Special Lecture: What Have We Learned From Gravitational Waves?

Initial question: what have we learned from the direct detection of gravitational waves?

Since the first direct detection of gravitational waves (or, interchangeably, gravitational radiation), on September 14, 2015, close to 100 events have been detected and many more are on the way due to the continued improvement of gravitational wave detectors. Gravitational waves give us a unique view of the universe: the events we have seen have, with one exception, been electromagnetically dark, which means that GW are the *only* way we could see those events. Given that the energy per time emitted at the end of a comparable-mass merger is vastly more luminous than anything else (by a factor of at least 10,000), that means that we are seeing the most luminous events in the universe for the first time.

Let's summarize a few of the highlights before we go into details:

1. Detection of a heavier class of stellar-mass black holes than was previously known. The masses of previously known stellar-mass black holes topped out at $\sim 20\ M_\odot$, but many above $30\ M_\odot$ have been seen in binary coalescences.
2. In particular, one event (GW190521) and a few other events may have had masses in the "pair instability gap", which was expected to not have black holes if they were born from stars.
3. The first double neutron star merger, GW170817, demonstrated that short gamma-ray bursts can come from the coalescence of two neutron stars, as suggested by numerous people, and validated the predictions of a "kilonova", which is the emission produced by the debris of the collision.
4. Some people are very excited about the prospect for numerous future events to provide an estimate of the Hubble parameter that is independent of electromagnetic estimates.

Now let's think about the basics of gravitational radiation, followed by some thought about how we can get binary sources (which are the only sources thus far detected). We'll start with a review of radiation in general.

Radiation in general, and gravitational radiation in particular

By definition, a radiation field must be able to carry energy to infinity. If the amplitude of the field a distance r from the source in the direction (θ, ϕ) is $A(r, \theta, \phi)$, the flux through a spherical surface at r is $F(r, \theta, \phi) \propto A^2(r, \theta, \phi)$. If for simplicity we assume that the radiation is spherically symmetric, $A(r, \theta, \phi) = A(r)$, this means that the luminosity at a distance r

is $L(r) \propto A^2(r)4\pi r^2$. Note, though, that when one expands the static field of a source in moments, the slowest-decreasing moment (the monopole) decreases like $A(r) \propto 1/r^2$, which implies that $L(r) \propto 1/r^2$ and hence no energy is carried to infinity. This tells us two things, regardless of the nature of the radiation (e.g., electromagnetic or gravitational). First, radiation requires time variation of the source. Second, the amplitude must scale as $1/r$ far from the source.

We can now explore what types of variation will produce radiation. We'll start with electromagnetic radiation, and expand in moments. Suppose that we are far from some distribution of electric charges, which could be in motion. For a charge density $\rho_e(\mathbf{r})$, the monopole moment is $\int \rho_e(\mathbf{r})d^3r$. We assume that the volume over which we perform the integral encompasses the entire system; no charges can enter or leave. As a result, the monopole moment is simply the total charge Q, which cannot vary, so there is no electromagnetic monopolar radiation. The next static moment is the dipole moment, $\int \rho_e(\mathbf{r})\mathbf{r}d^3r$. There is no applicable conservation law, so electric dipole radiation is possible. One can also look at the variation of currents. The lowest order such variation (the "magnetic dipole") is $\int \rho_e(\mathbf{r})\mathbf{r} \times \mathbf{v}(\mathbf{r})d^3r$. Once again this can vary, so magnetic dipole radiation is possible. The lower order moments will typically dominate the field unless their variation is reduced or eliminated by some special symmetry.

Now consider gravitational radiation. Let the mass-energy density be $\rho(\mathbf{r})$. The monopole moment is $\int \rho(\mathbf{r})d^3r$, which is simply the total mass-energy. This is constant, so there cannot be monopolar gravitational radiation. The static dipole moment is $\int \rho(\mathbf{r})\mathbf{r}d^3r$. This, however, is just the center of mass-energy of the system. In the center of mass frame, therefore, this moment does not change, so there cannot be the equivalent of electric dipolar radiation in this frame (or any other, since the existence of radiation is frame-independent). The counterpart to the magnetic dipolar moment is $\int \rho(\mathbf{r})\mathbf{r} \times \mathbf{v}(\mathbf{r})d^3r$. This, however, is simply the total angular momentum of the system, so its conservation means that there is no magnetic dipolar gravitational radiation either. The next static moment is quadrupolar: $I_{ij} = \int \rho(\mathbf{r})r_i r_j d^3r$. This does not have to be conserved, and thus there can be quadrupolar gravitational radiation.

This allows us to draw general conclusions about the type of motion that can generate gravitational radiation. A spherically symmetric variation is only monopolar, so it does not produce radiation. No matter how violent an explosion (even a supernova!) or a collapse (even into a black hole!), no gravitational radiation is emitted if spherical symmetry is maintained. In addition, a rotation that preserves axisymmetry (without contraction or expansion) does not generate gravitational radiation because the quadrupolar and higher moments are unaltered. Therefore, for example, a neutron star can rotate arbitrarily rapidly without emitting gravitational radiation as long as it maintains stationarity and axisymmetry and rotates around the axis of symmetry.

This immediately allows us to focus on the most promising types of sources for gravitational wave emission. The general categories are: binaries, continuous wave sources (e.g., rotating stars with nonaxisymmetric lumps), bursts (e.g., asymmetric collapses), and stochastic sources (i.e., individually unresolved sources with random phases; the most interesting of these would be a background of gravitational waves from the early universe).

Frequency limits for different sources

Remarkably, it turns out that the maximum gravitational wave frequency from a given source essentially depends only on its average density! We can motivate that for binaries by remembering Kepler's Third Law: $f \sim (M/a^3)^{1/2}$, where M is the total mass of the binary and a is the semimajor axis. Clearly a can't be any smaller than the sizes of the objects, so if that size is R then the maximum frequency is $f_{\max} \sim (M/R^3)^{1/2}$, and since M/R^3 is proportional to the density, $f_{\max} \sim \rho^{1/2}$. What about rotation? If something is bound by gravity (as all big things in the universe are), then the acceleration holding the thing together is $\sim GM/R^2$. The centrifugal acceleration is v^2/R, and for an angular velocity f, $v = Rf$, which means that the centrifugal acceleration is Rf^2. Something can obviously rotate slowly, but at the maximum rate the gravitational and centrifugal accelerations balance; faster than that, and the thing would fly apart. This again gives $f_{\max} \sim (M/R^3)^{1/2} \sim \rho^{1/2}$. Finally, for pulsations or bursts, it isn't obvious but you also get $f_{\max} \sim \rho^{1/2}$ (with some caveats; here we focus on pulsations or bursts that involve most of the mass of the star). If we think about black holes in particular, then because the event horizon radius $r_g \sim M$, the "density" is like $\rho \sim M/r_g^3 \sim M/M^3 \sim M^{-2}$, and the maximum accessible frequency is therefore $f_{\max} \sim M^{-1}$.

As a result, different frequency ranges probe different categories of sources. At the LIGO/Virgo/KAGRA frequencies of $\sim 20 - 2000$ Hz, only neutron stars and stellar-mass black holes (and associated events such as core-collapse supernovae) can be seen. At lower frequencies, such as those of LISA ($\sim 10^{-4} - 10^{-1}$ Hz), those binaries can still be seen, but at wider separations. In addition, we can see double white dwarf binaries, some supermassive black hole binaries ($\sim 10^4 - 10^7\ M_\odot$), and so-called extreme mass ratio inspirals (EMRIs), where a stellar-mass object spirals into a supermassive black hole. At the lower frequencies of pulsar timing arrays ($\sim 10^{-9} - 10^{-6}$ Hz), even more massive black holes can be seen (up to $10^{10}\ M_\odot$), and even those will not merge in band.

Ways to produce binary sources of gravitational radiation

So far, only binaries have had their gravitational radiation detected directly. We therefore now talk about how you might get binary compact objects to spiral together; we'll focus on the stellar-mass objects that can be seen with ground-based detectors (LIGO/Virgo/KAGRA) and say just a bit at the end about massive black hole binaries. There are a few basic mechanisms which have been discussed, of course with details and subvariations:

1. Evolution of a massive binary. This is the simple and obvious one. Massive stars, of the type that would leave behind a neutron star or black hole, occur with high frequency in binaries ($> 80 - 90\%$). So, all that is necessary is to have both evolve into black holes (say) and then come together. But the reason it's not that simple is that if you go with what might be your initial impulse and require that their separation be larger than their size as giants (i.e., more than 1 au or a few au), then the inspiral time is many times the age of the universe for black hole masses of tens of solar masses. So you need tricky ways to bring them together without having them merge before both are black holes.

2. Dynamical processes, particularly binary-single interactions. Say that a black hole is produced in a binary. Stellar densities are too low pretty much anywhere for direct collisions to happen, but if the binary interacts with a single star (in the sense that the single star more or less passes through the binary), then complicated interactions happen. It is normally the case that when the dust settles, (1) if the binary was "hard" (meaning that the binding energy of the binary is greater than the relative kinetic energy of the single) then after the interaction the binary is harder (meaning more bound), and (2) the final two objects in the binary are the two most massive of the three stars. Thus black holes can start single, or with a normal-star companion, and exchange into binaries. Lots of work has been done on this, too. This seems a promising mechanism, but only a small fraction of stars are in environments dense enough for this process to be efficient; $\sim 10^{-3}$ of stars, perhaps. Thus you need remarkably high efficiency for this to compete, rate-wise.

3. A recent entry has been mergers produced in the accretion disks of active galactic nuclei (AGN). The idea is that massive stars can form in AGN accretion disks, or massive stars in the nucleus of an active galaxy can be captured into a disk by repeatedly going through it in tilted orbits. Then, they evolve to become black holes. Once in the disk, they can interact and form binaries, which might spiral in due to gas drag in the disk.

4. It has even been suggested that primordial black holes, which have long been a dark-horse candidate for dark matter, might be clustered in their distribution and thus they might merge to produce the LIGO/Virgo/KAGRA events.

You might think it would be easy to tell the difference: mergers in the field of normal galaxies would tell you it's the first process, in globular clusters or nuclear star clusters would tell you it's the second, in AGN would tell you it's the third, and in the middle of nowhere would tell you it's the fourth. But localization of GW events is so poor (30 square degrees is the *lowest* solid angle yet obtained) that that doesn't help.

Moreover, all proposed channels have so many unknowns and free parameters that, to be frank, there are no predictions: only retrodictions. That is, after each event or each catalog of events, advocates of a given channel will point out how they can accommodate the new observation. What would be an actual prediction would be a firm statement that something *can't* happen; if any outcome works in your model, your model doesn't make predictions.

As an example, early in the era of direct GW observations, people who supported the first channel (massive binary evolution) swore up and down that a completely forbidden zone for them was black holes in the so-called "pair instability gap". The idea is that for a massive enough star, when it gets to the fusion stage of having a core that is primarily oxygen, the required temperature is so large that many of the photons (which were providing pressure whose gradient supported the star against gravity) pair-produce to make low-energy electrons and positrons, which have very low pressure. Thus the core contracts more, the fusion rate goes up, and there is an instability that blows the star apart like a souped-up Type Ia supernova. This means that we shouldn't have black holes between masses of around $50 - 120\ M_\odot$. Then, GW190521 was observed, which had best-estimate black hole masses of $66\ M_\odot$ and $85\ M_\odot$. Seems pretty clear-cut! Obviously, the first channel was eliminated for this one; prior to the detection, all modelers agreed with this. Another channel was necessary, right?

Nope. As soon as the masses of the black holes in GW190521 were announced, people came up with many ways to circumvent the "hard" limits. It's frustrating and disappointing. It also isn't limited to that channel; people who work on the other channels are also happy to accommodate any result. To be a little fair to the modelers, there are a *lot* of uncertainties. But this does mean that a huge amount of basic astrophysics needs to be understood so that the flexibility of the models can be reduced and we can eventually get back to true predictions.

Now, as promised, a brief word about supermassive black hole binaries. Most (maybe nearly all) massive galaxies in the universe have a central supermassive black hole. Galaxies merge with each other. Thus, if the black holes in the merged galaxy can get together, you have a supermassive black hole binary. It's not as easy as it might sound, because the multi-kiloparsec trip to the center of a galaxy requires a lot of dynamical interactions, but detailed studies suggest that this is successful most of the time if the galaxies and thus the black holes are comparable in mass. These are expected to be detected directly with LISA, and indirectly and in aggregate using pulsar timing arrays.

Electromagnetic counterparts

If a gravitational wave event is accompanied by an electromagnetic counterpart then we get a lot more information. For example, since we can localize EM signals beautifully (sub-arcsecond for optical events, for example), then an EM counterpart will tell us where the event happened. This happened with the first double neutron star GW event, GW170817; more about that below. EM counterparts are a lot tougher with two black holes, but it turns out (ask me if you want details!) that although when the black holes have stellar mass it's very likely next to impossible to get a detectable counterpart, when the two black holes are supermassive then there can be enough mass around (as in, AGN accretion disks) to be detectable. There is a lot of work currently on possibilities. One thing I worry about is that if we are to identify which galaxy hosts the GW event (remember, the localization is poor, and will be even with LISA despite it probably managing 1 square degree), we can't just think about ways that the binary SMBH will be visible. We have to have clear evidence that the variation, spectrum, or whatever is actually because of the binary rather than just being part of the seemingly infinite variety of AGN weirdness!

GW170817

On August 17, 2017 (thus 170817), LIGO finally detected a double neutron star coalescence after seeing multiple double black hole coalescences (Virgo was on but the weakness of its signal can't be claimed as a detection). Nature was very kind: the merger happened just 40 Mpc away, which meant that this was (and still is!) the strongest signal yet seen. Moreover, this was accompanied a little less than two seconds later with a gamma-ray burst, and then over hours to days to weeks to years all other parts of the EM spectrum were detected. The source was localized to the outskirts of a particular galaxy. This single event led to multiple remarkable inferences:

1. As had been suspected for many years, double neutron star mergers can produce (short) gamma-ray bursts. This one was off-axis; that is, rather than us seeing it jet-on (which is typical for gamma-ray bursts), we saw it some 20-30 degrees to the side.
2. As had been predicted a few years before the event, this merger produced a lot of heavy elements (things like gold) because not everything fell into the central object (which probably became a black hole within a few seconds, although the case isn't airtight). The matter that spewed out was obviously very neutron-rich, and those neutrons captured onto nuclei and formed heavy elements. This led to a characteristic light curve which, to stress this again, had been predicted. It is possible that double neutron star mergers produce most of the elements heavier than hydrogen in the universe, although this is still debated.
3. Because neutron stars deform under the tidal gravity of a companion, their deformation should leave an imprint on the gravitational waveform; basically, as the stars spiral together, some orbital energy goes into the deformation and (as a weaker effect) the actual production of gravitational waves is changed by the fact that the stars aren't spheres anymore. There was marginal evidence for such effects, but this did tell us that the stars couldn't have been too large. In turn, this gave us new insight about the properties of the matter in these (and by extension all) neutron stars.
4. Some people have also been excited by the prospect that GW170817 and events like it

can be used to measure the Hubble parameter in new ways. The point is that to get the Hubble parameter you need a redshift and a distance. If you can identify the host galaxy (as in GW170817), then you have the redshift. Gravitational waves give you the luminosity distance (side note: in cosmology you have to be careful with your distance definitions; the "luminosity distance" is the distance in Euclidean space at which something of known luminosity would appear at a given flux). Thus, combined, the two measurements give you the Hubble parameter. But there are lots of uncertainties in these measurements. My opinion is that the continued improvement of purely EM measurements of the Hubble parameter will continue to be more precise than those from GW for very many years, but we'll see.

All in all, just six years after the initial direct detection, gravitational wave observations have told us a remarkable amount about the universe. Continued improvement of ground-based detectors, the gradual bettering of pulsar timing array sensitivity, the flight of LISA, and maybe even the analysis of the B-mode polarization of the cosmic microwave background, will give us continued surprises and insights. It's a good time to study gravitational radiation!