Frontiers: Gamma-Ray Bursts

We will now take a look at one of the most exciting and controversial fields in all astrophysics: gamma-ray bursts. These events may have a higher peak photon luminosity than any other events in the universe, and their physics is therefore extreme enough to have motivated a number of exotic suggestions. Their spectra and brightness also may make them great backlighting for the universe, meaning that in principle absorption lines in their spectra can tell us about the composition and evolution of the $z \sim 5 - 20$ universe (although currently no GRBs have been established spectroscopically beyond $z \approx 8.2$). The history of this field is also an object lesson in how new evidence can shift opinions dramatically. At one time or another, a substantial majority of people in the field have believed (with a fair degree of certainty) that gamma-ray bursts are (1) in the Galactic disk, (2) standard candles, (3) caused by merger and inspiral of two compact objects, and (4) the product of a special type of supernova. Current opinion favors (4) for one of the major types of GRB, and probably (3) for the other, but new evidence can always change this. To me, the study of gamma-ray bursts encapsulates much of what makes the scientific process unique. It is a subject filled with rancor and conflict, but the emergence of new data has had its say in a way not available with pure philosophy.

Brief summary of properties

Let's first summarize briefly what gamma-ray bursts are. Indeed, it is somewhat difficult, because unlike many of the sources and phenomena we've discussed, gamma-ray bursts are rather heterogeneous in their properties. The first two properties come from their name: the emission is primarily in gamma rays (with a spectral νF_ν peak in the hundreds of keV), and the events have a limited duration (from milliseconds to about a thousand seconds, as seen so far). There seems to be a broad bimodal distribution of durations, one peak being less than a second and the other being at 10-20 seconds. Unlike X-ray bursts, the profile of the flux with time is not universal. Many bursts have a "FRED" profile (fast rise, exponential decay), but others are more spiky, or have some emission, a long quiescent period, and then have more emission. Within the sensitivity of current instruments, there appears to be of order a few bursts per day in the universe, of which perhaps 10-20% are of the short hard variety.

The distribution of locations of bursts on the sky is consistent with isotropic. The positional uncertainties for the bursts can be many degrees if detected only in gamma-rays, but if followups detect afterglows in the optical or infrared, sub-arcsecond precision is possible. The flux observed at Earth has an extremely broad range between different bursts, from a maximum of about 10^{-3} erg cm^{-2} s^{-1} to the flux limits of detectors, which can be down to 10^{-8} erg cm^{-2} s^{-1}. All bursts that have been localized enough for pointed follow-up have X-ray afterglows lasting days (before they are too weak to detect), and about half have detectable optical afterglows. The spectrum and the time development of the bursts are adequately described by power laws with a few breaks in them. Redshifts (or at least lower limits to the redshift) have been obtained for a number of bursts, clearly indicating that many, perhaps all, bursts are at large cosmological distances.

History of detection

Gamma-ray bursts were first discovered as a byproduct of the Cold War. In the late 1960s there was a concern that the Soviets might test nuclear weapons in space. The US decided that it needed to be able to detect the gamma-ray emission that would result, and it therefore launched the Vela series of satellites. They were alarmed when, starting in 1968, the satellites detected gamma-ray flashes from space! The spatial resolution of the satellites was poor, but eventually it was determined that the flashes came from outside the solar system, so in 1973 the flashes were reported publicly.

In 1979 there was an apparent breakthrough in the study of gamma-ray bursts. On March 5, 1979, nine separate satellites detected a remarkably strong burst (impressive enough that this is simply known as the "March 5 event"). Many of these satellites were far enough away from the Earth that it was possible to localize the direction of this event by timing; an aid to this localization was that the event had an extremely sharp onset. This event came from the N49 supernova remnant in the Large Magellanic Cloud, and later was even more specifically determined to come from an X-ray hot spot in the cloud. This was exciting, because this was the first time that a GRB had been identified with a quiescent source. Moreover, this source repeated; 16 more bursts were seen over the following months. However, it is now accepted that this event was the first identified member of a separate class, soft gamma-ray repeaters, which were later identified as magnetars. At the time, though, this mislead people for a long time, because it appeared that this was clear evidence for a Galactic source of the bursts, and it was so clearly established that it appeared to be a fixed point in the data.

In the 1980s, other bits of evidence appeared to support the local origin of the bursts. Data from the Japanese satellite *Ginga* for several bursts suggested the existence of cyclotron absorption-like features in three bursts, one that appeared very secure. This also seemed to argue strongly for a relatively local origin. The point is that without any persistent sources or direct evidence of distance, a given flux is not informative about the distance (in the dark, a light could be a nearby firefly or a distant airplane). However, the argument was that if the distance was cosmological, the luminosity would be so high as to prevent the formation of lines near a compact object. At the end of the 1980s, virtually the entire community (with the notable exception of Bohdan Paczynski) was sure that gamma-ray bursts mostly came from neutron stars in the disk of the Galaxy.

In 1991, the Compton Gamma-Ray Observatory was launched, as one of NASA's Great Observatories program. The Burst and Transient Source Experiment (BATSE) was particularly well-suited for detection of GRBs, since it had a low flux limit and all-sky coverage. It also had better angular resolution than previous instruments, although even for bright bursts the location was no better than two degrees and for dim bursts it was 30 degrees or worse. Prior instruments had detected no deviation from isotropy, but it was expected that with BATSE's much more sensitive detectors a bias towards the Galactic plane and center would be seen. It was not. However, what was seen was a rollover at low fluxes compared to what would be expected in a Euclidean universe with a constant density of sources.

This radically changed the way that most people thought. The combination of isotropy with a deficit of dim sources is exactly what is seen in cosmological populations of all types. The expanding universe means that beyond a redshift of roughly unity, there is less volume to play with, so if there was a constant comoving density of sources then there would appear to be fewer sources at large distances. When this result from BATSE was established firmly, therefore, most people switched over to thinking that GRBs were cosmological. As a result, previous evidence in favor of a more local origin was discounted (e.g., it was now felt that the *Ginga* "cyclotron lines" were statistically insignificant). However, Don Lamb and colleagues pointed out that a population of high-velocity neutron stars in the halo of our Galaxy could also explain the observed isotropy and falloff. The majority of the community didn't agree with this, although a debate held in 1995 in Washington, D.C. helped convince people that the case wasn't open and shut in favor of a cosmological origin. The main problem was one of physics versus astronomy. Isotropy and a rollover in the brightness distribution has, historically, suggested a distant cosmological origin. On the other hand, it's a lot easier to figure out energy sources on the scale of 10^{38-42} erg than on the scale of 10^{51} erg of gamma rays. For the next breakthrough, a smoking gun was needed.

The main problem was that now that the March 5 event was considered separate, no quiescent counterparts of GRBs existed. This divorced the field of GRBs from the rest of astronomy, and made further progress difficult. One problem was that most GRB localizations were with BATSE, which could only do a couple of degrees at best, and in that kind of area there are an unlimited number of sources. Into this mix, in 1996, came the Italian-Dutch satellite BeppoSAX. Initially the goals with this satellite had nothing to do with GRBs. However, starting in 1997, it was able to localize $\sim$10 bursts per year to accuracies of a few arcminutes. This is a small enough area that optical and X-ray pointed observations could be brought to bear quickly. People then looked for initially bright sources that faded... and found them. In 1999, a rapid pointing even found an optical source that reached 8th magnitude just seconds after the GRB (and that record has since been broken; one burst got to an optical magnitude of 5, which means that if you were in a dark area, you could theoretically have seen it with your own eyes!). The optical observations localize the source to a fraction of an arcsecond, and has allowed spectra to be taken that prove the sources are at high redshifts (up to $z \approx 8$ in the most extreme cases). Thus one can at least say that most to all GRBs are cosmological.

The flux distribution of bursts themselves was originally fit by a model in which the rate of bursts per volume is unchanged throughout the history of the universe, and in which the bursts have a constant rest-frame luminosity (hence, in which they are non-evolving standard candles). This was the standard model for a while, but BeppoSAX observations and their follow-ups have shown that the rate per volume was much higher in the old days; in fact, it appears roughly consistent with the star formation rate as a function of redshift, so the new standard model is that the long bursts happen in star-forming regions. However, the *short* bursts follow a different pattern; they can happen in any type of galaxy, and several are clearly *not* hosted by a galaxy, although they might be within some tens of kiloparsecs of the nearest likely galaxy.

Fireballs and afterglows

In the mid-1990s a number of researchers realized that regardless of what the energy source was, the release of 10^{51} erg or more in a few seconds or less would produce an expanding fireball, and that the interactions of this fireball with the surrounding medium would yield potentially robust signatures, including afterglows. The observation of these afterglows is therefore something of a confirmation of the models, although the models have enough parameters (due to legitimate uncertainties!) and the observations are featureless enough that the association wasn't instantly convincing. The initial models had a very quick release of energy (the standard picture was of merging compact objects, taking milliseconds), and the gamma-rays as well as the afterglow were attributed to shocks and interaction with a clumpy interstellar medium. Ed Fenimore and colleagues showed, however, that this would not produce the observed properties of GRBs, so now it is thought that the gamma-ray bursts reflect the rest-frame duration of the event, and that internal shocks in the outgoing fireball account for the gamma-ray emission. The general success and robustness of the fireball model seems at this time to be a theoretical fixed point in a very uncertain situation.

The Central Engine

The most interesting question related to GRBs is what powers them. The acceptance that they originate from significant redshifts represents a major shift in what has to be considered. When they were thought to originate from ~ 1 kpc, this could be easily explained by any number of processes on neutron stars. But at $z = 1$, a 10^{-5} erg cm^{-1} s^{-1} burst means an isotropic luminosity of 10^{51} erg s^{-1}, which is another story entirely! In fact, this luminosity is much greater than the peak luminosity of supernovae. What, then, could do this?

In much of 1990s, attention focused on mergers of compact objects. The merger of two orbiting neutron stars releases some 10^{53} erg, and therefore would be enough to power a burst. However, there are some potentially major problems with this. First, the time scale of a merger is milliseconds, not tens of seconds as is usually seen (and is the rest-frame time, from Fenimore's work). Second, if there are too many baryons in the fireball, then the energy all goes to them and the resulting Lorentz factor is too low. This would mean that afterglows would take years instead of days, and the peak in emission would be at much lower energies than observed. Therefore, there has to be some way to have the burst occur in a "clean" direction that has a deficit of baryons. There were suggestions that neutrinos and antineutrinos would annihilate and produce leptons far enough away that the requisite cleanliness would be achieved, but this turns out to be too inefficient. In addition, the merger times of binaries can be quite long (billions of years, as is the case for three NS-NS binaries in our Galaxy). One would therefore expect cases in which the binary, having received kicks from the two supernovae, had traveled great distances from their host galaxy, and had delayed merger enough to occur long after star formation had dropped off. One would not expect such a strong evolution as is seen, although there are lots of uncertainties about that as well.

The new standard model for long bursts therefore involves a special type of supernova. The idea is that a massive star evolves quickly, so it doesn't have time to leave its birthplace or explode long after the starburst. The type that produce GRBs are sometimes called "hypernovae". The idea is that the formation of a rapidly rotating $\sim 10\,M_\odot$ black hole in such a supernova will establish a jet and a preferred axis that may clean away baryons. The total energy is perfectly adequate, as well, because these events have a lot of beaming; even so, the true luminosity is about 10^{51} erg s^{-1}.

But short bursts have to come from some other source. For one thing, unlike the long bursts, no short bursts have ever been identified with a supernova. Their afterglows are much fainter than the afterglows of long bursts, which makes host identification a lot tougher. Because some are outside their likely host galaxies, and because those host galaxies don't have to be forming stars actively, the most favored current idea is that the short bursts are produced by the merger of two neutron stars, or of a neutron star and a black hole. The supernovae that produce the compact objects can give a kick to the system that could eject it from the galaxy. The time to merger, from the initial production of the double compact object system, could be up to billions of years, which gives plenty of time for the system to leave the galaxy and for the star formation in the galaxy to quiet down.

Where to go from here

The big hope in the community is that there will be a detection of a gravitational wave event that is coincident with a short gamma-ray burst. This would tell us definitively that compact object mergers produce bursts... at least some bursts, since we have learned from the history of the subject that there could be quite a bit of heterogeneity in the class! The likelihood of actually seeing a short bursts with a gravitational wave event is, however, not that great. Bursts are beamed, so to see one it has to be beamed at us. For the event to be visible using ground-based gravitational wave detectors, the merger would have to be at least twice as close as the closest identified short gamma-ray burst. Mind you, we only see a minority of bursts, because detection depends on us looking at the right time, but all things considered there might only be 1-2 bursts per year within the LIGO detection volume, and we probably won't be looking at that time.

Thus people have spent a lot of effort thinking about other electromagnetic signals that could be associated with short bursts. If a neutron star is consumed by a black hole or if it merges with another neutron star, it's likely that a decent amount of mass will not go into the ultimate remnant that is formed. That mass will be very neutron-rich, and simulations suggest that it will propagate outward at $0.01 - 0.1c$.

What will happen with this outflow? Neutrons are unstable at low density, so neutrons will decay as the outflow expands. Those protons and electrons can combine with neutrons to produce alpha particles, and those particles can combine with others, and so on. Indeed, in such a neutron-rich flow you have the ideal conditions for *r-process nucleosynthesis*. Here the "r" stands for "rapid", and it means that nuclei can capture neutrons faster than those nuclei can decay to an equilibrium state (called beta equilibrium). This is how heavy elements, well past iron, are formed. There is an ongoing debate, but there is good evidence to suggest that much to most of the r-process elements in the universe (like gold! Scrooge McDuck cackle...) are formed in NS-NS or NS-BH mergers. The other candidate is core-collapse supernovae; a basic distinction between the r-process mechanisms is that NS-NS or NS-BH mergers are rare (perhaps one per hundred thousand or million years per galaxy) but produce a lot of r-process elements (maybe $10^{-3} - 10^{-2}$ solar masses), whereas core-collapse supernovae are much more common (on the order of one per hundred years per galaxy) but produce far smaller amounts of r-process elements per event (the amount is not clear). By the way, "s-process nucleosynthesis" is "slow" in the sense that between neutron captures there is time for decay of the nuclei to beta stability.

But back to potential quasi-isotropic EM counterparts. As the material flows out, energy continues to be injected by the formation of, and radioactive decay of, the nuclei that are produced. Because the outflow will be large and will be visible from any angle, this might be easier to see as a counterpart to a gravitational wave event than would be the highly beamed gamma-ray burst itself.

There has been considerable study of this process, with evolving conclusions. Early work assumed that the outflowing matter would have an opacity characteristic of what are called iron-peak elements (i.e., elements with atomic numbers similar to iron). The luminosity would then peak when the optical depth through the matter becomes of order unity, and with iron-peak opacities, optical depth unity would be reached in about a day, the emission would be peaked at optical wavelengths, and the maximum luminosity would be about a thousand times as great as a classical nova. That last property motivated some researchers to dub this possibility a "kilonova", and others preferred "macronova".

However, it is now understood that because the primary nuclei will have 2-3 electron shells more than iron peak elements, and therefore vastly more electronic transitions, the opacity will be tremendously larger (by at least factors of tens) than was previously thought. This means that (1) the outflow will become transparent much later, (2) as a consequence, the generated energy will largely go to expanding the outflow rather than to photons that escape, (3) because it takes longer and the energy is less the luminosity will be much less, and (4) because by the time the outflow becomes optically thin it will be larger than previously thought, that luminosity will be spread over a huge area and thus the temperature will be reduced a lot compared with previous expectations. In particular, the main emission will be in infrared rather than optical, and because our instrumentation is much less sensitive in infrared than in optical this is bad news all around. There might be some viewing angles (as in, near the orbital axis of the system) where optical emission dominates, but this isn't clear. There was an event in 2013 that was thought for a while to be a good kilonova candidate, but in light of current theory it seems likely to have been something else.

An isotropic-EM candidate that has not been given as much attention was suggested by our own David Tsang and his colleagues in 2012. Suppose that two neutron stars spiral in gradually toward each other via emission of gravitational waves. At some point, the orbital frequency will match with crustal modes, which means that the crust could be catastrophically cracked due to the resonant forcing. Dave et al. suggested that this might explain some otherwise puzzling apparent precursors to a few short gamma-ray bursts. This would emit a comparable amount of energy to a macronova, but the emission would be relatively prompt. He's still working on the exact properties.

We'll see! A coincident GW-EM event, which most bettors would guess would be related to a short gamma-ray burst rather than a long burst, would contain a lot of information. Keep tuned!