Detection of Gravitational Radiation

There is justified excitement about the imminent ($\leq$ 5 years, we hope!) detection of gravitational radiation by ground-based detectors, and possibly from pulsar timing arrays. However, direct detection has not happened yet. The reason is that gravitational waves are exceedingly weak and thus extremely precise instrumentation and excellent statistical analysis will both be needed to finally welcome this new astrophysical messenger to the table. In this lecture we will discuss several aspects of the detection of gravitational radiation, and then finish by indicating some of what we might learn from those detections.

The theorist's version: ideal detection and how it is possible

To work our way into how waves can theoretically be detected, let's suppose that we are a point particle. If we are shut up in a very small box in empty space and a gravitational wave passes by, can we tell by the accelerations it induces if we make our measurements over a small time? No! This is an aspect of the equivalence principle: over short distances and small times, an observer in free fall (which this still is) won't measure any accelerations. Thus we need to make observations over nonzero distances to detect gravitational waves.

Given that understanding, the basics of detection are initially obvious, then inobvious, then (we hope!) obvious again. The initially obvious part is that because a passing gravitational wave stretches and squeezes spacetime, it seems to be a simple matter of measuring the distance between two separated locations as a function of time. The inobvious part starts when we consider *how* we are to measure the distance. The problem is that the measuring sticks you might initially consider are *also* stretched and squeezed by the same amount. For example, suppose that you have 4,000 one-meter sticks lined up next to one of the arms on either of the LIGO interferometers. If a gravitational wave comes by and compresses the 4 km distance by some factor, that's also the factor by which all of the sticks will be compressed. Thus the length of the interferometer, measured that way, doesn't change. Similarly, if you have determined that at rest the interferometer has a length that is a certain number of wavelengths of light of a specified frequency, the problem is that when a gravitational wave comes along those wavelengths are also compressed in the same way. Therefore, again, the length doesn't appear to change.

The solution to this conundrum lies in an understanding of what is *not* affected by gravity. For example, the speed of light is not affected by gravity, nor are fundamental constants such as the charge of the electron, the rest mass of the electron, or Planck's constant.

To see how the constancy of the speed of light helps, let's think about the measurement of distance that takes place in a one-armed interferometer. Working off of Bernard Schutz's lecture notes (several versions are on the Web), we note that for axes oriented such that the wave propagates in the z direction, and for the "+" mode of polarization, the metric becomes

$$ds^2 = -dt^2 + (1 + h_+)dx^2 + (1 - h_+)dy^2 + dz^2 \tag{1}$$

where we are using the standard shorthand in which $c = 1$. The perturbation h_+ is time-variable: $h_+ = A^{xx}e^{ik(z-t)}$ (of course we take the real part). Suppose that the origin is at $x = 0$ and the other station is at a coordinate location $x = L$; note that for the reasons argued above, the other station is *always* at that coordinate location. Then the *proper* distance between the stations is

$$\Delta s = \int_0^L [1 + h_+(x)]^{1/2} dx \tag{2}$$

and hence the proper time of propagation to the far station is

$$t_{\rm far} = \int_0^L [1 + h_+(t(x))]^{1/2} dx \ . \tag{3}$$

This is an implicit equation, but because $h_+ \ll 1$ we can expand:

$$t_{\rm far} \approx L + \frac{1}{2}\int_0^L h_+(x)dx \ . \tag{4}$$

If we now include the return time, and generalize to a starting time of t (instead of the 0 we have thus far assumed) we find

$$t_{\rm return} = t + 2L + \frac{1}{2}\left[\int_0^L h_+(t + x)dx + \int_0^L h_+(t + L + x)dx\right] \tag{5}$$

and (for this particular geometry, not for a generally oriented wave)

$$\frac{dt_{\rm return}}{dt} = 1 + \frac{1}{2}\left[h_+(t + 2L) - h_+(t)\right] \ . \tag{6}$$

For more general orientations, the rate of change of the return time with the time also depends on the wave amplitude at the far end.

In words: because the speed of light is constant, the changing proper distance *does* have an effect on the return time. Thus, in principle, if it were possible to tag each cycle of laser light (via modulation, say) and measure its transit time with sufficient precision, gravitational waves could be detected with a single arm down which light traveled and reflected. Indeed, in some sense this is akin to how bar detectors work in theory. For bar detectors, the passage of a gravitational wave squeezes and stretches the matter in the bar because the proper distance between the ends of the bar changes. If the frequency of the wave matches the resonant frequency of the bar, then oscillations can be set up that might be high enough amplitude to make their presence known to piezoelectric detectors on the bar. The problem is that it is difficult to get the required sensitivities with bars, and because they rely on resonance their best sensitivity is confined to a small range of frequencies.

This is why most of the big groups have focused on laser interferometers. The time measurement suggested in the previous paragraph is problematic because it would require utterly spectacular timing. Suppose, for example, that you want to detect a 100 Hz gravitational wave, and you have managed by clever arrangements of mirrors to effectively increase the length of your detector to the full 3×10^8 cm size of the wave. A wave with an amplitude of $h = 10^{-23}$ would require a measurement of a change in arrival time that is 10^{-23} times the travel time, or 10^{-25} seconds! That is not happening. Now, timing changes *are* the basis of pulsar timing arrays, where multiple millisecond pulsars are timed and residuals are examined for hints of gravitational waves. But we'll focus on ground-based laser interferometers here.

The approach taken by multiple groups around the world (LIGO, Virgo, KAGRA, GEO-600) has been to use interferometers. There are various possible designs, but the currently popular approach is to have an L shaped configuration in which the laser source is at the corner and there are mirrors at both ends. There are details that matter, but the toy version is that light is sent to both simultaneously and if the detector is *not* affected by a gravitational wave then the returned light interferes with itself to give one fringe pattern, whereas if the detector *is* influenced by a wave then because the proper distances to the two ends will differ from their configuration at rest, the fringe pattern will differ from the at-rest pattern in a time-dependent way that can signal the presence of gravitational waves.

Even here you might think that there are insuperable problems. Suppose that we have 4×10^5 cm long arms (as in LIGO) and an $h = 10^{-23}$ gravitational wave. The path differences are then at most 4×10^{-18} cm!!! Even if you increase the path length with mirrors as before to 3×10^8 cm, the path difference is just 3×10^{-15} cm, which is considerably less than the size of a proton, let alone the $\sim 1\mu$m wavelength of the light.

This might make the whole enterprise seem impossible, but the difficulty can be circumvented by the use of the statistics of many photons. The point is that a fringe actually consists of a very large number of photons, so by averaging over those photons you can find the centroid of the fringe to a precision that is vastly better than the precision with which you can localize a single photon. An astronomical analogy has to do with determining the centroid of a star. The Hipparcos telescope, which had an astrometric precision of 0.001", had a 29 cm diameter mirror. Assuming it used visible light with a wavelength of 5×10^{-5} cm, its formal diffraction limit was $1.22\lambda/D = 0.4$". So how did it do 400 times better for its astrometric precision?

The answer is that Hipparcos only looked at bright stars. The light from such point objects spreads out in a known pattern called a point spread function. This extends over many pixels, so by measuring the intensity at each of those many pixels and fitting the distribution to the point spread function, it is possible to determine the centroid of the light, and thus the direction to the star, to something like $N^{-1/2}$ times the size of a pixel if N photons are observed. Similarly, laser interferometers involve so many photons that they can, in fact, resolve fringe movements that are at the small size expected from gravitational waves.

That, at least, is the theory. How does this work in practice?

Sources of noise

For more details on interferometer noise, I strongly recommend that you read "Gravitational Wave Detection by Interferometry (Ground and Space)" by Pitkin, Reid, Rowan, and Hough (2011, Living Reviews in Relativity, **14**, 5).

Given the weakness of gravitational wave signals, it should come as no surprise that there are many sources of noise that must be overcome. Some of the major ones are:

- Seismic noise. Every quake, every treefall, every truck, every mischievous grad student jumping up and down near the detector (yes, that was me) will shake the detector and thus introduce noise. Amazingly, it is possible to shield against this fairly well by the use of pendulums. The point is that if you shake a pendulum above its resonance frequency, the motion of the suspended weight is damped like $1/(\text{frequency})^2$ compared to the motion at the suspension point (try it). Thus by combining multiple-pendulum suspension with active damping systems, it seems likely that seismic noise itself will not be a problem. However...

- Gravity gradient noise. *This* is likely to be the limiting noise source at low frequencies. If you think about what seismic activity actually is, you'll realize that it is pressure (and thus density) waves in the ground. Because the density varies, the gravitational field does as well. We *can't* shield against gravity, so this is a problem. To understand this even further, imagine that your detector is on the surface and a density wave passes underneath. When the dense part of the wave is to the left of a mirror, that mirror is drawn to the left; when the dense part is to the right, the mirror is drawn to the right. Thus the mirrors are forced to swing back and forth by the varying gravity caused by such waves. Luckily the spectrum of seismic noise drops off rapidly towards higher frequencies, so in the >10-15 Hz range that current ground-based detectors will explore, this won't be a dominant factor. The noise sufficiently far underground is less (part of that has to do with getting away from surface waves), so third-generation detectors such as the proposed Einstein Telescope will be able to push down to a few Hertz.

- Thermal noise. If the components of the detector vibrate thermally, this will obviously introduce noise into the system. It turns out that for Advanced LIGO and other second-generation detectors, the dominant source of this noise will actually be dissipation in the mirror coatings (the coatings are needed to achieve the required high reflectivity). There is a good deal of research into better coatings, but materials science is complex and thus there is not a clear path to the best option. This has also spurred the search for very low loss materials for the test masses and their suspensions, with fused silica and sapphire being the best current contenders. It appears that thermal noise will likely be the limiting factor at the most sensitive frequencies ($\sim 100-200$ Hz) of second-generation detectors, so improvements would pay significant dividends.

- Quantum noise. For this, we can start with shot noise, which is what you get when you have a finite number of photons. You'd like to determine the phase of the laser signal to unlimited precision, but the Poisson arrival times of photons means that this isn't possible. Those phase errors are what will limit the sensitivity at high frequencies (several hundred Hertz and above), where we expect to get most of the astrophysical information about the individual masses of compact binaries and even the radii of neutron stars (more on that later). To make it even worse (and this makes it "quantum" noise), there is an uncertainty-like relation between the phase uncertainty and the amplitude uncertainty of a wave. Thus although there are tricks you can play with nonlinear optics (where "nonlinear" can include photon number changes, or "two men enter, one man leaves!", to quote Mad Max Beyond Thunderdome) to reduce either the phase noise or the amplitude noise, their product cannot be decreased and in practice will increase if you mess with the photons. Nonetheless, because phase noise is important at different frequencies (above ~ 200 Hz, say, for Advanced LIGO) than is amplitude noise (below ~ 200 Hz), there are some potentially clever tricks you can play. For example, Cannon et al. 2012, ApJ, 748, 136 (see also Kyutoku and Seto, arXiv:1312.2953) suggest that it might eventually be possible to know that a binary coalescence is happening up to ten seconds *before* the actual merger. If so, various people including Rana Adhikari have proposed that it might be possible to wait until the signal passes 200 Hz, then squeeze the phase noise to improve high-frequency sensitivity.

The statistics of detection: detection versus parameter estimation

Not all astronomical electromagnetic sources are weak. This means that you can detect some sources without any foreknowledge of their properties: you don't need to know anything about nuclear fusion to detect the Sun! In contrast, all gravitational wave sources are so weak that we have not yet detected any of them directly. This means that in addition to the need for exceptionally precise instruments, the gravitational wave community needs to take advantage of the best available statistical methods.

One way that we can improve our detection prospects is to apply a *template* to the data. As an analog to this, suppose that you are at a really loud party and you are trying to talk with someone. It turns out (yes, I've done this) that even if you're not a lip reader, if you watch your friend's face and see their lips move you can pick out more of what they are saying than if you don't have those visual clues. That's because you have *some* idea of what they are saying, and this gives you a hint that can allow you to pick out their speech from the background noise.

Three of the four categories of gravitational wave sources that we discussed last time are amenable to this type of matched filtering. The exception is the burst category, which more or less by definition are sources with waveforms that we can't predict; for those, you need more general and less sensitive techniques such as searches for excess noise. For binaries, the templates are a combination of analytical functions obtained by careful expansions in small quantities such as the orbital speed over the speed of light (most useful for comparable-mass binaries) or the mass ratio (most useful for extreme mass ratio binaries). For continuous sources, you can guess at the form of the signal if you see an electromagnetic counterpart (such as a pulsar), or if you don't have a counterpart you can still look for a pulse rate that you can expand into a constant part and a first derivative (for example). For stochastic backgrounds, the natural approach is to look for a broad stretch of frequencies over which the signal behaves like a power law.

To get a handle on how you use templates, suppose that you have some data $x(t)$ and you have a best characterization $n(t)$ of the noise. You want to apply a certain template $h(t)$ to the signal to determine whether the signal is present. Your first thought might be that you could define a normalized version of your template and then compute the product between your template and the data

$$\int_{-\infty}^{\infty} h(t)x(t)dt \; . \tag{7}$$

Equivalently, you might imagine using a Fourier transform to do this in the frequency domain:

that is, you could define the Fourier transform of the data or the template as

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t)e^{2\pi i f t}dt \tag{8}$$

and the measure of the overlap of your template with the data would be

$$\mathcal{R}\left(\int_{0}^{\infty} \tilde{x}(f)\tilde{h}^{*}(f)df\right) \ . \tag{9}$$

Here the asterisk indicates a complex conjugate, and we ultimately take the real part.

The difficulty with this definition is that it does not take into account the properties of the noise. Suppose, to be extreme, that the noise is overwhelming except in a small range of frequencies around f_0. Surely you don't want to weight the data equally in all frequencies, as in the above formulation; you want instead to place much greater weight around f_0, because that is the frequency range that contains your useful data. These considerations lead one to adopt instead the *Wiener optimal filter*. To define this filter, we define first the one-sided spectral density of the noise

$$S_n(f) = 2\int_{-\infty}^{\infty} n(t)e^{2\pi i f t}dt, \qquad f > 0 \tag{10}$$

and then the scalar product between any two functions x and y as

$$(x|y) \equiv 4\mathcal{R}\int_{0}^{\infty} \frac{\tilde{x}(f)\tilde{y}^{*}(f)}{S_n(f)}df \ . \tag{11}$$

The appearance of $S_n(f)$ in the denominator gives harmonic weighting to the noise, which is what we want. If the noise is Gaussian and stationary, then the log of the likelihood function (which we want to use in standard Bayesian statistical analyses) is

$$\ln\Lambda(x) = (x|h) - \frac{1}{2}(h|h) \ . \tag{12}$$

Here we have relaxed the normalization requirement on the template h. The optimal signal to noise ratio, which determines the probability of detection, is $\rho = \sqrt{(h|h)}$.

Ideally we'd have a full set of templates that we would match to the data, and this would therefore give us optimal detection and characterization of the source. But it's not quite that easy. The problem is that (if we consider binaries) there are a lot of parameters to consider in our template models. A fully general double black hole binary has two masses, six parameters corresponding to the two spins (magnitude and direction), two parameters associated with the direction of the orbital axis, an eccentricity, a distance from us, and a direction on the sky. Searching over all those is not in the cards. There are some simplifications (e.g., the "extrinsic" parameters such as the direction on the sky can be factored out from the "intrinsic" parameters of the binary), but it is still necessary to cut some corners if you want to do analysis on a timescale short enough for electromagnetic followup or even the pre-detection we discussed in the previous section.

Thus it has been understood clearly in the community that fast *detection* of a source need not use identical templates to eventually precise *characterization* of the source. This has led to assessments of how *effectual* and *faithful* a given template bank is. A template bank is effectual [a terrible term, in my opinion; it should just be "effective", which would avoid monstrosities such as "effectualness"] if, for any real signal, there are templates in the bank with sufficient overlap for detection in a given fraction of signals (90% is often a goal). The bank is *faithful* (a much better term!) if the parameter values for the best-matched template are in agreement with the actual parameter values for the system.

To understand this with a simple example, consider a two-dimensional vector, and suppose that you want to represent this as a combination of basis vectors. You know that any such vector can be represented as a linear combination of any two basis vectors that are neither parallel nor antiparallel, so any such pair of vectors will be effectual. However, if you had some particular orientation of axes in mind, the coefficients for your pair of vectors would not be faithful to the "real" coefficients.

An example more directly related to gravitational waves has to do with the role of black hole spin. Suppose you have a coalescing binary of two black holes. Say that their spins are exactly parallel to each other and to the orbit; this guarantees that the binary will not precess. Recent analyses have found that the waveform can be well-modeled using templates of *nonspinning* black holes; the downside is that the masses (for example) that you infer with such templates will not be correct. Thus you can detect such binaries with a restricted set of templates, but your numbers will be wrong. Depending on your goals, this might not be a problem; for example, if your estimate of the *direction* to a source is not compromised by your simplistic template choice, then you might be able to (1) detect the source, (2) follow it up electromagnetically, and (3) at your leisure, later, use a full set of templates to figure out the real binary parameters and their uncertainties.

A more extensive program along these lines (led by Manuel Tiglio) involves reduced basis waveforms. To understand the principle, let us again turn to vectors. Suppose that you have an n-dimensional space and have computed $m \gg n$ vectors in that space (suppose also that you don't know anything about vectors!). You'd like to have a set of basis vectors from which you can build any of the m vectors in your template bank. A way you could do this is (1) pick one of the m vectors at random, and add it to your bank after normalizing it,

(2) go through the remaining $m-1$ vectors to determine which one has the *least* overlap with your original vector, (3) normalize that one and add it to your bank, (4) now go through the remaining $m-2$ vectors to determine which one has the least overlap with the best linear combination of your two vectors, (5) normalize *that* vector and add it to your band, and (6) repeat until every remaining vector can be represented by a linear combination of the vectors in your bank to some pre-set tolerance.

We happen to know that n vectors will suffice, but in the more complicated real case of gravitational waveforms we don't know this in advance. However, by using the same procedure as above (lots of waveforms computed, pick one, pick the next one with the smallest overlap, and so on), tremendous reductions are possible in the number of templates required in a bank so that detections are nearly optimal. Of course, unfolding the best overlap to get the actual parameters is a much more computationally intensive procedure, but given that we might have tens of detections in a given year, any detected source will be beaten to within an inch of its life to confess its properties!

As a final thought, what will we be able to learn from gravitational wave detections, assuming that our characterization and parameter estimation is ultimately correct? In addition to the obvious huge physics returns (GR works in strong gravity, horizons exist, gravitational waves travel at the speed of light), there will be many things we can learn astrophysically. The rate and properties (spins, masses, mass ratios) of stellar-mass binaries will give us insight into stellar evolution that we can't get any other way. Pulsar timing detections won't tell us the spins of supermassive black holes (the holes will be too far from each other), but if individual sources are seen they will give us key information about galaxy mergers and black hole coalescence. If high-frequency sensitivities are improved by squeezing light, we might well get information about neutron star radii (critical for nuclear physics) that is at least complementary to, and possibly better than, what we can find from X-ray observations. If we also get electromagnetic counterparts, we could help solve mysteries about short gamma-ray bursts for stellar-mass binaries, and for supermassive black hole binaries (such as with pulsar timing arrays or space-based detectors such as LISA) we'll have a tremendous cosmological probe because we will have both a redshift and a luminosity distance. If we're really lucky and there is a supernova in our Galaxy that kindly waits for gravitational wave detectors to reach full sensitivity (2020, give or take a couple of years), then we will be able to combine electromagnetic, neutrino, and GW information to revolutionize our understanding of core-collapse supernovae. It's trite, but gravitational wave detection really will open up a new window to the universe; let's wait to be surprised!