can be determined from spectral mea-
Nova observations not only helped determine the type Ia supernova explosion. Superluminal determinations of the apparent angular shift in the position of emission explains how Earth orbits the Sun. The technique does not work well for larger distances, as the angular shift due to Earth’s orbital motion becomes too small to measure.

For many objects beyond our galaxy, an important tool for measuring distances is the standard candle: an astronomical source whose intrinsic luminosity is assumed to be known. Suppose a star has luminosity $L$, and observers on Earth measure it to have a flux $F$. From the inverse square law and assuming the star radiates isotropically, you obtain the luminosity distance

$$D = \frac{L}{4\pi F}.$$  (2)

Nature does not provide observers with stars whose luminosities are precisely known. However, it does provide stars and other objects whose luminosities can be inferred accurately. Celebrated examples are the Cepheid variables, giant stars whose luminosities vary periodically. By studying a group of such stars in the Small Magellanic Cloud—a dwarf galaxy near the Milky Way—Henrietta Leavitt discovered in 1912 that each star’s oscillation period correlates with its intrinsic luminosity. Some Cepheids are close enough that their distances can be determined using parallax, and thus their luminosity can be calibrated. With the luminosity–period relationship empirically established, Cepheid variable stars can serve as standard candles for measuring distances beyond the limits of parallax.

By putting together multiple methods for measuring distances, astronomers construct what is called the cosmic distance ladder (see the article by Mario Livio and Adam Riess, PHYSICS TODAY, October 2013, page 41). On each rung of the ladder, objects thought to be of standard luminosity are identified and calibrated in terms of measurements contributing to the previous rung.

Various sophisticated methods now exist for measuring $H_0$, but many depend in one way or another on the distance ladder. One method relies on an important standard candle that can be seen very far away: the type Ia supernova explosion. Supernova observations not only helped determine $H_0$, they also implied nonlinear contributions to equation 1 that showed the expansion of the universe is accelerating. That result led to the awarding of the 2011 Nobel Prize in Physics to Saul Perlmutter, Adam Riess, and Brian Schmidt (see PHYSICS TODAY, December 2011, page 14).

The most recent measurement of the expansion using supernovae yields $H_0 = 73.24 \pm 1.74$ km s$^{-1}$ Mpc$^{-1}$. An alternative method, based on the Planck satellite’s measurements of fluctuations in the cosmic microwave background gives $H_0 = 67.74 \pm 0.46$ km s$^{-1}$ Mpc$^{-1}$. The two values are uncomfortably far apart if their uncertainties are to be believed. Given the many rungs on the distance ladder that must be empirically calibrated, it would not be surprising for one or both of the $H_0$ determinations to be affected by undiscovered systematic errors. Or the universe might be more complicated than the community now thinks. Perhaps it is more inhomogeneous; perhaps it is less isotropic; perhaps an important contribution to its mass–energy budget has been overlooked; or perhaps general relativity does not describe the universe well on the largest scales.

The discrepancy among measurements of $H_0$, one of the fundamental quantities of cosmology, may be the result of systematics, or it may hint at new physics. Both possibilities motivate new measurements to resolve or confirm the tension.

**Binary inspiral, a standard siren**

The box on page 38 highlights some important features of gravitational waves. Its equation B3 shows that the amplitude of a...
MEASURING COSMIC DISTANCES

KEY PROPERTIES OF GRAVITATIONAL WAVES

Gravitational waves are described by a tensor field $h_{\mu\nu}$ that characterizes the dynamics of gravity in general relativity. The indices $\mu$ and $\nu$ range over the four coordinates of space and time. Many of the properties of $h_{\mu\nu}$ are analogous to those of the vector potential that characterizes electromagnetic radiation.

For sources moving at much less than the speed of light, electromagnetic radiation is described by the vector potential $A$ that arises from a source’s time-varying electric dipole moment $p$:

$$A_j = \frac{\mu_0}{4\pi} \frac{1}{D} \frac{dp_j}{dt}. \tag{B1}$$

Here, $D$ is the distance from the source, $\mu_0$ is the permeability of free space, and the index $j$ labels one of the three spatial dimensions. The dipole moment, in turn, is the integral of the charge density $\rho$, over the volume of the source, weighted by the position $r$:

$$p = \int \rho \, r \, dV'. \tag{B2}$$

In 1918 Albert Einstein showed that the analogous result for gravitational waves is the quadrupole formula

$$h_{jk} = \frac{2G}{c^4} \frac{1}{D} \frac{d^2 I_{jk}}{dt^2}, \tag{B3}$$

where $G$ is Newton’s constant and $c$ is the speed of light. The indices $j$ and $k$ in equation B3 are purely spatial. Much like electromagnetic radiation’s electric and magnetic fields, gravitational waves are orthogonal to their direction of propagation. As a result, the components of $h_{\mu\nu}$ with time indices—$h_{\theta\phi}$ and $h_{\phi\theta}$—do not radiate and they can be ignored when discussing gravitational waves. The tensor $I_{jk}$ is the source’s mass quadrupole moment

$$I_{jk} = \frac{\int \rho_m \left[ r' r_k' - \frac{1}{3} (r')^2 \delta_{jk} \right] dV'}{D}, \tag{B4}$$

where $\rho_m$ is the source’s mass density and $\delta_{jk}$ is 1 when the indices match and vanishes otherwise. Notice that gravitational radiation involves two time derivatives of the relevant moment rather than the single derivative appropriate for electromagnetic radiation. Notice also that the different constants that connect source and radiation reflect the different fundamental forces involved.

Both electromagnetic and gravitational waves have two polarizations. The electromagnetic basis polarizations point along two orthogonal axes in the plane perpendicular to the direction of propagation, and the electric force that a passing wave exerts on charges can be decomposed into components along those basis directions. Gravitational waves also exert forces normal to the propagation direction, but they act tidally, stretching along one axis as they squeeze along the perpendicular axis. If the wave propagates along the $z$-axis, one polarization stretches and squeezes along the $x$- and $y$-directions. That polarization is conventionally labeled $h_\phi$. The other polarization, $h_\theta$, stretches and squeezes along axes rotated by $45^\circ$ from the $x$- and $y$-axes. The figure above shows how a circular ring of freely floating particles would be distorted over time by a gravitational wave propagating toward the observer with one or the other of those polarization states.

gravitational wave falls off inversely with the distance to the source. If it were somehow possible to learn how the source’s mass quadrupole moment (defined in the box) varies with time, then a measurement of the gravitational-wave amplitude would reveal that distance.

As was first shown by one of us (Schutz), for gravitational waves generated by binary stars it is indeed possible to take the measured data and derive how the quadrupole moment varies. In other words, binary inspiral allows for a determination of the distance to the source without any reference to the cosmic distance ladder. The only empirical calibration needed is of the gravitational-wave detector, to make sure it reports the amplitude of the wave correctly. Beyond that, the only assumption is that general relativity is valid.

Consider a binary in a circular orbit. Its members will circle one another with a frequency $\Omega$ that depends on the binary members’ separation and their masses $m_1$ and $m_2$. The gravitational waves it emits take energy from the orbit and cause the binary’s components to spiral toward one another. As the separation decreases, the orbital frequency increases, which causes increased energy loss due to gravitational waves, which leads to a further decrease in the orbital separation, and so on. The binary thus chirps; the gravitational waves go from low frequency to high at an increasing rate, all the while increasing in amplitude. From Kepler’s law and a well-known formula that relates the power emitted in gravitational waves to the binary’s changing quadrupole moment, it is possible to show that

$$\frac{d\Omega}{dt} = \frac{96}{5} \left( \frac{G M}{c^3} \right)^{5/3} \Omega^{11/3}, \tag{3}$$

plus correction terms that don’t alter the substance of our story. In equation 3, we have introduced the chirp mass $M = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$.

The rate of change of the frequency depends only on one parameter: the chirp mass. Once you know $M$, you know how quickly the frequency is changing at any point in the evolution of the binary system. All binaries with the same $M$ will have, to leading order, the same chirping sweep from low frequency to high, although the corrections to equation 3 do introduce a dependence on the individual masses of the stars. The chirp
time, $Gm/c^2$, characterizes how rapidly $\Omega$ changes due to gravitational-wave emission.

Let’s turn now to the waves’ amplitude and consider a circular binary oriented such that the normal to its orbital plane makes an angle $i$ to our line of sight. With that convention, $i = 0^\circ$ means the binary is viewed head on, and $i = 90^\circ$ corresponds to an edge-on view. As described in the box, gravitational waves come in two polarization states. With the convention that the normal to the orbital plane is in the $xz$-plane, their amplitudes are given by

$$h_+ = \frac{2c}{D} \left( \frac{Gm}{c^3} \right)^{5/3} \Omega^{2/3} (1 + \cos^2 i) \cos 2\Phi(t),$$

$$h_\times = \frac{4c}{D} \left( \frac{Gm}{c^3} \right)^{5/3} \Omega^{2/3} \cos i \sin 2\Phi(t),$$

where $\Phi(t)$ is the accumulated orbital phase found by integrating the orbital frequency $\Omega$ over the duration $t$ of the measurement, and the factor of 2 multiplying $\Phi(t)$ is due to the waves’ quadrupolar nature.\(^{10}\) The amplitudes depend on the masses $m_1$ and $m_2$ only through the chirp mass.

Once observers have measured gravitational waves from a binary, they can accurately match the waves’ phase evolution to that of a model template, as in figure 1a. Doing so determines $M$, typically with high precision. If it is possible to measure more than one polarization, then the ratio of their amplitudes determines the inclination angle $i$. Once $M$ and $i$ are known, the distance to the source is determined, according to equation 4, by measuring the waves’ amplitude.

Binary inspiral thus acts as the gravitational-wave analogue of a standard candle, but it does not require the cosmic distance ladder. No empirical calibrations are needed to obtain the source distance; the only fundamental assumption is that general relativity is valid. Because gravitational-wave detection is more like hearing a signal than seeing an image, several of us in the field independently came up with the name standard siren. The term first appeared in print in a paper by two of us (Holz and Hughes), and it seems to have stuck.\(^11\)

The “standard” in standard siren arises because the waves’ frequency sweep and amplitude both depend on the same mass $M$. That is no coincidence. It follows from the fact that the intrinsic gravitational luminosity—a combination of amplitude and frequency—depends only on the number of orbits remaining until coalescence.\(^12\) That number, in turn, depends only on $\Omega$ and $d\Omega/dt$. Determining those two quantities from the gravitational waveform thus yields the binary’s instantaneous luminosity. Measuring the amplitude directly then gives the luminosity distance $D$.

### A redshift degeneracy

The chirp mass enters the equations for the gravitational waveform and frequency evolution in a combination with units of time: the chirp time $Gm/c^2$. That time scale experiences the usual cosmological redshift, so there is a fundamental mass–redshift degeneracy: A binary with masses $m_1$ and $m_2$ at redshift zero produces gravitational waves that fit exactly the same waveform template as the waves from a binary with masses $m_1/(1+z)$ and $m_2/(1+z)$ at redshift $z$.

The degeneracy can be broken in several ways. Perhaps the simplest is to assume values for $H_0$ and other cosmological parameters. Then, from equation 1, a measurement of distance, as determined from the amplitude of a gravitational wave, yields an estimate of the redshift. The true masses of the binary can then be inferred. That approach has been applied for most of the sources that LIGO and Virgo have measured to date, including GW150914.

If, however, the gravitational-wave event has an electromagnetic counterpart, a measurement of the spectrum of the counterpart or of an associated host galaxy determines the redshift to the source without the need of additional assumptions. In that case, telescopes and gravitational-wave detectors do the tasks to which they are best suited, measuring redshift and distance, respectively. Given those quantities, equation 1 yields the Hubble constant $H_0$.

### Gravitational-wave event GW170817

The field of standard-siren cosmology was conceived\(^6\) in 1986, and after more than 30 years of gestation, it finally arrived with triplets: the detection of GW170817 by the LIGO and Virgo observatories, followed 1.7 seconds later by the discovery of an associated gamma-ray burst, followed 11 hours later by the discovery of an optical counterpart.\(^13\) GW170817 comprised two compact objects with masses in the range of 1.36–2.26 solar masses and 0.86–1.36 solar masses, consistent with a binary neutron star system. The coalescence of the stars occurred at a

---

**FIGURE 3. A STANDARD-SIREN DETERMINATION** of the Hubble constant $H_0$. The blue curve shows the probability distribution for the value of $H_0$, as determined from measurements of the event GW170817. Vertical dashed lines mark the 68% credible interval; vertical dotted lines mark the 95% interval. Also shown are measured values and error intervals for supernova observations (brown) and for the Planck satellite’s determination (green) based on observations of the cosmic microwave background. The standard-siren value is consistent with the supernova and CMB measurements; standard-siren measurements have the potential to resolve the tension between them. (Adapted from ref. 5.)
luminosity distance of about 40 Mpc. With an observed signal-to-noise ratio of 32, GW170817 is by far the closest and loudest gravitational-wave source detected to date.

The optical counterpart to GW170817 (see figure 2b) was found within 10 arcseconds of the center of the galaxy NGC 4993, an angle that corresponds to a separation of about 2 kpc. The redshift of the galaxy is 0.009. A Bayesian analysis\(^5\) that combines the galactic redshift with the LIGO–Virgo measurement of distance to GW170817 leads to an estimate of the value of the Hubble constant: \( H_0 = 70.1^{+1.2}_{-1.0} \text{ km s}^{-1} \text{ Mpc}^{-1} \). Figure 3 shows the full Bayesian probability distribution for \( H_0 \).

In principle, the combined LIGO and Virgo distance estimate is limited by the detectors’ amplitude calibration. The current precision of a few percent will improve as the detectors approach their design sensitivity. However, as equation 4 shows, the measurement of \( D \) and hence of \( H_0 \) depends on determining the inclination \( \iota \). Uncertainty in \( \iota \) increases uncertainty in \( D \) and thus increases uncertainty in \( H_0 \).

The determination of \( \iota \) comes primarily by measuring gravitational-wave polarizations. The two LIGO detectors are closely aligned in orientation, so they do a poor job constraining polarization. Virgo is at a different orientation, so it can provide additional constraints on polarization. Unfortunately, Virgo had little sensitivity to the sky position of GW170817. Polarization was thus not well constrained even by the combined LIGO and Virgo measurements. The result was a large uncertainty in \( \iota \) and a roughly 15% error in the measured distance to GW170817. If additional information can constrain the inclination—for example, from studies of the associated gamma-ray burst, kilonova, or afterglow—then the distance estimate and derived value of the Hubble constant will correspondingly improve.

### A loud, bright future

The event GW170817 is the first in what we expect to be a rich catalog of standard sirens. LIGO and Virgo are currently being upgraded (see the photo on page 34) and should resume observing in early 2019. At design sensitivity, they will be able to detect neutron star coalescences at a distance large enough to boost the event rate by a factor of 5 to 10. It will be more difficult to find electromagnetic counterparts at larger distances, though, because of the \( 1/D^2 \) falloff in electromagnetic flux and the \( D^2 \) growth in the number of possible host galaxies in a fixed localization angle.

Accuracy in the determination of cosmological parameters should improve as roughly \( 1/N \) as \( N \) sirens are measured. With as few as 30 events, standard sirens may be able to resolve the tension between current competing measurements of the Hubble constant.\(^14\) With more events, it may become possible to do standard-siren science without electromagnetic counterparts by averaging over all potential host galaxies, by exploiting knowledge of the mass distribution of merging neutron stars, or by using the effect that the neutron star equation of state has on the binary waveform at late times.\(^15\)

Other detectors will soon join LIGO and Virgo. KAGRA (Kamioka Gravitational Wave Detector),\(^16\) a kilometer-scale interferometer being built in Japan’s Kamioka mine, is expected to begin operating in 2020, and LIGO-India, planned for the Hingoli district of the Maharashtra state, should begin operating within the next decade. Additional detectors improve the signal-to-noise ratio and make it possible to better determine source parameters. The community has already benefited from detector synergy with the measurement of GW170817: The weak signal measured by Virgo indicated that the event was near a null in the Virgo antenna sensitivity pattern. That information significantly aided in locating the event direction and greatly facilitated the search for the optical counterpart.

Planning for the next generation of gravitational-wave detectors is under way.\(^17\) Ground-based detectors, such as the proposed Cosmic Explorer and the Einstein Telescope, will operate in the largely the same 10–10 000 Hz frequency band as LIGO and Virgo but will be able to measure binary inspiral to much larger distances—essentially to all neutron star mergers in the universe. (See PHYSICS TODAY, October 2018, page 25.) The space-based Laser Interferometer Space Antenna (LISA), a European Space Agency mission that includes NASA participation, is set to launch by 2034. LISA will operate in the millihertz frequency band, have high sensitivity, be self-calibrating, and measure standard-siren events involving the coalescence of black holes with masses from about \( 10^4 \) up to about \( 10^7 \) solar masses for distances corresponding to redshifts as large as 20. Especially on the low end of its mass range, LISA may be able to determine the source position accurately enough to pin down the galaxy cluster or even the galaxy hosting the event.

Just over one year ago, GW170817 not only allowed a measurement of the distance to its source but also provided a proof of principle that the standard-siren technique could measure the Hubble constant. Future measurements will transform standard sirens into an important tool for studying the expansion history of the universe.

### References

17. B. Sathyaprakash et al., Class. Quantum Grav. 29, 124013 (2012); B. P. Abbott et al., Class. Quantum Grav. 34, 044001 (2017).