INTRODUCTION

Serendipity is jumping into a haystack to search for a needle, and coming up with the farmer’s daughter. —Julius H. Comroe Jr.

1.1 Serendipity during the Cold War

Before Mythbusters and The A-Team made big explosions cool, big explosions were decidedly uncool. The threat of nuclear war between the United States and the USSR (and, perhaps, China)—made blatantly real during the Cuban Missile Crisis in October 1962—had become a fixture in everyday life. One year after the crisis, seeking to diffuse an escalating arms race and the global increase of radioactive fallout from nuclear weapons testing, Soviet Premier Nikita Khrushchev and U.S. President John F. Kennedy agreed to the Partial Test Ban Treaty. Ratifying nations agreed that all nuclear weapons testing would be conducted underground from then on: no longer would tests be conducted in oceans, in the atmosphere, or in space.

The United States, led by a team at the Los Alamos National Laboratory, promptly began an ambitious space satellite program to test for “non-compliance” with the Partial Test Ban Treaty. The existence of the Vela Satellite
Program was unclassified: the rationale, experimental design, and satellite instrumentation were masterfully detailed in peer-reviewed public journals while the program was on going. The concept for this space-based vigilance endeavor was informed by the physics of nuclear explosions: while the optical flash of a nuclear detonation could be shielded, the X-rays, gamma rays (sometimes written as $\gamma$-rays), and neutrons that are produced in copious numbers in the first second of an explosion are much more difficult to hide; we call the measurement of these by-products the “signature” of a nuclear detonation. Going into space for such surveillance was a must: the Earth’s atmosphere essentially blocks X-rays, gamma rays, and neutrons from space.

While the signatures of nuclear detonations were well understood, the background radiation of light and particles in space was not. To avoid false alarms caused by unknown transient enhancements in the background, satellites were launched in pairs—both satellites would have to see the same very specific signatures in their respective instruments for the alarms bells to sound. Widely separated satellite pairs also had the advantage that most of the Earth could be seen at all times. While the Vela orbits provided little vantage point on the dark side of the Moon—a natural location to test out of sight—the gamma rays and neutrons from the expanding plume of nuclear-fission products would eventually come into view. In total, six pairs (Vela 1a,b through Vela 6a,b) were launched between 1964 and 1970.

As evidenced by the Vela Satellite Program, the U.S. was obviously very serious about ensuring compliance.
That the capabilities of the program were open was also a wonderful exercise in cold war gamesmanship—you are much less inclined to break the rules if you are convinced you will get caught.

While hundreds of thousands of events were detected by the Velas—mostly from lightning on Earth and charged particles (cosmic rays) hitting the instruments—the telltale signatures of nuclear detonation were thankfully never discovered. Those events that were obviously not of pernicious or known origin were squirreled away for future scrutiny.

Starting in 1969, Los Alamos employee Ray Klebesadel began the laborious task of searching, by eye, the Vela data for coincident gamma-ray detections in multiple satellites. One event, from July 2, 1967, stood out (figure 1.1). Seen in both the gamma-ray detectors of Vela 4a and Vela 4b (and weakly in the less sensitive Vela 3a and Vela 3b detectors), the event was unlike any known source. Though there was no known solar activity on that day, the event data themselves in one satellite were incapable of ruling out a Solar origin, especially if it was a new sort of phenomenon from the Sun. Over the next several years, other intriguing events similar to the July 2nd event were seen in the Vela data. By 1972, Klebesadel and

*GRBs are conventionally identified by the date that they are detected. The first two numbers correspond to the year, the next two correspond to the month, and the last two correspond to the day. In the case that more than one GRB is detected on the same day, letters are appended to the burst name. For instance, the first GRB detected was GRB 670702. GRB 101221B would be the second GRB detected on December 21, 2010. (This naming convention clearly runs into problems after a hundred years of observations!)
Figure 1.1. The first gamma-ray burst, GRB 670702, detected by the Vela 3a,b and 4a,b satellites. Shown is the gamma-ray light curve of the event, which is the instrumental brightness (counts per second) versus time measured since the event triggered the on-board instruments. The count rate before the burst is not zero due to persistent gamma-ray sources in the sky and random instrumental events in the detector. But when the event arrives at the detector, it vastly outshines the background. GRB670702 was a long-duration GRB, lasting more than eight seconds and showing variability on timescales less than one second. Adapted from J. Bonnell, A Brief History of the Discovery of Cosmic Gamma-Ray Bursts. http://antwrp.gsfc.nasa.gov/htmltest/jbonnell/www/grbhist.html (1995).

his colleagues Ian Strong and Roy Olson had uncovered sixteen such events using automated computer codes to aid with the arduous searches.

What were these bursts of gamma rays? To answer that question, the Los Alamos team recognized that it had better determine where on the sky the events came from. Pinpointing the direction of a light source is easy if you
can focus it: this is what cameras used for photography and the human eye do well with visible light. But X-rays, and especially gamma rays, are not amenable to focusing: the energies of these photons are so high that they do not readily interact with the free electrons in metals and so cannot be reflected to large angles. The focusing of light without large-angle reflection is exceedingly difficult. The best the X-ray and gamma-ray detectors on the Velas could do was stop those photons, recording both the energy deposited in the detectors and the time that the photon arrived at the satellite.

The arrival time of the photons from specific events held the key to localization. Just as a thunderclap is heard first by those closest to the lightening bolt, an impulsive source of photons would be seen first in the satellite closest to the event and then later, after the light sweeps by, with the more distant satellite. Light (and sound, in the case of thunder) has a finite travel speed. Since the Vela satellites were dispersed at large distances from each other (approximately 200,000 kilometers) the difference in the arrival times of the pulses could be used to reconstruct the origin on the sky, the location on the celestial sphere. As figure 1.2 shows, an event seen in two satellites produces an annular location on the sky, and an event seen in three satellites produces a location in two patches on the sky.

It turns out that, with precise-enough timing, a network of satellites can localize sources within the Solar System not only on the sky (two dimensions) but also in the third dimension (distance). A few GRBs have been shown to be unequivocally at distances beyond the Solar System using such a timing technique.
Figure 1.2. Triangulation of gamma-ray bursts using the arrival time of light at different satellites. Image (a) shows the geometry of localization in two dimensions. The GRB propagates down through the Solar System as a plane wave depicted here with dashed lines. The event triggers the satellite at left first, then the satellite at right. By measuring the difference in trigger time ($\Delta t$) and knowing the distance $r$ between the satellites, the angle $\theta$ toward the event can be inferred. Here $c$ is the speed of light. Image (b) shows the localization on the celestial sphere (gray circle) using three satellites. Two independent angles are determined using the difference in arrival times between satellites 1–2 and 2–3. The uncertainty in those angles $\delta \theta$ is directly determined by the uncertainty in the precise time difference between satellite triggers. The direction toward the GRB is determined by the regions on the sky where the annuli overlap.

This triangulation capability, albeit crude, was sufficient to convince the Los Alamos team that it had uncovered a class of events that was not coming from the Earth, Sun, Moon, or any other known Solar System object. In 1973, Klebesadel, Strong, and Olson published their findings in the Astrophysical Journal, one of the venerable peer-reviewed journals used for describing scientific
results in astronomy. The paper\(^5\) titled “Observations of Gamma-Ray Bursts of Cosmic Origin” marked the beginning of the gamma-ray burst (GRB)\(^6\) enigma that to this day captivates the imagination and keeps astronomers scratching their heads.

The word serendipity is overused and misused in science. Most mistake a serendipitous discovery to be synonymous with an unexpected (and unforeseen) discovery. But, as Julius Comroe’s colorful analogy in the epigraph describes, serendipity demands both an unexpected discovery and an entirely more pleasant discovery than the one being pursued. While GRBs certainly were unexpected and unforeseen,\(^7\) they were also much more scientifically valuable than what was being sought after: instead of the detection of a nuclear test by an enemy, a discovery that in the 1960s would have set the world down a dangerous and dark path, GRBs were a fresh light from the dark heavens. Indeed, their mysterious nature would captivate a generation of astronomers. The discovery of GRBs—not just their detection but the recognition that the events represented a new phenomenon in nature—was truly a serendipitous moment in modern science.

1.2 A New Field Begins

Members of Klebesadel’s team announced the discovery of GRBs at the June 1973 meeting of the American Astronomical Society, a few days after the publication of their seminal paper. In that meeting (and in the paper) they described their observations testing the hypothesis that
GRBs originated from supernovae (SNe) in other galaxies; this was the only physical model for the origin of cosmic bursts of gamma rays available at the time. By trying to correlate a GRB in time and sky position to all known SNe, the attempt to connect GRBs to the then-brightest explosions in the universe “proved uniformly fruitless.”

Determining what objects and what events on those objects produced GRBs quickly became a hot topic. By the end of 1974, more than one dozen ideas for the origin of GRBs had already been published. The theories spanned an astonishing range of possibilities, from sunlight scattering off fast-moving dust grains to comets colliding with white dwarfs (WDs) to “antimatter asteroids” smashing into distant stars. All viable models necessarily accommodated the available data, but the GRB data were simply too sparse to constrain a talented and imaginative group of eager scientists.

More data would be needed to narrow down the range of plausible models. By the end of 1973, the Los Alamos team had found a total of twenty-three GRBs. Teams working with other satellites equipped with gamma-ray detectors also began reporting detections of GRBs, even some of the same events seen by the Vela satellites. New programs were conceived to find more GRBs and observe them with more sensitive detectors. The supposition—if not just a hope—was that with better data some telltale signature of the origin of the events would emerge. Unknown to those sprinting to find the answer, for all but a few special events, those telltale signatures would take over thirty years to uncover (a veritable marathon in modern science).
Light does not easily betray its origin: there is nothing in a gamma-ray photon itself that can tell us how far it traveled, nor can we learn directly just how many of those energetic photons streamed away from the event that produced the GRB. Without a measurement of the distance to a source, the pool of possible culprits is simply too broad: since we have a general sense of the types and the spatial distribution of objects in a given volume of space, if we knew that GRBs arose from distances on the size scale of the Solar System (for example), then there could be only a select set of objects responsible (comets, asteroids, planets, etc.). At a more fundamental level, without knowledge of distance, it is all but impossible to know how much energy the source put out. And without that knowledge the range of physical mechanisms that could be responsible for the sudden release of all that energy is also too broad. Case in point: a street lamp appears about as bright as the Sun, yet the scales of energy output are vastly different as are the physical origins of the light.

Since light does not directly encode distance, how do astronomers determine distance to astrophysical entities? If sufficiently nearby, objects appear to be in slightly different places on the sky for observers at different places. This measurement of parallax yields a direct triangulation of distance but is exceedingly difficult to determine for most objects beyond a few hundred light years away from Earth. Beyond that, for all but a few special cases, we must infer distance by associating some source with a source whose intrinsic brightness or size we think we know (usually because we think there is an analogous system within the parallax volume).
The key, then, for GRBs would be to associate the events with something else whose distance we could more readily infer. In this respect, the inability to measure a precise two-dimensional position of a GRB on the sky directly hampered the ability to measure the all-important third dimension. Getting better positions of GRBs on the sky became the driving impetus behind the next several decades of GRB observational projects.

1.3 Precise Localizations and the Search for Counterparts

By the late 1970s, not only were there more satellites flying with higher-sensitivity detectors, but some of these satellites were far from Earth (in particular, near Venus and the Sun). This interplanetary network (IPN) gave a significant improvement on the timing localizations of GRBs (see figure 1.2). At a distance of up to $d = 2$ astronomical units (AU) (twice the distance from the Earth to the Sun), a pair of satellites with the capability to determine the time of the onset of a GRB to an accuracy of $\delta t = 0.1$ seconds would be able to produce an annular localization ring of thickness $\delta \theta \approx \delta t \times c / d = 10^{-4}$ radian $\approx 1/3$ arcminute. By 1980, there were a handful of well-localized (to tens of square arcminutes or better) GRBs, and by the end of the 1980s there were dozens of well-localized GRBs using the interplanetary timing technique.

In a spatial area on the sky, while millions of times more accurate than the first GRB positions, these square-arcminute localizations proved insufficient to rule out most
models. If all *error boxes* on the sky contained a bright star or a bright galaxy, the association with a certain physical class of objects would be secure. This was not the case. Instead, GRBs must have been associated with something faint or unseen. The enormity of the Universe and its bountiful constituents is a real shackle in this respect: in even the most empty directions looking out through our Galaxy, a single error box would contain tens of thousands of faint stars and tens of thousands of faint and distant galaxies. This amounted to a line up of culprits simply too big to get any significant traction on the question of distance and, ultimately, the origin of GRBs.

Observing at gamma-ray wavelengths is just about the worst idea if the goal is to localize an event precisely. But if a *counterpart* at some other wavelength could be associated positively with a specific GRB, then the location of the GRB could be more precisely identified. The most credible counterpart would be an event, consistent with the GRB position, that seemed to happen at around the same time as the GRB—it is actually quite natural to expect that some energy should be pumped into channels other than gamma rays, but just how much energy and on what *timescales* that energy would emerge across the *electromagnetic spectrum* were not well known. As mentioned, no (visible-light) supernova counterparts were found by Klebesadel’s team during the early years of the field. And, despite several efforts in the 1970s and 1980s to discover a concurrent signal from radio to infrared to optical to X-ray wavelengths, no convincing counterparts were found. There was another possibility: if the “engine” (see §2.3)
that produced the GRB had been active previously, then perhaps a transient counterpart could be found in the old image archives of the same place on the sky. Some tantalizing archival transients were indeed uncovered, but none proved robust under detailed scrutiny.

1.4 The March 5th Event and Soft Gamma-Ray Repeaters

On March 5, 1979, an intense gamma-ray event triggered the IPN satellites distributed throughout the inner Solar System. Within the first tens of milliseconds, the event became so intensely bright that the detectors on board all nine satellites—even those pointing away from the event direction—saturated: photons arrived at such an appreciable rate that they could not be recorded fast enough. This blinding was only temporary, however, as for the next few minutes some detectors recorded a fading signal with an unusual character. Unlike all the other GRBs that had been seen to date, this decaying tail appeared to vary periodically. The fact that the initial pulse “turned on” so rapidly suggested that the size of the emitting surface was small, less than the size of the Earth. The eight-second periodicity in the signal was also an important clue for understanding the progenitor. In nature there are only a few classes of physical configurations that give rise to periodic brightness changes; of the most interest are the pulsations of an emitting surface, oscillations through an emitting object, and rotation. The natural (most physically simple) timescale for changes in pulsations and oscillations
is the time $\tau$ it takes for sound waves to cross the object, $\tau \approx l/c_s$ (where $l$ is the characteristic size of the object and $c_s$ is the speed of sound in the object). For rotation, that timescale is the period of the rotating object. Ordinary stars, like the Sun, have much longer sound-crossing times and rotation periods than eight seconds. On timing arguments alone, one is quickly pushed to consider a very dense (and hence large $c_s$) and/or small object as the likely origin of such an event.

The March 5th event, by virtue of its very short and intense pulse, was a superb client for timing localization by the IPN. Initial analyses led the event to be localized to a few square arcminutes, which was further refined to 0.1 square arcminutes with multiply redundant consistency checks, given the number of satellites participating in the timing network. Remarkably, the event appeared to lie on the outskirts of a well-known supernova remnant in the very nearby galaxy called the Large Magellanic Cloud (LMC).

While the distance to the event could not be unambiguously determined, the evidence connecting it to something small/dense and the SN remnant was overwhelming. It became readily apparent that a neutron star (NS)* very likely was the cause of the March 5th event. First, NSs are small and dense enough to account for the rapid rise-time observation. Second, NSs were known to be rapidly spinning, some now observed as fast as five-hundred times

*A neutron star (NS) is an inert mass supported by degeneracy pressure from neutrons and nuclear forces. It is a close analog of WDs (supported by electron degeneracy) but at higher mass and density.
per second (to be sure, the eight-second periodicity was actually longer than most known NS rotation periods). Third, rotating NSs were known to exhibit pulsed emission, though usually at radio wavelengths and not at gamma-ray energies. Fourth, NSs are the by-product of most types of supernovae, receiving substantial “kicks” of up to $\sim 1,000$ km/s during their formation. This natal kick provided a natural explanation for why the event would be located off-center from the SN remnant—in the time since the SN (reckoned to be about 5,000 years from analysis of the SN remnant), the neutron star would have traveled an appreciable distance from the center of the explosion.

If the March 5th event occurred in the LMC, about $r_{\text{LMC}} = 75$ kiloparsec ($kpc$) from the Sun, the total energy release ($E_{\text{March 5th}}$) could now be easily calculated using the $1/r^2$ law:

$$E_{\text{March 5th}} = 4\pi r_{\text{LMC}}^2 S \approx 7 \times 10^{45} \text{ erg},$$

where $S \approx 10^{-4}$ erg cm$^{-2}$ was the total fluence of the event recorded.* This $E_{\text{March 5th}}$ is a fantastically high energy release for an event dominated by a giant subsecond pulse of gamma rays. This amounts to as much energy as the Sun emits in more than a thousand years. Still, this is a small fraction of the kinetic energy available to the NS and an even smaller fraction of the available restmass energy (see §2.2).

*When discussing physical quantities, I generally use the conventional centimeter-gram-second system (cgs) beloved by astronomers. Units of energy are often given as erg (from the Greek word, ergon, meaning "work"). One erg is approximately the kinetic energy of a mosquito buzzing around at $\sim 30$ cm/second. One erg equals $10^{-7}$ Joules.
While the detailed physics of the mechanism(s) at play would be debated for years, the inference of the March 5th event as having arisen from a neutron star was a remarkable feat. Nevertheless, the March 5th event was quickly realized as anomalous when compared to the other dozens of GRBs observed previously. Aside from the periodicity and the localization in a satellite galaxy of the Milky Way, (a) it was hundreds of times brighter than the brightest events seen in the entire decade of GRB observations, (b) the initial pulse was shorter in duration than in all but $\sim5$ percent of known events, (c) the pulsating tail was “softer”—having a lower characteristic frequency where most of its energy was radiated—than other events. Perhaps the most unusual observation was that there were more events from the same location on the sky, the first just fourteen hours after the March 5th event, then more in the following month. Repetition of other GRBs had never been witnessed previously. The inferred energy release was difficult to reconcile with the demographics of the GRB population as a whole: if all GRBs were variants on the March 5th event, then either March 5th was too bright (if all GRBs were of Galactic origin) or too faint (if all GRBs were of extragalactic origin).

Rather than claim victory in finally understanding the origin of GRBs, those that wrote about the March 5th event immediately touted its oddities compared to the rest.

*More than one decade later, the quiescent X-ray emission from the location of the March 5th event was found, confirming the basic expectations of a hot NS progenitor.*
of the population. Indeed, while some events have been even more energetic than that of March 5th, the event of March 5th is now considered a prototype of a giant flare from a special subclass of GRBs called *Soft Gamma-ray Repeaters* (SGRs). Over the past three decades, six more SGRs were discovered and another two had similar characteristics to the March 5th event without confirmed repetition. All these events appear to have occurred from neutron stars within the Milky Way. In 1998, the spin of another SGR was observed to slow at a rate consistent with braking due to strong magnetic fields. It is now believed that all SGRs in our Galaxy (and a class of anomalous pulsars that radiate at X-ray wavebands) are so-called *magnetars*, NSs with extraordinarily high magnetic fields.

1.5 BATSE and the Great Debate

Despite marked advances in understanding the origin, distance scale, and energetics of the March 5th SGR, the mystery for the majority of events raged. Throughout the 1980s, data collection on new GRBs proceeded apace and, by the end of that decade, hundreds of events had been reasonably well localized on the sky and their basic properties measured. Two distinct global properties of “classical GRBs” began to emerge—the location and the brightness distributions—both with important implications for the distance scale of GRBs and hence their origin.

The locations of GRBs on the sky appeared to be randomly (*isotropically*) distributed: that is, there was no indication that any one direction on the sky was especially
more apt to produce GRBs than any other (see figure 1.3). If GRBs were due to neutron stars strewn throughout the disk of the Galaxy, for instance, the locations of GRBs on the sky should have been preferentially located near the Galactic plane (as is seen with SGRs). If associated with older stars in the roughly spherical “bulge” of the Milky Way, GRBs would have been preferentially located in the direction toward the Galactic center and less so toward the opposite direction.\textsuperscript{21} The inference that the Sun was roughly at the center of the GRB distribution in space, while casting aside some models, still allowed for a variety of distance scales: from a fraction of a light year to billions of light years.

The brightness distribution of GRBs appeared to show that we were seeing out to the edge of the GRB population: there were too few faint GRBs relative to the number expected if GRBs were uniformly (“homogeneously”) distributed in space. Brightness was most straightforwardly measured as the peak flux ($P$, with units [erg s\(^{-1}\) cm\(^{-2}\)]) in the light curve of a GRB. The brightness distribution is usually measured as the number, $N(>P)$, of GRBs brighter than some peak flux $P$ per year. If the peak luminosity ($L$, with units [erg s\(^{-1}\)]) of all GRBs is the same, then, using the 1/r\(^2\) law, for a given flux $P$ we would see all the GRBs within a maximum distance:

$$d_{\text{max}} \approx \sqrt{\frac{L}{4\pi P}} \propto P^{-1/2}. \quad (1.1)$$

All the GRBs to that distance would be brighter than $P$ by construction. The number of GRBs we would detect
Figure 1.3. Distribution of different classes of objects and phenomena on the sky. Both top and bottom are projections of the celestial sphere such that the center of the projection is the direction toward the center of the Milky Way (in the constellation Sagittarius) and each line of constant latitude and longitude are spaced at 30-degree intervals. (top) Globular clusters, which appear diffuse, but centered around the Galactic center. The locations of the nine known or suspected SGRs (as of July 2010) are also shown (star symbol). All but the March 5th SGR (which is in the LMC), are within 6 degrees of the Galactic plane (represented by the longest horizontal line). This strongly indicates that the progenitors of SGRs are related to objects in the disk of our Galaxy. (bottom) In contrast, the distribution of GRBs (taken from the first few years of the BATSE experiment) appears randomly distributed throughout the celestial sphere; that is, GRBs appear to be isotropic. Adapted from B. Paczyński, PASP 107, 1167 (1995).
to that brightness (or brighter) in one year would just be the volume times the intrinsic rate ($R$, in units of [event yr$^{-1}$ per volume element]): $N(>P) \propto V \times R \propto R \times d_{\text{max}}^3 \propto R \times P^{-3/2}$. So with a homogeneous distribution, we expect that the number of faint GRBs should grow as a powerlaw proportional to $P^{-3/2}$, where the constant of proportionality scales directly with the intrinsic rate $R$: for every ten times fainter in flux we observe, we would nominally expect about thirty-two times more GRBs. While this was indeed seen for the brightest events, there was a flattening at the faint end of the brightness distribution. This flattening was highly suggestive that we were seeing the “edge” of the GRB distribution in space, an important clue in understanding the distance scale. But without knowing the intrinsic luminosity $L$, we could only infer the shape of the distribution, not the scale. It was like seeing a picture of a building but not knowing if it was of a miniature in a snow globe or the life-sized version.

One of the primary criticisms of the interpretations of apparent isotropy and inhomogeneity was that the population of GRBs observed thus far were observed with a variety of instruments on different satellites. Each detector had its own range of sensitivity to GRBs in energy and duration. One GRB might trigger one detector but not get recorded by another. Even in the smaller subsets of GRBs detected by a single instrument, where some of the concerns of a heterogeneous dataset could be mitigated, fainter events were not detected with as high efficiency as

*Plotting the logarithm of $N(>P)$ versus the logarithm of $P$ would show a distribution following a line with slope $-3/2$. 
brighter events; detector threshold effects were a natural explanation for the apparent paucity of faint events. The robustness of the isotropy measurements were also in question: what if the instrument spent more time looking in one direction than another? In that case, the observed distribution on the sky would not represent the true distribution. By 1990, given these observational caveats, there was enough wiggle room in the global statistics to allow for almost any distance scale and hence a broad range of progenitors.

To settle the questions about the peculiarities of the data collected over multiple telescopes, a single experiment was required, whose sky coverage and trigger efficiencies were well modeled, capable of detecting GRBs to significantly fainter levels than before. Such an experiment had been in construction at Marshall Space Flight Center (in Huntsville, Alabama) throughout most of the 1980s, having been accepted for funding by NASA in the late 1970s. The Burst and Transient Source Experiment (BATSE) was launched by space shuttle Atlantis on April 5, 1991, on board the Compton Gamma-Ray Observatory (CGRO; along with three other high-energy experiments). BATSE was about ten times more sensitive than previous GRB missions, allowing it to find roughly one burst a day for what would be its nine-year mission. BATSE also had a unique configuration on the spacecraft that allowed it to localize individually the 2,700 GRBs it discovered: the BATSE detectors were placed at the eight vertices of CGRO, allowing at least three detectors to “see” any place in the sky at one time. Since the intensity of light is proportional to the angle it makes with a detector, the
relative GRB flux recorded in the detectors could be used to triangulate the location of the GRB on the sky. For the first time, a single satellite would be capable of positioning a GRB to \( \sim 10 \) degrees on the sky. Moreover, the efficiency for discovery of GRBs in different places on the sky was well understood, allowing for the cleanest reconstruction of the true spatial distribution of GRBs on the celestial sphere.

Given the strict criteria for triggering on a GRB, the BATSE efficiencies for detecting faint bursts were also well understood. The result was a remarkable and clean vindication of the two emergent properties of the GRB population. In the first few years of operation, the statistical consistency of isotropy of GRBs on the sky was confirmed beyond all reasonable doubt, as was the rollover of the brightness distribution. Any viable model for GRBs would need to place the Sun at (or very near) the center of a GRB population in space where the apparent rate of occurrence of the farthest events was less than those nearby.

Only two progenitor distributions survived the brutal BATSE scalpel. The first was a population of progenitors in the outer halo of the Milky Way, producing GRBs at such appreciable distance (\( \sim 100 \text{ kpc} \)) that the \( \sim 8 \text{ kpc} \) offset of the Sun from the Galactic center would not be noticed by BATSE as a slight anisotropy. The paucity of faint events would be explained by the finite volume of the Galactic halo. The other viable scenario was a distribution of progenitors associated with galaxies billions of light years from the Milky Way. The isotropy would then be a natural consequence of the nearly perfect homogeneity of the Universe on large scales in all directions. This homogeneity
had been observationally confirmed with the measurement of isotropy of radio quasars, bright and distant massive black holes (BHs) at the centers of distant galaxies. The brightness distribution of GRBs could be explained both by the finite age of the Universe and by effects due to the expansion of the Universe itself.

The results from BATSE nucleated two schools of thought: either GRBs were from “extragalactic” distances, and hence immensely and almost unfathomably bright, or from less extreme “Galactic” distances. The general consensus before the BATSE results was that GRBs were of a Galactic origin, but the BATSE results swayed many to the extragalactic camp. Each school had its vocal advocates. To commemorate the seventy-fifth anniversary of the famous 1920 Herber Curtis and Harlow Shapley debate over the then-controversial size scale of the Universe, a second “Great Debate” was held in 1995 in the very same auditorium as the first debate in the Natural History Museum in Washington, D.C. Arguing for the Galactic distance was Donald Q. Lamb from the University of Chicago. Arguing for the extragalactic distance scale was Bohdan Paczyński (pronounced “Pah-chin-ski”) of Princeton University. After more than an hour of point and counterpoint, a poll of the GRB-aficionado-laden audience in the crowded room showed a rough split between the scenarios. The crux of Paczyński’s argument was that, regardless of the details, the most natural explanation for the isotropy and inhomogeneity was an extragalactic origin. Lamb was more tactical, noting for each relevant observation how a Galactic origin could accommodate the data. Most important for the Galactic argument was
that, thanks to the March 5th event, we had an existence proof: we knew that at least some types of neutron stars were capable of making at least some types of gamma-ray transients. Dating back from the 1970s, there were some ideas of what could produce GRBs at extragalactic distances, but none was a particularly mature theory, and none was well tested by observations of the day.

In the social event immediately after the Great Debate, I, a precocious undergraduate student and fond of the Galactic picture, challenged Prof. Paczyński on some of the finer parts of his argument (which, of course, he parried well). He ended our conversation with a story about how most data had for years been showing that the Milky Way had a bar-like structure in the center, like many other galaxies, but that it took the death of the main proponent of the “no-bar” hypothesis for global consensus to glom on to the bar hypothesis. That albeit well-respected proponent had failed to convince the next generation of scientists why his story was compelling in the face of the data. As Professor Paczyński admonished, controversy and uncertainty in science are never settled by just talk, let alone a single debate. They are settled by new observations and insights about those observations. Eventually the “incorrect” view of the way in which the Universe operates is simply shown to be wrong, and all the remaining skeptics either change their minds or die.

1.6 The Afterglow Era Begins

It was abundantly clear in 1995 that new observations would be needed to settle the distance scale debate once
and for all. Just as always, the main hope remained to find a precise location of classical GRBs and associate those positions with known objects. Since the single-satellite precision of BATSE was about as good as that detector configuration could provide, the key was to find counterpart events at other wavebands. It was known from precise IPN locations that no long-lived (weeks to months) optical transient accompanied the events. But thanks to the nearly instantaneous relay of new GRB positions from BATSE, several groups began construction of optical telescopes awaiting email alerts from GRB satellites; this new generation of telescopes was capable of rapidly observing new GRB positions. If GRBs were accompanied by optical flashes lasting just a few minutes, the robotic “on-call” telescopes would have a chance of catching the fleeting light. Theory of GRB emission mechanisms, developed in the early 1990s, also predicted radio transients from GRBs (chapter 3); as such, some groups began a search at radio wavebands with premier facilities like the Very Large Array (VLA) in New Mexico. None of the early searches of BATSE positions proved fruitful.

A group from the University of California, Massachusetts Institute of Technology (MIT), and Los Alamos proposed a new satellite mission in the mid-1980s that would seek to localize GRBs using the X-rays that where known to accompany many events. In what shaped up to be a multinational collaboration with France and Japan, the High-Energy Transient Explorer (HETE) was designed to trigger on new GRBs using an instrument without positional capability, then use another instrument to localize the X-rays from that GRB. The concept behind
the X-ray camera was to create a mask pattern that would block some of the incoming X-ray photons but let the others fall on a grid of position-sensitive X-ray detectors. The mask would create a distinct shadow depending on the direction of the GRB, which in turn could be used to reconstruct the location of the event to IPN-level accuracies (tens of square arcminutes uncertainty regions) but within a few seconds of the trigger. Unfortunately, HETE was lost shortly after launch in 1996, but the second incarnation, HETE-2, was put together with mostly spare parts and launched in 2000; HETE-2 would quickly vindicate the new X-ray localization approach.

While GRBs were not the main scientific priority of a new Italian-Dutch experiment, the “Satellite per Astronomia a Raggi X” (commonly known as BeppoSAX), it did carry a Gamma-Ray Burst Monitor (GRBM) that stared at the same place on the sky as the X-ray-coded mask imagers (collectively called the Wide-Field X-ray Camera [WFC]); those instruments were capable of seeing roughly 1/15th of the sky at any time. Launched in April 1996, BeppoSAX demonstrated that summer that a burst triggered in the GRBM and seen in the WFC field of view could be localized to $\sim 5$–10 arcminute radius. Such localization accuracy was often obtained with the IPN (see §1.3), but BeppoSAX had the ability to determine such positions in a matter of a few hours, whereas IPN localizations typically took days to determine. The speed of finding a precise GRB position would prove to be a crucial capability.

On February 28, 1997, BeppoSAX localized a GRB using the WFC and the GRBM and found a good-enough position to command the satellite to repoint to allow its
Narrow-field X-ray Instruments (NFI) to image the GRB region. Just eight hours after the GRB, two of the NFI cameras detected a bright, new X-ray source consistent with the WFC position. Three days later, that source had vanished.

Not only had the BeppoSAX team found the first afterglow of a GRB, the X-ray afterglow was positioned well enough (to 50 arcseconds in radius) to allow for sensitive searches of counterparts at other wavelengths. Just twenty hours after the GRB, a group led by Jan van Paradijs at Amsterdam University, the Netherlands, obtained a set of “deep” optical images of the WFC error location from La Palma Observatory on the Canary Islands (Spain). One week later, my collaborator (Nial Tanvir) obtained the next set of images of that field from the same observatory. A faint source in the NFI error box had vanished between those two epochs: it was the discovery of the first fading optical (i.e., visible-light) afterglow of a GRB. The optical position, determined to better than 1 arcsecond, was by far the best localization of any GRB, including the March 5th (SGR) event.

As the optical afterglow faded, the Hubble Space Telescope (HST) was trained on the position and found a blue blob around the fading afterglow. To most, this “nebulosity” looked a lot like a faint low-surface brightness galaxy, thus confirming the extragalactic hypothesis; but to the stalwarts of the Galactic model, it looked like an NS-blown

*The adjective deep is often used to described long exposures on the sky, where fainter and fainter objects can be detected with higher and higher signal-to-noise; these fainter images probe a larger depth of the Universe than “shallower” exposures.
1.6 The Afterglow Era Begins

bubble (confirming the opposite). Despite some claims from a few that the afterglow appeared to be moving on the sky (a nominal expectation of the Galactic model) and that the nebulosity changed color and shape (only possible in the Galactic model), all doubts about the distance scale of classical GRBs would be soon laid to rest.

On May 8, 1997, BeppoSAX spotted another GRB, and an optical afterglow was promptly discovered consistent in location and time with another fading X-ray afterglow. Astronomers at the California Institute of Technology (Caltech) obtained a spectrum of the optical afterglow using the Keck II 10-meter telescope in Hawaii; this telescope (and its twin Keck I) was the largest optical telescope in the world at the time. A quick inspection of that data led to one of the biggest discoveries in modern astrophysics: notched out of an otherwise smooth spectrum were a series of absorption lines, characteristic of iron and magnesium in gaseous form. But instead of seeing lines at the specific wavelengths as they would appear in a laboratory, they all appeared shifted to significantly redder wavelengths. This effect is called redshift, and the amount of redshift is usually given with the symbol $z$. The relationship between the observed wavelength $\lambda_o$ and the emitted wavelength of light $\lambda_e$ is $\lambda_o = \lambda_e \times (1 + z)$. Similarly, since $\lambda \nu = c$, the relationship between the observed frequency and the emitted frequency is $\nu_o = \nu_e \times (1 + z)^{-1}$; so frequency decreases and wavelength increases with increasing value of $z$. If we wanted to associate the redshift of a source with its apparent velocity $v_{app}$ away from us, we can use the equation $1 + z = \left(1 - \frac{v_{app}/c}{1 - v_{app}/c}\right)^{1/2}$, where $c$ is the speed of light. Within the Milky Way, objects show redshifts (and even blueshifts, with $z < 0$) which are very small ($|z| < 10^{-3}$, where $|v_{app}| < 300$ km/s) but objects at large distances from the Milky Way can show large redshifts, $z > 1$. The next footnote discusses the various mechanisms leading to the apparent redshift of light.
redshift effect was immediately recognized as an effect due to the expansion of the Universe,* and, given a reasonably well-prescribed mapping between the redshift measurement and distance, this observation established that GRB 970508 must have occurred from a distance of more than about 5 billion parsec (i.e., 5 Gpc) away. Some gas cloud in a distant galaxy lay between us and the GRB,† and its absorbing metals provided enough of a unique fingerprint to measure an unambiguous redshift.

In one fell swoop of the telescope, the thirty-year marathon to measure the distance scale of classical GRBs was won. The cosmological distance scale, the disfavored choice of so many for so long, had triumphed. To top it off, the first radio afterglow of a GRB was found following GRB 970508. A detailed inspection of the way in which the radio afterglow behaved, coupled with the then-known distance, strongly suggested that the afterglow-emitting surface was expanding at a rate close to the speed of light.

*An expanding universe tends to stretch out the observed wavelengths of light, causing redshift. The further away a source is on cosmological scales, the larger its apparent redshift. There are two other explanations for the origin of redshift. First, redshift can be due to relative velocity differences between the absorbing (or emitting) source and the observer. When moving apart, this redshifting is called a Doppler shift. Second, when light passes near any object with mass, its wavelength changes depending on the distance to the mass. Light emitted near the surface of a neutron star, for instance, is perceived to be more red by observers at progressively larger distances from the neutron star. This effect is called gravitational redshift and is a manifestation of General Relativity. In the case of GRB redshifts, no corroborating evidence for either of these two explanations are viable, leaving only the cosmological-expansion explanation.

†Recalling the discussion of the uncomfortable energetics that a cosmological distance scale would require, we might note Galileo’s admonition: “Facts which at first seem improbable will, even on scant explanation, drop the cloak which has hidden them and stand forth in naked and simple beauty.”
This relativistic expansion, which we will explore in §2.2.1, was a basic prediction of most cosmological theories of GRBs—another triumph for observers and theorists alike.

The seminal afterglow discoveries of GRB 970508 were first presented at a workshop less than two weeks later, conducted on the island of Elba, Italy. Franco Pacini and the BeppoSAX collaboration had organized that workshop following GRB 970228 and, as a student involved in the first optical afterglow discovery of GRB 970228, I was immensely thrilled to attend. Having not been part of the GRB 970508 results, I was honored to have been asked along with veteran theorist Mal Ruderman by the British journal Nature magazine to report on the Elba meeting in a News & Views article that accompanied the seminal papers on the GRB 970508 discoveries. We ended our article with a reflection on the accomplishment and a speculation on the future of the field: “It is hard to point to any other such important astrophysical problem that should be so quickly solved as the origin of gamma-ray bursts. But that will probably only be the first extraordinary chapter of a book that promises to become even more exciting.”

1.7 Progenitors and Diversity

By the end of 1998, more than twenty GRBs had been rapidly localized to the several-arcminute level, and six had confirmed cosmological redshifts. So, while the distance scale was indeed confirmed, understanding the “origin of gamma-ray bursts,” as we had perhaps zealously overstated in our article, was far from solved. With redshifts now
in hand, however, the energetics of individual events could provide an important constraint. The total energy release\(^*\) of most events was in the range of \(10^{51} - 10^{52}\) erg, comparable to the energy release in a supernova (\(\sim 10^{51}\) erg) and reasonably consistent with a variety of progenitor models where no more than a small fraction (\(\sim 10^{-2}\) \(M_\odot\)) of stellar mass of energy\(^†\) is released in the event.

However, the implied energy release of GRB 971214 threw the community for a major loop. Shri Kulkarni and his collaborators (including me) at Caltech discovered a high redshift of \(z = 3.42\) (hence a very large distance) for GRB 971214; the simplistic calculation suggested that the gamma-ray energy released during the GRB—over the course of just a few seconds—amounted to the equivalent of an appreciable fraction (>10 percent) of the entire mass of the Sun! News outlets ranging from the PBS News Hour\(^27\) to the Drudge Report proclaimed the event to be the “biggest bang in the universe recorded,” with an “intensity second only to the Big Bang.” Hate email from theorists, worried about observers overturning the tidy stellar-mass progenitor hypotheses, piled up in Kulkarni’s inbox. A relaxation of those uncomfortable energy requirements (and the tempers of theorists) would come the following year, as it was recognized that GRBs could

---

\(^*\)Assuming that the burst emitted the same amount of energy in all directions. See §3.3 for a discussion.

\(^†\)For explosions that involve nuclear, rather than chemical reactions, it is natural to describe the energy release (\(E\)) in mass-equivalent (\(M\)) quantities, using the conversion of \(E = Mc^2\) from Relativity Theory. Here \(c\) is the speed of light and \(M_\odot\) is the mass of the Sun. The Sun has an equivalent rest-mass energy of \(1.78 \times 10^{54}\) erg.
be collimated (or “jetted”). The implication was that the inference of energy release was too high. The geometry of GRB explosions is discussed more fully in §3.3.

On the opposite end of the energy spectrum, an otherwise normal-looking GRB (from a high-energy perspective) was discovered by BeppoSAX on April 25, 1998. Titus Galama, in the Netherlands, noted a curious bright source in the WFC error box, apparently in a spiral arm of a nearby galaxy. Instead of the rapidly fading optical afterglow of other GRBs, this source got brighter with time. Later recognized as a peculiar and bright supernova (designated SN 1998bw), the association both in place and time of this counterpart with GRB 980425 implied an energy release for the GRB several orders of magnitude smaller than the other GRBs previously studied. Nature, in producing the oddity that was GRB 980425, had provided another clear insight into a GRB progenitor. As the March 5th event showed us that NSs were capable of making some forms of GRBs, we now knew that massive stars (the likely origin of SN 1998bw) were also capable of making GRBs. Unfortunately, no convincing case of another extremely low-luminosity, 980425-like GRB has been found to date.†

The question of whether GRB 980425 resides on

---

* Supernovae are named in the order they are officially discovered in a given year, starting alphabetically with the designation “A.” After twenty-six SNe are found in a given year, two lower-case letters are used, starting with “aa.” SN 1998bw was the seventy-fifth SN found in 1998.

† To be sure, there have been a number of “underluminous” GRBs that bridge the energy release between GRB 980425 and other cosmological GRBs, GRB 060218 and GRB 100316D, both at low redshift and associated with supernovae, had total gamma-ray energies about 10–100 times more than GRB 980425. See, for example, R. L. C. Starling et al., *arXiv* 1004.2919 (2010).
the extreme tail of the “classical GRB” set or represents a physically distinct class of GRB-producing progenitors is still not a settled question.

Observations of the galaxy hosts of GRBs and the locations of GRBs around those galaxies provided strong circumstantial evidence that classical GRBs were more associated with younger stars than older stars. Primarily on two grounds this was troublesome for the popular view was that cosmological GRBs arise from the merger of NS binaries. First, many scenarios for NS coalescence posited a long delay (approximately one billion years) for the event after the birth of the progenitor stars. In that time, the star formation in that galaxy could subside. So why GRBs were all associated with blue galaxies (the observational hallmark of active star formation) was a mystery. Second, NS binaries were expected to travel far from their birthsite before coalescence, so GRBs from such progenitors would have no business being associated spatially with star formation.

The first direct evidence for a massive-star progenitor of a cosmological GRB came with my discovery of a curious supernova-like feature in the late-time light curve following GRB 980326. Many other “SN bumps,” some exhibiting supernova features to a high degree of confidence, were subsequently detected in other GRBs; this led to the speculation that most, if not all, classical GRBs could be connected to massive star explosions, rather than NS binary mergers. However, it was the

*This was true from 1997 to 2005. Even today only a minority fraction of GRBs are associated with older galaxies. See chapter 4.*
spectroscopic observations following GRB 030329, first by groups at Harvard University and in Copenhagen, that the “smoking-gun” evidence directly connecting classical GRBs to massive stellar death was uncovered. Embedded in the spectrum of the afterglow light of that GRB were telltale supernova spectral features. Amazingly, that supernova strongly resembled SN 1998bw, but unlike GRB 980425, this event occurred at a “cosmological” distance.

Both SN 1998bw and SN 2003dh (as the SN associated with GRB 030329 was designated) were of an unusual type of broad-line “Type Ic” SN, thought to arise from a star once more than 30–40 times the mass of the Sun but having expelled its hydrogen and helium envelopes as it neared the end of its life cycle. These are some of the most extreme SNe in nature coming from some of the most massive (and rare) stars. Massive stars are like the James Dean of the stellar population: they live fast and die young. Though they have much more fuel to burn than stars of mass comparable to the Sun, massive stars burn fuel at much higher rates than low-mass stars. So, when a group of stars of different masses is formed, the most massive stars expend their fuel first, typically within 10–50 million years after nuclear burning begins. Hence, when we say that GRBs are connected with young stars and ongoing star formation, we are also implying a connection to massive stars.

Broadly speaking, the association was a wonderful full circle for the field. The first theory of GRBs (before GRBs had been discovered) involved supernovae, and it was the association with SNe that Klebesadel and
his collaborators first tested. However, the mechanism whereby the supernova shockwave itself would produce a GRB cannot account for the energy release in most GRBs. Instead the observations were more consistent with a progenitor scenario—commonly referred to as the “collapsar” model—first put forward by Stan Woosley in 1993.\(^1\) In the current collapsar picture, a fast-moving jet is launched from a newly created BH at the center of a collapsing massive star. The GRB is produced as material in the jet shocks against itself. The SN is created as heat released from freshly minted radioactive elements powers\(^2\) the explosion of the star.

Supernovae have also been manifest after several other GRBs since 2003, but, like the GRBs themselves, there appears to be a large range in the basic properties of the SNe (energy releases, velocities, synthesized radioactive mass, etc.). In a challenge to the notion that all classical GRBs are from massive stars,\(^3\) two low-redshift GRBs were found in 2006 that did not have accompanying SNe to

---

\(^*\)We know now that their null result was due to a mismatch of the (large) distances of GRBs and the more nearby supernovae found in the 1960s and 1970s.

\(^1\)This original scenario failed to produce a supernova concurrently with the GRB but it was later refined.

\(^2\)Radioactive material is, by definition, unstable. Over time the nuclei of radioactive material lose protons and neutrons as that material transforms into different flavors of the same element (called “isotopes”) and to other elements altogether. In this process of decay, energy is released; and, since different isotopes decay at different rates, the power available due to radioactivity depends sensitively on the original composition of the radioactive material.

\(^3\)To be sure, most GRBs do not appear to be accompanied by supernovae-like events. However, most detected GRBs originate from \(z > 1\) where no SN features could have been discernible by current follow-up facilities: essentially, any SN that far away would be too faint to detect.
very deep limits (see §5.1.3). Still unsettled is whether these SNe-less GRBs represent yet another progenitor mechanism or simply are a manifestation of the original collapsar scenario where too little radioactive material is created to power an observable supernova.

The discoveries and insights that accumulated for the first eight years of the afterglow era were relevant to the majority of classical GRBs. Known since the 1970s was that a minority of events were especially short in duration and yet not apparently SGR flares. In 1992, observations with BATSE showed two distinct types of GRBs: the short events, those lasting less than about two seconds, appeared on average to have a “harder” spectrum than the long events (see figure 2.4 in the next chapter). By 2004, a bona fide “short-hard” GRB (SHB) had never been localized well by BeppoSAX or HETE-2 in part because the gamma-ray detectors were significantly less sensitive than BATSE to short-duration events. All the same questions persisted. Were the afterglows of SHBs the same as from the majority “long-soft” bursts (LSBs)? Was the distance scale also cosmological? Were SHBs also due to massive stellar death?

On May 9, 2005, the new NASA GRB satellite called Swift, responded autonomously to a short burst that had triggered its onboard gamma-ray detector called the Burst Alert Telescope (BAT). Swift was a novel facility in that BAT-discovered GRBs would not only be recognized quickly (<few seconds) by onboard computers and the discovery broadcast to the ground but also that the satellite itself would respond to the event by autonomously repositioning itself to capture the new GRB position with X-ray and ultraviolet/optical imagers. Pointing its
\textbf{X-ray Telescope (XRT)} at the GRB localization, Swift astronomers discovered the first X-ray afterglow of a short burst that day. With the precise X-ray location promptly relayed to the world, telescopes big and small were trained on the (<10 arcsecond) X-ray position. While no optical afterglow was found, the X-ray position was curiously close to a massive red galaxy containing little-to-no ongoing star formation. Though the physical association of GRB\textsuperscript{050509b} was not unassailable, the spatial coincidence with a “red-and-dead” galaxy was a remarkable departure from what was seen for long-burst galaxy associations. My group based at the University of California, Berkeley (and others) argued that the GRB must have originated from an older population of progenitors than most LSBs; and, given the appreciable physical offset of the X-ray position from the galaxy, one of the basic predictions of the NS-merger model had been borne out by those observations.\textsuperscript{32}

More than two dozen SHBs have been well localized since the summer of 2005, and the verdict is still out on their origin and even the rate of SHBs throughout cosmic time. As will be discussed in chapter 5, while not all SHBs are associated with red galaxies, the inference that they are due to an entirely different class of progenitors than that of LSBs seems to hold up. Unlike the collapsar scenario, which posits that a certain class of supernova should occur contemporaneously with a GRB, if the merger model is correct for SHBs, then a definitive observation confirming the model will be very difficult to obtain. The only obvious smoking gun for such progenitors—because they involve the violent coalescence of compact masses—would be the
detection of gravitational waves coincident with the GRB (see §5.2.3 and §6.4.1). Detection of such an event is a primary astrophysical impetus for the construction of next-generation gravitational-wave observatories capable of detecting NS mergers to large-enough distances in space.

1.8 Gamma-Ray Bursts in a Universal Context

Throughout this book, we explore our understanding of the diversity of GRB progenitors and the physics that gives rise to the prompt high-energy emission and the afterglows. Some of this understanding is a work in progress. Figure 1.4 shows some extragalactic progenitors that appear not only plausible but very likely to be involved in producing GRBs. Observational confrontation with theory has already opened up new vistas on the life cycle of massive stars and the nature of radiation processes from fast-moving material.

Yet irrespective of the theoretical underpinnings of the events themselves, there are some robust inferences about GRBs that hold vast implications for astrophysics as a whole. First, GRBs are for a brief time the brightest events in the Universe across the electromagnetic spectrum. At optical wavelengths, some of the brightest GRB afterglows are more than ten thousand times brighter than the brightest known quasar in the Universe and millions of times brighter than any supernova ever observed. Second, they occur in and around galaxies throughout the history of the Universe. Last, the progenitors that make GRBs were producing GRBs less than one billion years after the Big Bang. All these allow GRBs to serve as unique probes
Figure 1.4. Schematic of the various viable progenitors of cosmological GRBs. All involve the sudden release of energy in a small volume of space, triggered by the mechanisms noted in italics. Aside from extragalactic magnetars, the GRB arises from the rapid conversion of mass to energy. A black hole (BH) at the center of an accretion disk of in-spiraling material (shown at center) lies at the heart of many scenarios for energy liberation. The “central engine” of GRBs will be discussed in more detail in §2.3. Progenitor scenarios are discussed in chapter 5. The progenitors sizes are not shown to scale.

of other grand questions about the Universe that are not directly related to the events themselves.

While the afterglow light is bright, GRBs make for exceptionally useful lampposts: the light penetrates the gas and dust intervening in the line of sight from the burst location to the detector. Recall that the redshift of
GRB 970508 was found using absorption lines by gaseous metals. The detailed study of those lines, sometimes due to gas in galaxies that are spatially distinct from the GRB region, has started providing interesting views of the chemical state of galaxies in the distant past and information on how the chemical enrichment of galaxies is changing with cosmic time. The most distant GRBs hold the promise of revealing the state of the Universe itself, allowing us to measure the way in which the gas in the Universe transitioned from being in a fully neutral state to a fully ionized state. By the end of the first decade of 2000, GRBs were the redshift record holders, having catapulted past the most distant known quasars and galaxies with the discovery of a $z = 8.2$ event (GRB 090423). We explore in chapter 6 the ways in which GRBs are reinvigorating other pursuits in astronomy as well as opening up new possibilities to test long-held beliefs, such as the notion that the speed of light is constant in vacuum.

Like any maturing scientific endeavor, the development of an understanding of the phenomena that initially draws in the lessons from a variety of different disciplines eventually leads to new tools to study the world in a larger context. Along the way, there are triumphs and missteps. But, like all science, the story is never complete—we can only hope to close the chapters in front of us as we open up new ones.